Universität Regensburg
Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Mikroökonomik

Economic Experiments on Pro-Environmental Behavior and Gender Differences

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft
eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von:
Vanessa Schöller, M.Sc. with Honors

Berichterstatter:innen:
Prof. Dr. Andreas Roider
Prof. Dr. Lea Cassar

Tag der Disputation: 31.10.2023

For my mother.

Acknowledgement

Economics is shaped by individuals and their actions. Aside from the numerous anonymous study participants, I would like to acknowledge the support of everyone who accompanied me over the last four years.

First, I want to emphasize the countless support of my first supervisor Professor Andreas Roider. While developing research ideas and pursuing research projects, he helped me excel by providing valuable feedback and the freedom to explore. I would also like to thank my second supervisor Professor Lea Cassar for the guidance on experimental design and the fruitful discussions to position my research. I am immensely grateful to Professor Jana Gallus for her warm hospitality and the invaluable opportunity to visit her at UCLA. The research visit proved to be an incredibly enriching and transformative experience, fostering extensive discussions that have left a lasting impact on my Ph.D. and beyond.

Second, thanks are due to the Elite Network of Bavaria and its international doctorate program "Evidence-Based Economics" for generously funding the dissertation project. Conducting scientific experiments with numerous participants and presenting at international conferences was only possible due to the financial support received. I also thank the Coordination Office for Equal Opportunities of the University of Regensburg for partially funding my research visit at UCLA. I am grateful for the Bavarian Graduate Program in Economics and the inspirational courses offered. They were a source of knowledge that shaped the scope of my dissertation. In addition, I thank the participants at the following conferences for their valuable comments: Young Economists' Meeting, Brno; 12th International Conference of the French Association of Experimental Economics, Lyon; 6th Solomon Lew Conference on Behavioral Economics, Tel Aviv; 2022 North-American Economic Science Association Conference, Santa Barbara; Bavarian Micro Day, Munich; and Workshop on Behavioral Environmental Economics, online.

Third, I am deeply grateful to my colleagues. I owe special thanks to Clara, Helena, Alex, Brit, Silvio, and Lars for the collaborative efforts bringing forth our joint research projects. It was a great pleasure to work with you at any project stage and celebrate the publications that shaped my dissertation.

And last, my friends and family have given me their time and steady support, for which I will be ever thankful. Thank you, Mom and Nicolai, for always being there and giving me a foundation to build upon. Special thanks go to Dani for celebrating the highs and lifting me up in the lows. I can always count on you and am grateful to have you in my life.

Contents

List of Tables

List of Figures

Chapter I

Introduction

We should teach the students, as well as executives, how to conduct experiments, how to examine data, and how to use these tools to make better decisions.

Dan Ariely

As we navigate the 21st century, society encounters numerous environmental, social, and economic challenges that are increasingly complex and amplify each other. According to the World Health Organization (WHO), global warming poses the most existential threat to worldwide health. Global warming is expected to lead to a higher occurrence of extreme weather events such as floods, droughts, and storms, along with rising temperatures, increased heat-related fatalities, outbreaks of infectious diseases, and detrimental effects on water quality and crop yields (Intergovernmental Panel on Climate Change, 2022, hereafter IPCC). Already in 2030, the WHO estimates direct costs due to global warming to be between two to four billion US dollars a year (WHO, 2021).

Economic research provides answers for the multifaceted challenges of our time, aiming for a detailed understanding of the situation as it is and shaping the future. Environmental economics views climate change as a market failure that opposes efficient allocation (Stern, 2007). Greenhouse gas emissions represent a global negative externality that is insufficiently internalized. As a result, social welfare is diminished since too much carbon is emitted (Chichilnisky and Rezai, 2020). To internalize the external costs of emissions, economists advocate for policy interventions such as introducing a global sector-wide emissions trading scheme or taxing greenhouse gas emissions. To date, only 23% of all greenhouse gas emissions are covered by emissions trading schemes or taxes (World Bank, 2023). Moreover, carbon prices are mostly below the recommended levels to limit global warming to below 2°C (Pathak et al., 2022). As an emissions trading scheme defines an upper limit on the amount of greenhouse gasses emitted, it effectively reduces emissions while supporting flexibility in doing so. Companies have to possess allowances corresponding to their emissions, and those seeking to increase emissions must purchase allowances from other firms. The market-based system puts a price on greenhouse gas emissions that is the same for all market participants. In addition to the

emissions trading system that maintains a fixed total emissions level, a tax-based regulatory approach establishes a fixed price for emissions (Weitzman, 1974). The extent of emission reduction depends on the tax rate applied (Taschini et al., 2013). To maximize market efficiency, the tax rate should reflect the actual cost of emissions (Atkinson and Stern, 1974). Carbon taxes and trading systems both create incentives for polluters to reduce their emissions by putting a price on carbon emissions. The price signal also decreases demand for goods and services associated with high emissions and stimulates investment in low-carbon technologies. Moreover, both approaches lead to an efficient market quantity of emissions if uncertainties regarding the costs and benefits of emission reductions are absent (Taschini et al., 2013). However, both incentive-based approaches can benefit from leveraging insights from behavioral economics. Guiding the decisions of producers and consumers through price signals is most effective when considering individual decision-making (Camerer, 1999; Sunstein and Thaler, 2008). For example, introducing a carbon tax not only sets financial incentives but also conveys social values by framing carbon-intensive behavior as undesirable (Croson and Treich, 2014). Furthermore, no international organization can currently compel states to implement carbon mitigation systems (Altemeyer-Bartscher et al., 2010). As a result, voluntary pro-environmental actions become indispensable, even if they do not guarantee optimal solutions (Kesternich et al., 2017).

By applying behavioral economics to environmental questions, we can gain a deeper understanding of how policy interventions affect human behavior, how to encourage mitigating behavior, and how to design environments that encourage sustainable shifts in behavior (Gifford et al., 2011). The IPCC estimates that behavioral change, supported by appropriate policies, has the potential to rapidly reduce global greenhouse gas by a minimum of 5%. By implementing comprehensive strategies focusing on demand, global greenhouse gas emissions could be reduced by 40-70% by 2050 (IPCC, 2022). Hence, behavioral considerations play a vital role in climate change mitigation, especially in bridging the transition period until market approaches are adopted, and low-carbon technologies are refined (Dietz et al., 2009). Numerous studies demonstrate that behavioral interventions can yield substantial positive environmental outcomes (e.g., Allcott, 2011; Delmas et al., 2013; Ferraro and Price, 2013; Kallbekken and Sælen, 2013; Pichert and Katsikopoulos, 2008; Sunstein and Reisch, 2014).

Behavioral economics delivers insights into how and why people behave as they do. It acknowledges that an individual's decision-making is impacted, among others, by emotions, lack of willpower, prevailing circumstances, social preferences, and limited cognitive processing capacity (Samson, 2014). Consequently, humans deviate from predictions made by neoclassical theory (Mullainathan and Thaler, 2000). Behavioral economics incorporates economic theory with a psychological foundation to enhance its explanatory and predictive performance (Angner and Loewenstein, 2007). The resulting economic theory aims to be accurate in assumptions and predictions with reality, generally applicable across various contexts, and tractable, which means an analytical solution

should exist (Camerer et al., 2004). Economic experiments are a method used in behavioral economics that allows examining human decision-making in a controlled way to test theory predictions, reveal underlying principles, and identify causal relationships. Research in behavioral and experimental economics is growing, and its importance is underlined by various awarded Nobel prices, e.g., Vernon L. Smith and Daniel Kahneman (2002), Elinor Ostrom (2009), Alvin E. Roth (2012), Richard H. Thaler (2017), and Abhijit Banerjee, Esther Duflo, and Michael Kremer (2019). Improved forecasts regarding economic behavior also allow for more effective policy recommendations (Camerer, 1999). Several governments and international organizations have set up advisory groups that provide behavioral insight to support policy-making. For example, the European Union (EU) founded the "Competence Centre on Behavioural Insights", and the Organization for Economic Co-operation and Development (OECD) established the "OECD Network of Behavioural Insights Experts in Government" that connects over 100 government officials in more than 40 countries (European Commission, 2022; Observatory of Public Sector Innovation, n.d.).

Against this backdrop, this dissertation is motivated by the importance behavioral and experimental economics can have in shaping a better future. In all my research projects, I conduct economic experiments to deliver relevant behavioral insights. Chapters II and III of this thesis lie at the intersection of behavioral and environmental economics and identify barriers to pro-environmental behavior and how to overcome them.

Chapter II is motivated by the increasing number of companies that publish information about their sustainability in annual reports and on products. Providing information can complement market-based approaches and is frequently used to induce sustainable behavior, such as printing eco-labels on products. However, it is unclear which information induces pro-environmental behavior most effectively. Therefore, we conduct an experiment to compare the effect of carbon display in kilograms, abatement costs, and social costs, respectively, on individuals' purchasing decisions. It is found that the type of display has no significant impact on purchasing decisions, which is contrary to previous literature. Nevertheless, most participants believe that social cost information leads to the largest carbon reduction by consumers, and many prefer this information. The results can inform managers and policymakers interested in setting standards for CO_2 information.

Chapter III studies the dynamic effects of pro-environmental behavior. When designing behavior change interventions, dynamic effects are often overlooked. This chapter investigates moral balancing in pro-environmental behavior and whether it occurs only for substantial moral acts or even if the positive effect is negligible. In a two-stage economic experiment, we find that participants who successfully acquired a moral license offset less carbon than those who failed. The experiment exogenously varies the magnitude of the initial pro-environmental act and discovers that the magnitude does not systematically affect moral balancing. Participants with the greatest environmental con-

cerns do not engage in moral balancing, which indicates that environmental concerns moderate moral balancing.

The adverse impact of climate change is disproportionately borne by disadvantaged groups such as women (Habtezion, 2016). Therefore, climate change highlights gender inequality as another pressing issue society faces. Women encounter constrained opportunities in various domains of life, including education (Myers and Griffin, 2019), employment (Cortés et al., 2021), and leadership positions (United Nations Entity for Gender Equality and the Empowerment of Women & Department of Economic and Social Affairs, 2022, hereafter UN Women & DESA). According to the United Nations (UN), female representation in science, technology, engineering, and mathematics (STEM) fields is disproportionately low, with fewer women completing STEM school subjects, earning STEM degrees, entering STEM professions, and occupying senior leadership and academic positions within STEM disciplines. Women represent just 35% of STEM students and hold only 20% of STEM jobs globally (UN Women & DESA, 2022). Women are also underrepresented in the paid workforce; e.g., the EU reports a gender employment gap of 10.8 percentage points in 2021 (Eurostat, 2023). The gender employment gap in the EU leads to an estimated annual economic loss of 370 billion Euros, emphasizing that inequality has both adverse societal implications and significant economic consequences (Eurofound, 2016). Moreover, the gender pay gaps persist, and women in the EU generally earn about 11.4% less than their male counterparts for similar work in 2018 (Leythienne and Pérez-Julián, 2021). Across 39 countries worldwide, women only accumulate on average 74% of the wealth men possess by the end of their working careers (Willis Towers Watson, 2022). Women are underrepresented in leadership roles (i.e., hold only about 28% of managerial positions in 2020) and face significant hurdles in accessing decision-making positions (UN Women & DESA, 2022). Representation of women in parliament stands at 27% globally in July 2023, and less than a quarter of Cabinet Ministers (23%) are women (Inter-Parliamentary Union & UN Women, 2023).

Despite remarkable improvements toward gender equality in recent years, equality is not yet reached, and differences in labor market participation, earnings, and access to leadership positions prevail. Based on the current pace of advancement, the UN estimates that it may take up to 140 years to achieve gender parity in leadership positions in the workplace and at least 40 years to accomplish equal representation in national parliaments (UN Women & DESA, 2022). Achieving greater gender equality by 2050 is estimated to increase per capita GDP in the EU by about 6 to 10% (European Institute for Gender Equality, 2023).

To develop solutions that effectively promote gender equality (e.g., affirmative actions such as gender quotas), it is crucial to identify the factors driving gender inequality. Factors frequently discussed are, e.g., discrimination (e.g., Altonji and Blank, 1999), gender norms (e.g., OECD, 2015) and stereotypes (e.g., Ellemers, 2018), child penalty (e.g., Kleven et al., 2019), and behavioral traits (e.g., Marianne, 2011). A vast liter-

ature in behavioral economics shows that willingness to compete and risk preferences measured in experimental settings have external relevance and predict educational and occupational decisions, along with real-life earnings (e.g., Buser et al., 2014, 2017, 2018, 2021; Cortés et al., 2021; Reuben et al., 2015; Zhang, 2013). Since various papers show that men are more willing to compete (e.g., Beblo and Markowsky, 2022; Niederle and Vesterlund, 2007), more risk-seeking (e.g., Charness and Gneezy, 2012; Thöni and Volk, 2021), and less altruistic (e.g., Andreoni and Vesterlund, 2001; Bilén et al., 2021) than women these behavioral gender differences could partially explain observed gender inequalities in education and labor market outcomes (Marianne, 2011). But how robust are the reported gender differences in willingness to compete, risk-taking, and altruism? And how much can be associated with gender, and how much with an individual's biological sex?

Chapter IV is motivated by the vast literature that identifies gender as a key driver of economic decision-making. In an experimental online study with cis- and transgender participants, we test for correlational differences between gender and sex for competitiveness, risk-taking, and altruism by comparing decisions across these different subject groups. Moreover, participants are primed with either a masculine or feminine gender identity to examine causal gender effects on behavior. We hypothesize that if gender is indeed a primary factor for decision-making, (i) individuals of the same gender (but different sex) make similar decisions, and (ii) gender priming changes behavior. Based on 780 observations, we concluded that the role of gender (and sex) is not as decisive for economic behavior as originally thought.

The dissertation is structured as follows: Chapter II presents the first research project on different types of carbon information and if they can promote pro-environmental behavior. Chapter III provides insights into the second research project analyzing dynamic effects in pro-environmental decision-making. Chapter IV reports on the third research project investigating the robustness of gender differences in economic decision-making. Chapter V concludes by summarizing the mentioned research projects, exploring limitations, and suggesting avenues for future research.

Chapter II

Can monetized carbon information increase pro-environmental behavior? Experimental evidence[1]

1 Motivation

Since the Paris Agreement in 2016, participating countries have been obliged to limit the increase in global average temperature to 1.5°C compared to pre-industrial levels (UN Framework Convention on Climate Change, 2015). Despite various actions to achieve this goal, global emissions have increased over the last years, and the risk of missing the 1.5 degrees goal is high (UN Environment Program, 2019). To date, policy interventions that price emissions have been implemented and proven successful, such as emission trading and carbon taxes (e.g., Andersson, 2019; Martin et al., 2016; Sen and Vollebergh, 2018; Venmans, 2012). Providing information complements market-based approaches and is frequently used to induce sustainable behavior, e.g., through eco-labels on products[2] (e.g., Big Room Inc., 2021).

Moreover, employing renewable energy sources and changing national laws and policies takes time, whereas introducing changes in behavior through nudges can potentially be accomplished much faster (Wee et al., 2021; Wynes and Nicholas, 2017).

Sustainability reporting has been increasingly used by companies in recent years (e.g., Dhaliwal et al., 2011; KPMG, 2017; Maas et al., 2016), and it is also part of the UNs' Sustainable Development Goals since 2015 (UN, 2020). Thereby, sustainability reporting refers to "an organization's practice of reporting publicly on its economic, environmental, and/or social impacts, and hence its contributions – positive or negative – towards the goal of sustainable development" (Global Reporting Initiative, 2020*a*, p.3).[3]

[1] This chapter is based on Schöller and Ulmer (2023) published in Ecological Economics.

[2] Eco-labels are defined as "an official symbol that shows that a product has been designed to do less harm to the environment than similar products" (Cambridge Dictionary, n.d.).

[3] Other terminologies include triple bottom line reporting (Elkington, 1997), full cost accounting (Atkinson, 2000), corporate social responsibility (CSR) reporting, non-financial reporting and environmental, social and governance (ESG) reporting (Global Reporting Initiative, 2020*b*). A definition for the term sustain-

For example, companies use sustainability reporting to legitimize corporate activities, to improve corporate reputation and branding, to gain and signal competitive advantages, to be transparent and accountable, to motivate employees, and as an internal information and control system (e.g., Herzig and Schaltegger, 2011). Moreover, potential investors may positively react to information on corporate social responsibility (CSR). In particular, Martin and Moser (2016) show in an experimental setup that investors' valuation is higher if the company reports on their CSR investments. In their setup, managers anticipate the interest of potential investors and act accordingly by disclosing their investments. Interestingly, managers prefer highlighting the social benefits of CSR investments instead of their costs. Even though it is unlikely that such reporting will lead to a complete internalization of negative externalities (Cohen and Viscusi, 2012), Maas et al. (2016) remark that various stakeholders, including governments, predict that sustainability reporting may induce firms to become more sustainable.

Similarly, consumers frequently encounter eco-labels in their daily life. In the near future, eco-labels focusing on carbon emissions will likely be increasingly used as consumers demand them (Carbon Trust, 2019), governments encourage their use (Li et al., 2017), and companies have already started to provide them, thereby putting pressure on their competitors to do the same. For example, Oatly and Quorn indicate the CO_2 content of their products, and Unilever started to do so for all its products in 2021 and will extend carbon labeling to its entire product range within the next five years (Cohen, 2021). Moreover, consumers are informed about the CO_2 emissions of their flights when searching with Google flights (Compton, 2021) and Klarna, an online financial service provider, lists carbon emission estimates for every purchase (Klarna Bank AB, 2021). This specific form of nudge (Thaler and Sunstein, 2021) aims to inform consumers about the environmental consequences of their purchase and can lead to choosing low-carbon products. In fact, Lehner et al. (2016) document in their review that simplifying and framing information can lead to more sustainable behavior, but its success varies with the context considered. Furthermore, a majority of individuals state that they aim to act environmentally friendly, but the actual fraction of people taking pro-environmental actions falls short (Kollmuss and Agyeman, 2002; Prothero et al., 2011). One potential reason for this "value-action gap" (Blake, 1999) could be that individuals fail to process the available information. Nudges, for example, in the form of carbon labels, could reduce this gap and lead to more sustainable behavior.

However, it is still an open question how a firm's sustainability should be reported. Furthermore, it is unclear which type of information supports consumers best in translating an environmentally friendly attitude into environmentally friendly behavior. In the firm as well as the consumer context, a large variety of possible measures exist. With respect to reporting, there is currently a harmonization in the large number of existing models, metrics, and typically complex approaches to measure and report on a

able development is given by the World Commission on Environment and Development (1987) as "meeting the needs of the present without compromising the ability of future generations to meet their own needs" (p. 24).

company's impact (Barby et al., 2021). Concerning eco-labels, we also see efforts to standardize the large variety. In particular, the European Commission aims to develop a standard to assess the environmental footprint of products and services, which corporates will have to use to substantiate their claims (European Commission, 2020).

Expressing environmental and social effects in monetary terms is one particular approach discussed.[4] Whereas monetary valuation and aggregation is less frequently applied in other fields (e.g., in natural science), it is widely used in economics (Singh et al., 2009). Particularly, monetization is commonly applied in cost-benefit analyses, which are typically used for public good decisions (e.g., Boardman et al., 2017; OECD, 2006), and occasionally used for life cycle assessments (LCA) (Pizzol et al., 2015). In both cases, costs and benefits that ought to be considered in the decision-making process are initially not in financial terms. Non-monetary factors, such as air or water quality, are monetized through estimation based on economic tools (e.g., contingent valuation studies, the travel cost method, or hedonic pricing, see, e.g., Ness et al., 2007). Eventually, costs and benefits have the same unit and can henceforth be directly compared. In principle, both environmental (e.g., air quality) and social factors (e.g., health and safety) can be monetized, but currently, most companies that have already experimented with monetary impact valuation focus on environmental impacts (Serafeim et al., 2020).

Whether environmental impacts *should* be monetized is a debate beyond the scope of the paper.[5] Being aware of this criticism, we would like to stress that monetization is currently used and will potentially be used even more often as firms strive to report on their societal impacts instead of their resource use (KPMG, 2017). Moreover, it might be that sustainability receives more attention when monetized (Herbohn, 2005) and might lead to easier communication of different divisions in a firm due to a common unit of measurement (Bebbington et al., 2007).

It is, therefore, essential to study how individuals decide once monetized information is available. To the best of our knowledge, Hummel and Hörisch (2020) is the only other study that systematically compares individuals' choices when facing different types of carbon information, including monetized carbon impacts. In particular, Hummel and Hörisch (2020) conduct a non-incentivized survey experiment with German business students and focus on the difference in the willingness to invest in two types of hypothetical machines, one having a high, the other a low environmental efficiency. The carbon intensity of the machine is displayed in three ways relative to a benchmark machine: qualitative (better/worse), kilograms CO_2, and in terms of abatement costs (costs faced if allowances at the EU Emissions Trading System (ETS) would be bought). Hummel and Hörisch (2020) find that qualitative and kilogram CO_2 information leads to significantly larger changes in the reported willingness to invest compared to abatement cost information.

[4] For overviews of sustainability assessment methods, see, e.g., Ness et al. (2007); Singh et al. (2009).

[5] For an overview, see, e.g., Atkinson and Mourato (2008). The critique includes method specific concerns of cost-benefit analyses (e.g., Ackerman and Heinzerling, 2002; Baram, 1979) and sustainability reporting (e.g., Gray and Milne, 2002; Milne and Gray, 2013) in general and monetization in particular (e.g., Gsottbauer et al., 2015; Kallis et al., 2013, 2015).

We contribute to the literature by investigating which type of information on carbon emissions most frequently leads to pro-environmental decisions. For that, we conduct an online experiment where participants can purchase an emission-intensive virtual product. The product's carbon intensity is displayed either in kilograms, abatement costs, or social costs. We extend the study by Hummel and Hörisch (2020) in various aspects. First, our study is, to the best of our knowledge, the first that examines two types of monetized information: abatement costs and social costs. Second, we incentivize individuals' actions in the experiment, i.e., their decisions have real-world consequences. Third, we consider a more heterogeneous subject pool, which might provide insights into the external validity of the previous results. Fourth, subjects decide to buy or not to buy a product instead of reporting their willingness to invest. Stating one's purchase decision is arguably less cognitively demanding and therefore less susceptible to inconsistencies (Auger and Devinney, 2007; Ryan and San Miguel, 2000).

We find that the type of information display has no significant impact on the likelihood of purchasing an emission-intensive virtual product. The likelihood of purchasing the product tends to be lower when carbon emissions are displayed in abatement costs, but the difference is not statistically significant. Importantly, our results contrast previous literature by Hummel and Hörisch (2020) both qualitatively and by its statistical significance. In particular, we do not replicate their finding that information on kilograms CO_2 leads to significantly stronger effects on decisions than information on abatement costs. Our results stress that future research is required to understand the drivers of the observed differences and investigate their robustness. Moreover, our evidence suggests that monetized carbon information does not lead to pro-environmental behavior per se, but it also does no harm. Studying heterogeneous effects, we find that the specific type of information leads to similar purchase decisions for almost all considered subgroups. Our post-experimental questionnaire reveals that subjects believe that carbon measured in social costs has the largest potential to change behavior. Moreover, in their role as consumers, many subjects reveal to prefer this type of information. Thus, there seems to be a substantial gap between the interest in social cost information and its effectiveness.

The rest of the paper is structured as follows. We first review the related literature in Section 2. Our experimental design and hypotheses are described in Section 3 and Section 4, respectively. We present the results in Section 5 and conclude in Section 6.

2 Related literature

Our paper contributes to two strands of literature: (i) investigations on individuals' decisions if they learn about a firm's sustainability performance, and (ii) studies on the relative effectiveness of different types of information to induce sustainable behavior.

2.1 Individuals incorporate environmental performance of firms into their decisions

Firms can provide information on their sustainability performance to external stakeholders on two levels: on the firm level, for example, via external reporting, and on the product level, for example, through eco-labels. Both types of information may influence individuals' decisions, such as their willingness to invest in a firm or purchase its products.

On the firm level, laboratory experiments show that individuals have a higher willingness to pay for goods produced by sustainable firms, but individuals do not fully compensate firms for their higher costs (Barreda-Tarazona et al., 2011). Moreover, individuals display a higher willingness to invest in sustainable firms but similarly do not reward firms for their higher costs (Bonnefon et al., 2022). Thereby, consumers respond stronger to irresponsible corporate behavior than to responsible behavior (Bonnefon et al., 2022), and they reward corporate social performance only in a good state, but not in a bad state of the economy (Brodback et al., 2020).

Based on a natural experiment exploiting the introduction of mandatory disclosure of environmental, social, and governance (ESG) performance, Grewal et al. (2019) show that investors reward the sustainable performance of firms through higher market valuations.

On the product level, eco-labels may influence individuals' behavior. Overall, the evidence to date shows that individuals are willing to pay a premium for products that are classified as sustainable (e.g., Echeverría et al., 2014; Michaud et al., 2013). However, the vast number of available eco-labels, each having its own meaning and standard, might reduce their effectiveness (Yokessa and Marette, 2019). Moreover, Leire and Åke Thidell (2005) note that further research is required to investigate when and why individuals do (not) take eco-labels into account.[6]

We add to this literature by conducting an incentivized experiment to investigate whether individuals incorporate information on carbon emissions into their decisions. Thereby, we focus on types of information that are increasingly used for reporting on the firm level and will potentially also be used for eco-labels on products.

2.2 Types of information and their effectiveness in inducing sustainable behavior

The exact type of information available to decision-makers might be crucial to induce sustainable behavior. In particular, the unit and precision of display might alter individuals' behavior: Do individuals react more to information on the financial aspects of consumption or on carbon emissions? Do individuals react stronger to precise carbon

[6] A related strand of literature investigates whether individuals avoid information on purpose, e.g., by not taking the time to read the information. In case information on sustainability is freely available and decisions have real-world consequences (like in our experiment), the evidence suggests that the risk of information avoidance is low (Lind et al., 2019; Pace and van der Weele, 2020).

emissions in kilograms CO_2 than when environmental impacts caused by the emissions are imprecisely stressed? How effective is information on abatement costs (e.g., to offset emissions) or social costs?

The meta-analysis by Delmas et al. (2013) suggests that feedback on financial aspects and financial incentives[7] can even lead to increases in energy use. In contrast, stressing emissions' environmental and health impacts is more effective in decreasing energy consumption (Asensio and Delmas, 2015, 2016; Chen et al., 2016) and leads to higher enrollment rates in an energy-saving program (Schwartz et al., 2015). According to Schwartz et al. (2015), this can be explained by individuals thinking about financial aspects themselves, whereas thinking about environmental reasons to preserve energy needs to be activated by an environmental framing. Moreover, individuals report to feel better when receiving environmental messages (Bolderdijk et al., 2013). Interestingly, they nevertheless prefer information on financial aspects (Schultz et al., 2015).

Few studies do not only stress environmental impacts but provide exact carbon emission information associated with one's actions. Compared to information on financial aspects, information in terms of kilogram CO_2 tends to be more effective, but the difference is less strong than when stressing environmental and health impacts. In particular, information on financial aspects and CO_2 have no direct effect on energy savings (Steinhorst and Klöckner, 2018). However, CO_2 information leads to higher intrinsic motivation than information on financial aspects (Steinhorst and Klöckner, 2018). Moreover, carbon information increases climate change salience (Spence et al., 2014). The increase in intrinsic motivation and salience might lead to pro-environmental decisions in the long run.

Contrary to providing information on the financial aspects of consumption or potential energy savings, the environmental effect can be expressed in monetary terms. For example, carbon emissions can be monetized based on abatement costs to mind or offset the emissions or based on the social costs induced by the emissions. Hummel and Hörisch (2020) display the carbon efficiency of a hypothetical machine in both kilograms CO_2 and its associated abatement costs, i.e., the costs of allowances required to compensate the emissions. They find that kilogram CO_2 information has a stronger effect on the reported willingness to invest.[8]

Monetization based on social costs has, to the best of our knowledge, not been studied yet. However, we hypothesize that monetization based on social costs is perceived differently compared to monetization based on abatement costs. In particular, social costs are an estimate of the social harm caused by carbon emissions, which could be perceived rather like stressing future health impacts of current consumption like in Asensio and Delmas (2015) and Chen et al. (2017).

We contribute to the existing literature by investigating the effectiveness of infor-

[7] For a review on financial incentives, see, e.g., Maki et al. (2016); Mi et al. (2021).

[8] The authors further test a qualitative information measure, i.e., whether the machine is more/less efficient than a benchmark machine, which shows similar effects as kilogram CO_2 information. Since qualitative information has already been tested, we did not examine it in this study.

mation about kilograms CO_2, abatement costs, and social costs on sustainable behavior. Extending previous literature, we investigate social cost information, which will increasingly be used but has not been studied yet. Moreover, abatement and social costs incorporate monetary and environmental aspects, which might lead to stronger impacts on individuals than information on financial aspects.

3 Experimental design

In an online experiment, we recruited 600 participants via Prolific [9] on the 17th and 18th of May 2021. The experiment is implemented in oTree (Chen et al., 2016). Participants are randomly assigned to one of three treatment conditions, each having 200 observations. Participation is only restricted to British individuals (country of birth) that live in the United Kingdom (country of residence) to ensure that individuals understand the English instructions well and are equally familiar with the measurement units. As we study different types of carbon information, using an online experiment is well suited due to the more diverse subject pool compared to experiments restricted to students (Charness et al., 2013; Crump et al., 2013; Horton et al., 2011). Moreover, experimenter demand effects are lower given the higher level of anonymity in online studies compared to laboratory settings (de Quidt et al., 2019; Mummolo and Peterson, 2019).

The experimental design is based on Pace and van der Weele (2020), who study the connection between beliefs, uncertainty, and information in the context of carbon emissions. Similar to their study, individuals decide on purchasing a carbon-intensive virtual product.[10] The purchase leads to a positive payoff for the participant, mimicking consumer surplus, as well as CO_2 emissions. The carbon emissions are implemented by canceling a donation to a carbon-reducing charity if the product is purchased (details see below). We use an incentivized setting as it has been shown that the hypothetical bias can skew results (e.g., Harrison, 2006; Löschel et al., 2013; Murphy et al., 2005).

To study the impact of alternative carbon measures on behavior, our treatments vary how the product's carbon emissions are displayed. The experiment is set up in six stages (see Figure 1), for the complete instructions, see Appendix A.2. To complete the study, participants need, on average, about 12 minutes and receive a show-up fee of £1.9. The experimental design, as well as the planned analysis, were preregistered on As-Predicted.org (No. 65739) before data collection (see https://aspredicted.org/y7tu5.pdf).

[9] Prolific is an online platform that connects researchers with more than 130,000 individuals from OECD countries willing to participate in online studies to earn money. When signing up, participants answer demographic questions that researchers can use to pre-screen and restrict access to their study. Eligible participants are not informed about the applied screening categories. Studies show that compared to other online research platforms, such as Amazon Mechanical Turk, participants on Prolific produce responses of higher data quality (Peer et al., 2017, 2021).

[10] An alternative design would be a choice between two products: one high in price and low in carbon emissions, the other low in price and high in carbon emissions. However, this design would only represent certain real-life decisions since there are products that are both expensive and environmentally harmful (e.g., beef or dairy products) and others that are cheap and environmentally friendly (e.g., lentils, beans, or oats).

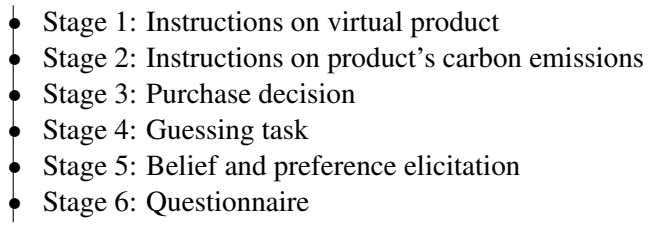


Figure 1: Stages of the experiment.

In Stage 1, participants are informed that they can purchase a virtual product, which yields an additional payment of £1. The additional payment mirrors the consumer's surplus and is computed as the difference in the product's value of £2 and its cost of £1. Compared to the show-up fee of £1.9, the additional payment increases the total payment by more than 50%. In case the product is not purchased, individuals do not receive an additional payment.

In Stage 2, the product's carbon emissions are explained. For each participant, we arranged a donation to the non-profit organization Compensators*[11] that would result in a carbon offset. If the product is purchased, this donation is canceled, effectively leading to increased carbon emissions (similar to Pace and van der Weele, 2020). Compensators* offsets emissions by purchasing allowances from the EU ETS without emitting CO_2, thereby ensuring that companies in the European Union cannot release these CO_2 emissions.[12] We chose the CO_2 reduction via the EU ETS as it does not lead to positive external effects such as a potential increase in biodiversity, which would, for example, be the case when planting trees. Therefore, we do not need to assume that participants' preferences do not interact with the type of carbon information. During our study period, Compensators* reduced one ton of CO_2 at a constant price of 45.45 €. We donate 1.40 € (£1.20)[13] if the individual does not purchase the product. This amount is more than the individual would receive as an additional payment to ensure that individuals have no incentive to purchase the product and donate outside the experiment (which we cannot observe).

It is crucial that participants believe the instructions, i.e., that any emissions are real, and that they are not misled. We ensure that the implementation of the carbon offset is credible for participants in two ways. First, we inform participants that the study does not use deception, in particular, that their decision has real-world consequences implemented as described in the instructions. Second, subjects are informed that they will receive a link via Prolific private message that includes the calculation of the total donation amount and the official donation receipt. We inform participants that we will publish the link on our official university websites to increase credibility. We include control questions on both aspects.

[11] https://www.compensators.org/en/compensators/

[12] Due to the market stability reserve, the actual carbon reduction might even be higher than the purchased allowances (European Commission, 2019). As the potential donation is identical in all treatments, the mechanism does not bias our estimates and is, thus, not further considered.

[13] Exchange rate £ to €: 1.1639 on May 10th 2021 (Exchange Rates UK, 2021).

The three treatments in the between-subject design differ in the unit in which the product's carbon emissions are displayed: kilograms and reference value (henceforth KILOGRAM), abatement costs (ABATEMENT COSTS), and social costs (SOCIAL COSTS). In each treatment condition, the respective carbon measure is briefly explained to the participants, and we include control questions to ensure that they understand the instructions.[14] For the relevant parts of the instructions, see Table 1.

Table 1: Overview of treatments.

Treatment	Information on carbon emissions	Carbon emissions of the product
KILOGRAM	The product's carbon emissions are measured in kilograms (kg). The kilograms express the mass of CO_2 molecules emitted into the atmosphere. For example, one emits 1 kg of CO_2 when burning about 0.43 litres of petrol.	If you buy the product, you will emit 30.80 kg of CO_2 – the mass of CO_2 molecules emitted into the atmosphere. This is equivalent to burning 13.33 litres of petrol.
ABATEMENT COSTS	The product's carbon emissions are measured by its abatement costs. The abatement costs express the amount one would need to invest to offset its emissions. Carbon offsetting means to compensate for the product's emissions by funding an equivalent CO_2 saving elsewhere.	If you buy the product, you will emit CO_2 that will lead to abatement costs of £1.20 – the amount one would need to invest to offset the emissions.
SOCIAL COSTS	The product's carbon emissions are measured by its social costs. The social costs express the estimated social harm caused by the carbon emissions in £ (for example, the monetary equivalent of damages caused by coastal floodings due to rising sea levels). As multiple ways of estimation exist, the social costs can be of different levels within an estimated range.	If you buy the product, you will emit CO_2 that will lead to social costs ranging from £1.03 to £6.70 – the estimated social harm caused by the carbon emissions.

In KILOGRAM, the product's carbon emissions are accompanied by a reference value in liters of gasoline that, when burned, would cause an equivalent amount of carbon emissions.[15] Since individuals are likely much more familiar dealing with £ values than

[14] More than 81% of participants answered all control questions on the carbon measure correctly at the first attempt and only 3.17% needed three or more attempts to answer all control questions correctly.

[15] Although the imperial system is mostly used in the UK, the metric units are used for CO_2 and gasoline (e.g., GOV.UK, 2021*a*,*b*).

with kilograms of CO_2, we add a reference value in the latter treatment to make the CO_2 emissions in kilograms less abstract. We want to avoid measuring an effect solely because individuals cannot evaluate the magnitude of the CO_2 emission. Thereby we follow the literature investigating behavior and CO_2 emissions (e.g., Falk et al., 2021; Hummel and Hörisch, 2020; Pace and van der Weele, 2020). If carbon information in kilograms is used more widely, e.g., in supermarkets, people are likely to form their own reference values.

In ABATEMENT COSTS, participants receive information on the costs that would need to be invested to offset the product's emissions.[16]

In SOCIAL COSTS, social costs associated with the product's purchase are displayed as an interval. The estimate is calculated based on the frequently cited study by Nordhaus (2019)[17] and is lower than the very recent estimates by Rennert et al. (2022) published after our experiment was conducted. Therefore, our study provides a lower bound for the effect of social cost information. Contrary to the other two treatment conditions, we do not provide a single value in SOCIAL COSTS. We decided to present social costs as an interval to take into account that available social cost estimates vary in size and that there is uncertainty in the estimation. As a result, the difference between a single value and the uncertainty communicated to the participants by providing an interval is part of the estimated effect.

In Stage 3, participants decide whether to purchase the virtual product. In Stage 4, we ask them to guess the number of beer bottles that results in the same CO_2 emissions as the virtual product.[18] This guessing task allows us to compare participants' perceptions of the environmental impact of different types of CO_2 information. If the guess is less than 5% away from the actual value, subjects receive a bonus of £0.20. We compare participants' guesses with the life-cycle emissions of CO_2 equivalents estimated by Amienyo and Azapagic (2016) for beer production in the UK, which suggests that about 129 bottles yield the same level of CO_2 emissions as the virtual product.

In Stage 5, we provide participants with all three types of carbon information. We ask them which information they would prefer as consumers and which information they expect is least likely to lead to a product purchase. We use this information to explore which information is rated most positively and to study participants' beliefs on the impact of the carbon labels.

[16] We provide information on the amount that would have been donated to Compensators* if the product was not purchased. Due to the experimental setup, we have provided information on the actual abatement costs, i.e., the amount that would have been donated to Compensators* if the product was not purchased. This amount differs from the marginal abatement costs, i.e., the estimated marginal costs per unit of CO_2 emitted (for an overview, see, e.g., Huang et al., 2016).

[17] Particularly, carbon emissions of one ton in 2020 are associated with social costs between 43 and 278 USD in 2018, which we convert to 2020 £ using the average conversion rate in 2018 provided by Exchange Rates UK (2021) and inflate to 2020 based on the average yearly inflation rate by Bank of England (2021).

[18] Imai et al. (2022) shows that individuals make more precise guesses for CO_2 emissions of beer compared to other products. Participants in our experiment were instructed to consider the CO_2 emissions of a beer bottle from production up to waste disposal. The beer is produced in the UK, filled in a 0.33-liter glass bottle without further packaging, bought unchilled in the supermarket, and consumed at home.

In Stage 6, individuals are asked to fill out a short questionnaire in which we elicit (i) demographic characteristics, (ii) environmental awareness based on the Six Americas Super Short Survey (SASSY)[19] (Chryst et al., 2018), (iii) beliefs that climate change is human-caused (Howe et al., 2015), (iv) beliefs regarding the importance of personally taken actions to fight climate change, and (v) altruism and patience (Falk et al., 2018).

4 Hypotheses

This study focuses on the relative effectiveness of carbon information on pro-environmental decision-making. We compare the purchase decision of an emission-intensive virtual product if the product's carbon content is displayed in terms of kilogram, abatement cost, and social cost information.

Our hypothesis regarding kilogram and abatement cost information is based on the finding of Hummel and Hörisch (2020). The authors report that decision-making is impacted more by kilogram than by abatement cost information. Therefore, we expect that the purchase rate in ABATEMENT COSTS exceeds the one in KILOGRAM:

$$\textbf{H1: } purchase\ rate_{\text{ABATEMENT COSTS}} > purchase\ rate_{\text{KILOGRAM}}.$$

Regarding the effect of social cost information on decision-making, there is, to the best of our knowledge, no empirical evidence so far. Our approach is explorative, and we refrain from formulating further hypotheses.

5 Results

Presenting our results, we use the following abbreviations: Chi-squared test (χ^2), Chi-square goodness of fit (GOF), Kruskal-Wallis test (KW), two-tailed Mann-Whitney U test (MWU), and Robust Wald test (W). We summarize multiple p–values by $p's$.

5.1 Summary statistics

In the following, we report the results of our pre-registered analyses. Table 2 provides an overview of the control variables showing that randomization worked well: The three treatments are balanced with respect to demographics, environmental awareness, offsetting in the past, altruism, and patience ($\chi^2(2)$/KW, $p > 0.06$). Compared to the general British population, our sample is more female (69%) and younger (34 years) (share of women in the UK: 51%, median age: 40 years Office for National Statistics, 2022).

[19] The SASSY classifies individuals into the following six segments depending on their concerns regarding global warming: dismissive (least concerned), doubtful, disengaged, cautious, concerned, alarmed (most concerned).

Table 2: Descriptive statistic.

Variable	Mean/ median	KILOGRAM	ABATEMENT COSTS	SOCIAL COSTS	Test statistics (p-values)
		(N = 200)	(N = 200)	(N = 200)	
Demographics					
Female	Mean	0.73 (0.45)	0.67 (0.47)	0.67 (0.47)	2.06 (0.38)
Non-binary gender	Mean	0 (0)	0.01 (0.10)	0.03 (0.16)	5.49 (0.06)
Age	Mean	36.35 (13.34)	36.76 (12.95)	36.57 (14.10)	0.31 (0.86)
Has children	Mean	0.47 (0.50)	0.45 (0.50)	0.38 (0.49)	3.36 (0.19)
Religious	Mean	0.28 (0.45)	0.26 (0.44)	0.26 (0.44)	0.27 (0.87)
Highest education achieved	Median	Undergraduate degree (BA/BSc/other)			1.78 (0.41)
Income category	Median	£30,000 - £49,999			3.27 (0.19)
Job in science	Mean	0.045 (0.21)	0.07 (0.26)	0.05 (0.22)	1.35 (0.51)
Job in business	Mean	0.15 (0.36)	0.23 (0.42)	0.17 (0.38)	4.06 (0.13)
Political orientation: Left	Mean	0.55 (0.50)	0.53 (0.50)	0.51 (0.50)	0.66 (0.72)
Political orientation: Right	Mean	0.30 (0.46)	0.26 (0.44)	0.29 (0.45)	0.94 (0.62)
Political orientation: Green	Mean	0.11 (0.36)	0.12 (0.33)	0.12 (0.33)	0.29 (0.86)
Environmental awareness					
SASSY segment	Median	Cautious		Concerned	4.03 (0.13)
Global warming caused by humans	Mean	0.62 (0.49)	0.63 (0.49)	0.63 (0.48)	0.10 (0.95)
Actions matter to fight climate change	Mean	0.65 (0.48)	0.70 (0.46)	0.67 (0.47)	1.19 (0.55)
Often takes env. friendly action	Mean	0.64 (0.48)	0.61 (0.49)	0.64 (0.48)	0.61 (0.74)
Has offset in the past	Mean	0.15 (0.36)	0.12 (0.33)	0.17 (0.38)	2.02 (0.36)
(Strongly) agrees to emission trading	Mean	0.42 (0.49)	0.44 (0.50)	0.45 (0.50)	0.50 (0.78)
Altruism	Mean	0.03 (0.81)	-0.02 (0.79)	-0.02 (0.87)	1.05 (0.59)
Patience	Mean	-0.005 (0.96)	-0.005 (1.00)	0.01 (1.04)	0.31 (0.85)

Note: Standard deviation in parentheses. For binary variables, we use Chi-square tests; for ordinal and continuous variables, we use Kruskal-Wallis tests.

5.2 Purchasing behavior

On average, about one-third of individuals purchase the product in Stage 3. This rate is in line with Pace and van der Weele (2020), who find average purchasing rates of 35%. Figure 2 displays the share of individuals that purchase the carbon-intensive virtual product by treatment. Individuals purchase the product least often when information on CO_2 emissions is displayed in abatement costs (purchasing share: 29%). If information is displayed in kilograms or social costs, the purchasing share is slightly higher at about 35%. However, the purchasing rate is not significantly different across the three treatment conditions ($\chi^2(2) = 2.216$, $p = 0.330$, Cohen's $w = 0.059$).[20] Given our experimental design, we could have detected a minimum difference between two purchasing rates of about 12.6 percentage points, assuming a commonly used power of 0.8 and a significance level of 5%.

Running linear probability models including a large set of control variables does not alter this result. In particular, we regress the purchase decision *(=1 if purchased)* on the information received using kilogram as the reference category in specification (1) of Table 3. We include two dummy variables ABATEMENT COSTS and SOCIAL COSTS to estimate the difference in purchasing likelihood in comparison to receiving carbon information in kilograms. In specification (2), we add control variables for individual's demographic characteristics.[21] We further add controls for environmental awareness in

[20] Cohen's w is computed as $w = \sqrt{\sum_{i=1}^{m} \frac{(P_{1i} - P_{0i})^2}{P_{0i}}}$ where m denotes the number of cells in the contingency table, P_{1i} denotes the observed proportion in cell i under H_1 and P_{0i} describes the expected proportion under H_0 (Cohen, 2013).

[21] As described in the related literature, individuals act pro-environmentally only in a good state of the

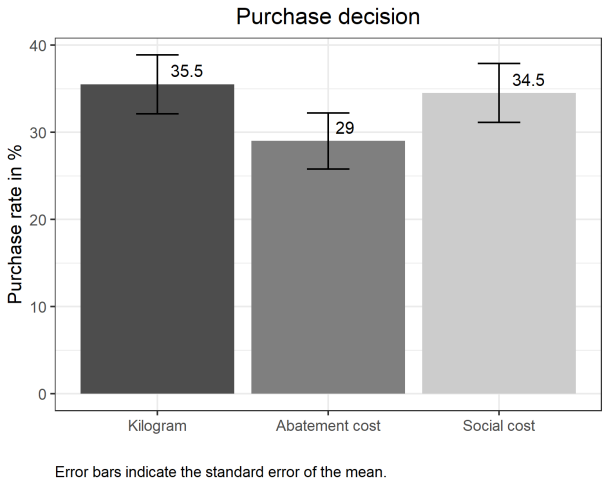


Figure 2: Purchase decision by treatment.

specification (3) and account for having offset emissions in the past in specification (4). In specification (5), we further control for individual's altruism and patience.[22]

Table 3: OLS regressions.

	Dependent variable: Purchase (0 or 1)				
	(1)	(2)	(3)	(4)	(5)
ABATEMENT COSTS	-0.065	-0.063	-0.050	-0.054	-0.060
	(0.047)	(0.047)	(0.044)	(0.044)	(0.043)
SOCIAL COSTS	-0.010	0.001	0.002	-0.003	-0.004
	(0.048)	(0.048)	(0.046)	(0.046)	(0.045)
Constant	0.355***	0.913**	0.802**	0.810**	0.794**
	(0.034)	(0.358)	(0.384)	(0.389)	(0.361)
Wald-test: ABATEMENT COSTS	1.141	1.313	0.980	1.115	1.225
= SOCIAL COSTS = 0 (p-value)	(0.320)	(0.269)	(0.376)	(0.329)	(0.295)
Demographics	No	Yes	Yes	Yes	Yes
Environmental awareness	No	No	Yes	Yes	Yes
Offset in the past	No	No	No	Yes	Yes
Altruism & patience	No	No	No	No	Yes
N	600	600	600	600	600
R2	0.004	0.109	0.222	0.230	0.255

Note: Robust standard errors in parentheses; The following control variables are included in Demographics: *gender, age, has children, religious, highest education level achieved, income category, job in science, job in business, political view* and in Environmental awareness: *SASSY segment, global warming caused by humans, (strongly) agrees to actions matter, often takes env. friendly action*; *** p<0.01, ** p<0.05, * p<0.1

Across all specifications, we find similar results. Abatement costs are associated with a lower purchasing probability, but the effect is not significantly different from zero ($p's > 0.165$). The estimates suggest that the purchasing likelihood is about five to six

economy (Brodback et al., 2020). Applying their findings more broadly to the state of an individual, we control for observable states (e.g., income, education level) and assume that unobserved states are equally distributed across treatments.

[22]We do not control for agreement to emission trading in our specifications as individuals might use disagreement as an excuse when buying the virtual product. The results are very similar if we account for agreeing to emission trading.

percentage points lower if an individual received information on abatement costs, which is a reduction of about 16 to 18 percent relative to the purchasing rate in KILOGRAM and SOCIAL COSTS, respectively. Thus, the effect seems to be economically relevant but lacks statistical significance.

Additionally, the estimate for SOCIAL COSTS is tiny, fluctuating around zero, indicating no significant difference to the purchasing likelihood in KILOGRAM ($p's > 0.834$). We therefore reject H1. Our result can be seen as a lower bound since the very recent estimate by Rennert et al. (2022) suggests even higher social costs.

The evidence might be due to multiple and opposite effects resulting from the display of social costs as an interval. Related uncertainty from the interval might, e.g., lead to higher purchasing rates due to motivated reasoning or reduced purchasing rates due to risk-averse individuals preferring the certain outcome.[23] Simultaneously, the social cost information itself might result in lower purchasing likelihood as they could be perceived as stressed societal impacts (see Section 2). To disentangle the effects, future research could display carbon information as both an average value and an interval. The two monetary types of information are also not jointly significant in the five specifications (W, $F(2)$, $p's > 0.269$). Excluding the four participants who failed the attention check in the questionnaire does not change the result. Logit regressions confirm these findings and can be found in Table A.2. We, therefore, summarize our first result based on the evidence of our experiment:

Result 1: *Carbon information in kilograms, abatement costs, and the interval of social cost estimates lead to similar purchasing decisions.*

Interestingly, this result differs from Hummel and Hörisch (2020), who find a significantly stronger effect of the kilogram information than information on abatement costs. Thus, we can neither confirm significant differences between the types of information nor do we find qualitatively the same effect. Various reasons for this divergence are possible, such as the different subject pools, the tasks participants are asked to complete, or the fact that the task in Hummel and Hörisch (2020) was not incentivized.

In Table A.3, we test for heterogeneous effects across gender (specification 1), political orientation (specifications 2-4), education (specification 5), income (specification 6), individual's altruism (specification 7), pro-environmental behavior (specification 8), and having a job in a science-related (specification 9) or business-related field (specification 10).[24]

We do not find significant heterogeneous effects for any subgroup in the specifica-

[23] In a comparable setup, Pace and van der Weele (2020) show that decreasing uncertainty of carbon information leads to more pro-environmental behavior. The broader literature on pro-social decisions under uncertainty finds contradictory effects (Butera and List, 2017). E.g., Kappes et al. (2018) find that uncertainty about the extent of harm caused to others increases pro-social behavior, whereas Exley (2016) documents that uncertainty decreases pro-social choice.

[24] Education is measured by whether the subject attained an A level or a higher level of education. Income is considered by including a dummy variable that is one if the individual reports earning more than £30,000 after taxes, the median household disposable income in the UK (Office for National Statistics, 2021). Pro-environmental behavior is considered by including a dummy variable that equals one if the participant (very) often engages pro-environmentally.

tions (1)-(8) (ABATEMENT COSTS: $p's > 0.314$, SOCIAL COSTS: $p's > 0.274$). Hence, we suspect that our findings are not limited to the specific sample studied but hold more generally, e.g., for representative samples. However, we can gain interesting insights by considering heterogeneity with respect to occupation. First, specification (9) indicates that kilogram information leads to significantly lower purchasing rates for individuals with a science-related job compared to non-science-related individuals ($p = 0.001$). Moreover, purchasing rates for these individuals are higher if the carbon emissions are displayed in terms of abatement costs or social costs ($p's < 0.035$). Even though this result is sensible as they might be more familiar with kilogram information, we do not overemphasize this finding as only 33 individuals in our study self-reported working in a scientific profession. Second, we cannot replicate Hummel and Hörisch (2020)'s evidence derived from an experiment with business students when we focus on individuals with a job in a business-related field (specification 10) ($p = 0.316$). As a result, future research could investigate how experts in sustainability assessments, such as experienced employees or investors, decide based on the respective carbon measures. We summarize:

Result 2: *Carbon information leads to similar purchasing likelihoods for almost all considered subgroups.*

Summarizing, information on abatement costs leads to about 6 percentage points lower purchasing rates than information on kilogram CO_2 or social costs. As the effect is, however, not statistically and qualitatively opposed to the effect found by Hummel and Hörisch (2020) we stress that future research is required to investigate potential reasons and the robustness of these results.

5.3 Beliefs and preferred information

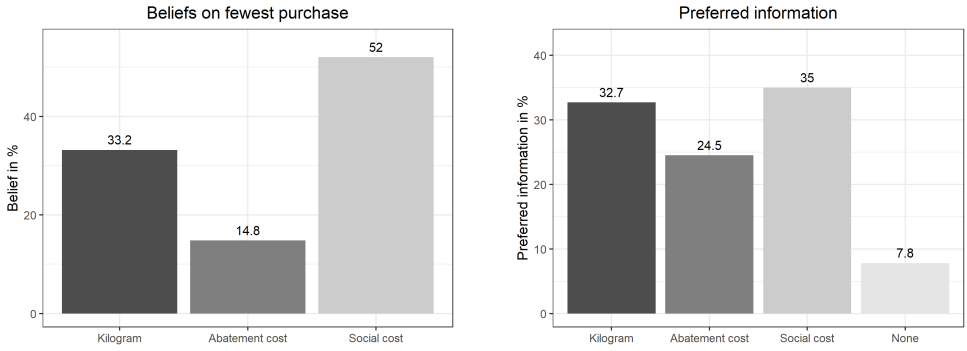
On an exploratory basis, we pre-registered to investigate which information individuals believe leads to the fewest purchases of the carbon-intensive product and which information individuals would prefer as consumers (Stage 5 in the experiment). Both questions were not incentivized. We ensured that individuals understood all three types of carbon information by including control questions.[25]

Figure 3a displays individuals' beliefs concerning the information that leads to the lowest purchasing rate of the emission-intensive virtual product. Interestingly, more than 50% of the individuals state that social cost information leads to the fewest purchases. The observed fractions deviate significantly from the hypothesized proportions that the beliefs are equally distributed over the three categories (GOF, $\chi^2(2) = 123.34$, $p = 0.001$). Importantly, these findings do not match with the results for the purchasing behavior in Section 5.2. Whereas we find that abatement costs tend to lead to the fewest purchases, individuals choose this type of information as most effective least often. In

[25] We report the pooled results as the treatment is not significantly related to individuals' beliefs ($\chi^2(4) = 6.970$, $p = 0.137$) and preferences ($\chi^2(6) = 8.658$, $p = 0.194$). Thus, individuals do not seem to be biased towards the information they received for their purchasing decision. For beliefs and preferences by treatment, see Figure A.1 and A.2.

contrast, individuals believe that social cost information leads to purchasing the product least often. To sum up, individuals hold incorrect beliefs as there seems to be a remarkable difference between the perceived effectiveness of information and actual purchasing behavior. We summarize the observation in our third result:

Result 3: *Individuals incorrectly believe that social cost information leads to the fewest purchase.*



(a) Beliefs **(b)** Preferences

Figure 3: Beliefs and preferred information.

Due to the increasing use of carbon labels, the question arises, which information on carbon emissions consumers prefer. As they might not prefer any of the three carbon labels, we provided the opportunity to choose *none* when asking for their preferences.

The responses are shown in Figure 3b. We can observe substantial variation across preferred labels, indicating that consumers have heterogeneous preferences for carbon labels. The stated preferences differ significantly from a hypothetical distribution where all three carbon labels are preferred with equal probability (GOF, $\chi^2(2) = 11.87$, $p = 0.003$). Moreover, participants report preferring the social cost information most frequently. Information on kilogram CO_2 is named second most often, and the abatement cost is preferred least. Thus, many participants prefer the social cost information, which most individuals also believe to induce pro-environmental behavior. We summarize:

Result 4: *Individuals have heterogeneous preferences for carbon labels, with social costs being the most preferred information.*

To summarize, individuals believe that social cost information leads to the fewest purchase of the carbon-intensive product, and many individuals would prefer this type of information as consumers. However, the beliefs differ from the observed purchasing behavior, i.e., abatement costs tend to lead to the fewest purchases, but many individuals do not believe it to be the most effective. Moreover, subjects rather prefer information displayed in kilograms or social costs over carbon measured in abatement costs.

5.4 Perception of environmental impact

As individuals tend to underestimate the emissions of products (e.g., Camilleri et al., 2019), it is questionable whether other types of information could correct their perception. To get a suggestive insight into how consumers perceive the respective carbon information, we ask them in Stage 4 to guess the equivalent number of beer bottles they could consume, which would lead to the same carbon emissions as purchasing the virtual product. The focus of this task thus does not lie in eliciting participants' knowledge on precise carbon emissions but in how their perceptions differ across treatment conditions. We pre-registered to investigate this measure on an exploratory basis. The cumulative distribution functions of the guesses are shown in Figure 4.

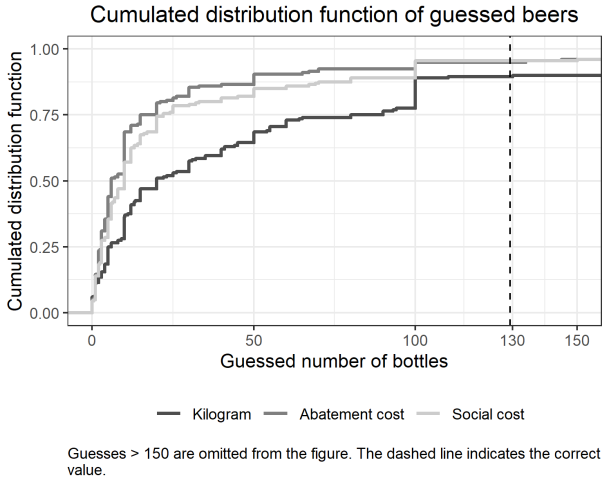


Figure 4: Guessing task.

Most guesses are far from the correct value of about 129, and only two individuals received the bonus. Interestingly, most individuals underestimate the equivalent number of bottles they could consume.[26] Since previous studies have found that individuals underestimate the CO_2 emissions associated with a given product (e.g., Camilleri et al., 2019), we were expecting that individuals overestimate the number of bottles they could consume instead of purchasing the virtual product. However, we find that individuals underestimate the potential consumption for a given level of CO_2 emissions. Thus, it seems that individuals have, in general, difficulties understanding the relative magnitude of CO_2 emissions.

This finding is well in line with previous research showing that individuals have difficulties assessing CO_2 emissions (e.g., Capstick et al., 2019; Grinstein et al., 2018; Kaklamanou et al., 2015) and translating greenhouse gas emissions from one action to another (see the survey by Wynes et al., 2020). However, this might change in the future if carbon labels are more widely applied and people form their own reference values based on their typical consumption. Whether the underestimation occurs due to

[26] We pre-registered to investigate the probability of overestimating the equivalent number of bottles. Given the results, we focus on the guesses directly; see Table A.4 for the regression results focusing on the probability to overestimate.

individuals' beliefs that the virtual product's CO_2 emissions are low or whether beer was perceived to have high emissions is not distinguishable. Based on our experimental findings, we summarize:

Result 5: *Individuals underestimate the potential consumption for a certain level of CO_2 emissions if CO_2 emissions are displayed in kilograms, abatement costs, or social costs.*

Individuals in KILOGRAM make significantly higher guesses than in the ABATEMENT COSTS (MWU, $z = -5.911$, $p < 0.001$) and SOCIAL COSTS treatment (MWU, $z = -4.411$, $p < 0.001$) and are thus closer to the correct value. The higher guesses in KILOGRAM indicate that participants associate a higher environmental impact with this type of information. As the probability of overestimating the emissions of a beer as well as its magnitude, should be equally distributed over the treatments, we can compare and interpret the differences between the treatments causally. However, we cannot differentiate between potential divers of the effect, such as higher familiarity with kilogram information or an effect of the reference value (liters of gasoline). Moreover, the anchoring effect (Tversky and Kahneman, 1974) may play a role as individuals see the magnitude of the carbon measure before making their guess, which is higher for KILOGRAM than in the ABATEMENT COSTS and SOCIAL COSTS treatment. We cannot present a distinct explanation for the pattern observed, but all provided explanations are probably part of the overall observed effect. We summarize:

Result 6: *Carbon information in kilograms leads to higher perceived environmental harm than abatement or social cost information.*

We find no statistically significant difference between the underlying distributions of the guesses of subjects who buy the product and those who do not (MWU, $z = -0.032$, $p = 0.975$). In addition, we test the correlation between guess and purchase decision with a linear probability model in Table A.5. In column 1, a significant ($p = 0.001$) but very small coefficient of $-9.56 * 10^{-6}$ is estimated, which suggests that higher guesses are associated with a slightly lower purchasing likelihood. We test the robustness of this result with a logit regression and find that the correlation is insignificant ($p = 0.260$) (column 2). Similarly, the effect remains small and is just marginally significant ($p = 0.056$) (insignificant ($p = 0.123$)) in the linear probability model (logit model) when we exclude extreme guesses over 1000 (columns 3 and 4, respectively). We therefore conclude:

Result 7: *The perceived environmental harm does not explain different purchasing rates.*

To subsume, individuals seem to have difficulties associating actions with a carbon measure, no matter how it is displayed. Therefore, information campaigns that inform on carbon emissions associated with actions are probably a necessary step towards knowledge of carbon emissions in daily life.

6 Discussion and conclusion

Motivated by the increasing use of monetized information about environmental effects in the corporate world, we aim to empirically investigate how individuals decide when facing this type of information. Due to an extensive debate on monetization but the lack of empirical evidence, we contribute to the existing literature by systematically investigating the effects of three types of carbon information. We hereto conduct an online experiment in which individuals decide whether to purchase a virtual product whose carbon emissions are randomly displayed in one of three forms: kilograms, abatement costs, or social costs. Differentiating between monetization based on abatement cost and social cost is a novel approach but might play a major role in how the information is perceived. Furthermore, to the best of our knowledge, we are the first to incentivize subjects' actions in the context of providing monetized carbon information. We increase the external validity of our findings by considering a heterogeneous sample that we recruit via Prolific.

Our findings contrast the only study we are aware of by Hummel and Hörisch (2020) that reports a significantly larger effectiveness of kilogram information compared to information on abatement costs. Contrasting their evidence, we find that the specific unit in which carbon emissions are displayed does not significantly affect individuals' purchasing decisions. Moreover, our finding even qualitatively contrasts Hummel and Hörisch (2020) as we find that information on abatement costs tends to have a higher effectiveness than carbon information in kilograms CO_2. Therefore, we stress that future research is required to investigate the robustness of the results and understand potential drivers, such as specific metrics and design choices.

The various forms of carbon information have similar effects among almost all considered subgroups of individuals. Interestingly, however, individuals believe that social cost information is most effective in deterring the purchase of the product. Moreover, many subjects would prefer this type of monetized information as consumers. Aiming to get a suggestive insight into individuals' perception of the environmental impact regarding the respective carbon measures, we find that individuals seem to have difficulties transferring a carbon measure to an associated action, irrespective of the type of carbon display.

Empirical evidence on behavioral differences to carbon information is relevant in multiple contexts. The context most closely linked to our experimental setup is eco-labels on products to induce pro-environmental choices by consumers. For example, the increasing use of carbon labels might be regulated in the future to enable comparisons between products (similar to nutrition information). Moreover, our results can inform the growing field of corporate reporting on social impacts, standard setters establishing general frameworks for carbon metrics, and be applied to internal management decisions (e.g., investment decisions). Based on our experiment in the consumer context, we would suggest that all studied types of carbon information induce similar behavior, but most individuals seem to prefer social cost information, and only a very small minor-

ity opposes carbon information. Since our experiment confirms previous evidence that individuals generally have difficulties understanding carbon information, we would expect that a consistent display across various decision contexts and choices would make it easier for decision-makers to learn about carbon emissions of products and related actions.

Our results originate from a stylized online experiment. Investigating the robustness of the findings in the respective contexts, in particular, by conducting field experiments with actual products, presenting social costs as a single value, using a representative sample of consumers, employees familiar with sustainability reporting, managers, or investors, would be a crucial step before implementing the respective carbon displays.

Overall, our study builds a stepping stone for further research on the optimal display of environmental information. I.e., future research could investigate which channels play a role when information is monetized: heuristics based on £ compared to kilogram information, mental accounting (Thaler, 1985) and motivated reasoning (e.g., review by Epley and Gilovich, 2016). Furthermore, future research could explore potential benefits of monetizing impacts that go beyond the scope of this experiment, which focuses on carbon emissions. For example, monetization could lead to more sustainable decisions if multiple environmental impacts (e.g., biodiversity and water usage) or social impacts (e.g., health and safety) are monetized and thus more easily comparable. Relatedly, the type of information might interact with additional information on products, such as their nutritional value, organic production, or price. Additionally, generating the relevant information might itself result in more sustainable performance of firms and products.[27]

Concluding, we see that companies' increasing interest in reporting on their social impact in terms of monetized values is mirrored in consumers' preferences for that type of information. However, we do not find supporting evidence for its success in inducing pro-environmental behavior. Hence, monetized information does not seem to lead to higher sustainability, but it also does no harm.

[27] Similarly, mandatory reporting seems to improve the environmental performance of firms (e.g., Downar et al., 2021; Jouvenot and Krueger, 2019; Tomar, 2023) and voluntarily eco-labels on products seems to lead to more sustainable products (e.g., Delmas and Grant, 2014; Egan and Waide, 2005).

Chapter III

Be green or feel green? An experiment on moral balancing in pro-environmental decision making[28]

1 Motivation

Many environmental problems such as global warming, air pollution, water shortage, or the loss of biodiversity have in common that they are primarily caused by human behavior (e.g., Steg and Vlek, 2009; Vlek and Steg, 2007). For example, around 70% of global greenhouse gas emissions are linked to household consumption (Hertwich and Peters, 2009). Despite technical innovations that boost sustainability, such as electric cars or energy-efficient buildings, behavioral change remains crucial (Steg and Vlek, 2009). Since innovations only contribute to sustainability if their positive impact is not overtaken by an increase in consumption (Gillingham et al., 2013).

With the increasing severity of environmental problems, people frequently face encouragement from governments, organizations, or peers to act more environmentally friendly. Policymakers introduce behavioral change initiatives to induce climate-friendly behavior using nudges, economic incentives, or information campaigns (Clot et al., 2022). For example, the UK government's 25-Year Environment Plan proposes "scoping out an evidence-based behaviour change strategy to enable further actions by individuals, communities, businesses, and government" (HM Government, 2018). From a government perspective, encouraging behavior change is attractive as, unlike regulations, it mostly does not lead to public backlash or diminishing votes (Whitmarsh and O'Neill, 2010).

Many environmental campaigns emphasize that "every change counts" and aim to motivate individuals to make small and painless behavioral changes (such as avoiding plastic straws, double-sided printing, or switching off the lights when leaving a room). The designers of these campaigns often hope small changes will result in higher-impact behavioral changes later on (Thøgersen and Crompton, 2009) in line with the foot-in-the-door effect (Freedman

[28] This chapter is based on Schöller and Schlereth (2023).

and Fraser, 1966). For example, the UK's Sustainable Consumption Round Table recommends "to drop new tangible solutions into people's daily lives, catalysts that will send ripples, get them talking, sweep them up into a new set of social norms, and open up the possibility of wider changes in outlook and behavior." (Sustainable Consumption Roundtable, 2006).

However, to evaluate the effectiveness and judge the overall welfare effect of an environmental campaign, one has to go beyond the immediate effect on the targeted behavior and also consider the campaign's indirect impact on future environmental behavior (Gilg et al., 2005; Grieder et al., 2021). Against this backdrop, we raise the following questions: What are the dynamic effects of pro-environmental behavior? Does promoting (small) behavioral changes affect future decision-making, and if yes, does it induce more or less pro-environmental behavior? In this paper, we analyze the dynamic effects of pro-environmental behavior and focus on whether the magnitude of the initial actions has a systematic effect on pro-environmental behavior later on. We provide insights on whether balancing occurs only for environmental actions with relevant consequences or already occurs from an individual's impression of having done something for the environment, even if it is not important on a grander scale.

According to the literature, pro-social decisions are not made in a vacuum but are affected by previous decisions (for an overview, see, e.g., Blanken et al., 2015; Kuper and Bott, 2019; Maki et al., 2019; Simbrunner and Schlegelmilch, 2017). In this paper, we focus on moral balancing (Nisan and Horenczyk, 1990).[29] Moral balancing is a cognitive bias that can go both ways; it is defined as moral licensing (cleansing) if someone is acting immorally (morally) after a moral (an immoral) action (for a theoretical consideration, see, e.g., Merritt et al., 2010; Mullen and Monin, 2016). Bénabou and Tirole (2011) introduce a theoretical model that explains the cognitive bias as balancing one's moral self-image, which positively impacts an individual's utility (Akerlof and Kranton, 2000, 2005; West and Zhong, 2015), and the cost of acting pro-social. Morally or immorally perceived actions are performed to close the gap between desired and perceived moral self-image (Mazar and Zhong, 2010). According to the concept of moral balancing, individuals are more likely to engage in ethical conduct when they feel a previous unethical decision threatens their moral self-image. In contrast, they are less likely to engage in moral behavior when they previously secured their moral self-image by a moral action (Ploner and Regner, 2013). A related theory explains intertemporally dependent altruistic behavior with a fixed 'altruistic budget' that determines individuals' altruistic acts over time (Gee and Meer, 2019). However, the literature is inconclusive whether the altruistic budget is fixed of flexible (Gee and Meer, 2019).

As individuals consider various domains to define their self-image, one not only observes moral balancing in the same behavioral domain (same-domain moral balancing) but also between seemingly unrelated behavior (cross-domain moral balancing) (Mullen and Monin, 2016). Moral balancing has been identified in various domains, suggesting that it can occur in any domain with positive normative connotation, e.g., pro-environmental behavior (Effron, 2016).

[29] The literature uses several terms rather interchangeably such as negative spillover effects (Engel and Szech, 2020), compensatory beliefs (West and Zhong, 2015), moral credentials (Monin and Miller, 2001), moral self-regulation (Sachdeva et al., 2009), or conscience accounting (Gneezy et al., 2014).

Besides moral balancing that predicts negative spillover effects on later environmental behavior, there is also evidence for positive spillover effects (e.g., Baca-Motes et al., 2013). Positive spillovers can be explained by the desire to act and to be perceived as consistent, adhering to cognitive dissonance theory (Festinger, 1957). The mixed evidence on whether positive or negative spillovers prevail and the limited number of work in the field of pro-environmental behavior stresses the importance of further analyses of the dynamic effects of moral behavior.

In an online experiment, we explore the following research questions. Do we find moral balancing in pro-environmental behavior, in particular carbon offsetting? Does moral balancing occur even if the initial moral or immoral act has close to no effect or only if the moral or immoral action has a substantial impact? Moreover, is the moral balancing effect moderated by an individual's moral values?

To test how pro-environmental decisions depend on past environmental behavior, we conduct an experiment that consists of two parts. In Part 1, participants perform a real-effort task. The treatments vary if and how much carbon offset participants receive for succeeding in the task. Thereby, we exogenously vary if a moral license is acquired and of which magnitude. In Part 2, we measure moral balancing by letting participants decide how much to donate for carbon offsetting. The treatments allow us to test whether pro-environmental behavior that has a negligible impact on the environment leads to a similar licensing effect than behavior that has a more pronounced impact on the environment.

We find that participants who successfully solve the real effort task subsequently donate, on average, less to carbon offset than those who failed. This difference is significant on the 10%-level and indicates that moral balancing influences pro-environmental decision-making, in particular in the domain of carbon offsetting. Regarding the magnitude of the initial moral act, we do not see systematic differences in moral balancing. Furthermore, our results indicate that moral balancing depends on individuals' moral values. We do not find moral balancing for participants with the greatest environmental concerns. Participants with somewhat lower concerns regarding global warming engage in moral balancing, but only if the initial act is substantial.

Most studies on environmental behavior only examine single actions, neglecting temporal context. We add to the literature by analyzing dynamic aspects of environmental decision-making. We consider that individuals regulate their moral choices, resulting in prior actions affecting subsequent decisions. Given the mixed evidence on the occurrence and determinants of moral balancing, the current experimental study aims to extend the literature on moral balancing and pro-environmental behavior in several ways. First, we analyze moral balancing in the domain of voluntary carbon offsetting, a growing market due to the increasing attention on carbon dioxide as the primary driver of climate change. In 2019, around $320 million worth of carbon offsets were purchased globally, reducing approximately 104 million metric tons of CO_2 emissions (Forest Trends' Ecosystem Marketplace, 2020). Second, we exogenously vary the magnitude of the initiating pro-environmental action to investigate whether moral balancing occurs only for substantial or also for neglectable moral actions. Moreover, we measure the initial action as well as the moral balancing effect in carbon offsets. Thereby, we can com-

pare both decisions on a joint and quantifiable scale. Third, we incentivize individuals' actions so their decisions have real-world consequences. Exciting work often relies on self-reported behavior (e.g., Chatelain et al., 2018; Lanzini and Thøgersen, 2014) or elicits intention to act (e.g., Jordan et al., 2011; Margetts and Kashima, 2017). However, self-reported behavior lacks reliability and might be biased in approximating actual decision-making, especially concerning moral behavior. Eliciting intention to act instead of actual behavior can lead to biased results, e.g., the meta-analyses by Maki et al. (2019) reports that pro-environmental behavior results in positive spillovers on intentions, whereas negative spillovers are observed for actual behavior. Fourth, we add to the research on moderators of moral balancing and study particularly the impact of environmental values.

The paper is structured as follows. Related literature is discussed in Section 2. Section 3 presents the experimental design and discusses the data collection. Hypotheses and results are presented in Section 4 and Section 5, respectively. We conclude and discuss in Section 6.

2 Related literature

From charitable giving (e.g., Brañas-Garza et al., 2013; Grieder et al., 2021) to food consumption habits (e.g., Wilcox et al., 2009), racial prejudice (e.g., Effron et al., 2009), or sexism (e.g., Monin and Miller, 2001) moral balancing is found in many domains (for meta-analyses see Blanken et al. (2015); Kuper and Bott (2019); Maki et al. (2019); Simbrunner and Schlegelmilch (2017)) as well as across domains (Hofmann et al., 2014).

Meta-analyses estimate moral licensing to be of small to medium magnitude (Blanken et al., 2015; Kuper and Bott, 2019; Simbrunner and Schlegelmilch, 2017). Regarding the welfare effects of moral licensing, Grieder et al. (2021) find moral licensing to decrease future charitable giving but that multiple opportunities to behave pro-socially positively impact aggregate donations.

Regarding our research question, we focus in the following brief literature overview on (1) moral balancing in the environmental domain and (2) moderators of moral balancing.

2.1 Moral balancing in the environmental domain

In the following section, we focus on studies in which the initial action and the moral balancing effect both lie in the environmental domain. Since the literature predominantly studies cross-domain moral balancing, we add evidence on less studied same-domain moral balancing. Compared to cross-domain, same-domain moral balancing has shown to be more likely and of greater magnitude (Dolan and Galizzi, 2015; Blanken et al., 2015). Besides same-domain moral balancing, pro-environmental behavior has been shown to affect behavior in domains such as pro-social decision-making (e.g., Engel and Szech, 2020; Hahnel et al., 2015; Mazar and Zhong, 2010). Similarly, also good deeds in other domains can affect sustainable behavior (e.g., Meijers et al., 2015; Sachdeva et al., 2009).

Moral balancing in environmental behavior has been reported regarding pro-environmental actions, intentions to act environmentally friendly (e.g., Burger et al., 2022; Geng et al., 2016),

and supporting climate-friendly policies (e.g., Noblet and McCoy, 2018). Maki et al. (2019) combine multiple studies in their meta-analysis and find a slightly negative moral balancing effect of pro-environmental behavior on future environmental actions and policy support. In contrast, they find positive spillover effects on intentions. The following overview focuses on moral balancing in actual behavior, as this is our study's primary objective.

Results from correlational analyses are mixed. A three-wave panel study with Danish consumers finds that some environmentally friendly behavior is related to more, others to less pro-environmental actions in subsequent years (Thøgersen and Ölander, 2003). Whereas correlational findings do not allow clean identification of moral balancing effects, our experimental design exogenously implements a pro-environmental action.

Several lab experiments confirm moral balancing regarding various forms of environmentally friendly behavior. For example, participants who performed a climate-friendly behavior by filling out a green instead of a conventional shopping list later conserved less water (Geng et al., 2016). Also, exposure to a green advertisement increases water consumption and lowers the intention to choose transportation with a low carbon footprint (Zhang et al., 2021). Randomly giving green-committed individuals positive feedback on the environmental friendliness of their shopping decisions reduced their recycling engagement compared to negative feedback. Receiving no feedback leads to a mid-range recycling rate (Longoni et al., 2014). A drawback of many lab experiments (e.g., Clot et al., 2013, 2016) is that, in particular, the initial action designed to induce moral balancing is often only imaginary and not performed, thereby limiting the informative value for real-world behavior.

Besides lab experiments, Tiefenbeck et al. (2013) report in a field setting that individuals participating in an environmental campaign to save water reduce their water consumption but consume more electricity. In contrast, Carlsson et al. (2021) find a reduction in electricity use after an information campaign that targets water consumption for individuals that had an efficient use of water prior to the intervention.

All mentioned studies test moral balancing from one pro-environmental decision context to an unrelated one, whereas we quantify the moral balancing effect and study the moral balancing of carbon offsetting on subsequent carbon offsetting decisions.

Besides studies documenting moral balancing, there is also empirical evidence finding no interdependences in pro-environmental decision-making (e.g., Liebe et al., 2021). Other studies report that one pro-environmental behavior increases the probability of further environmentally friendly behavior (e.g., Clot et al., 2016; Lanzini and Thøgersen, 2014; Margetts and Kashima, 2017; Panzone et al., 2021; Sintov et al., 2019). Panzone et al. (2021) is closest to our research design. They ask participants to recall past eco-friendly behavior and, similar to our research design, inform and congratulate them on the resulting amount of carbon savings. They find that the participants informed of their carbon savings purchased a food basket with a lower carbon footprint in an experimental online supermarket. Several replications of well-cited publications find null effects (e.g., Blanken et al., 2014; Urban et al., 2019), questioning the role moral balancing plays in decision-making.

Interestingly, not only one's own behavior but also an employer's good deeds can influence

participants' behavior (e.g., Grieder et al., 2020; List and Momeni, 2021). Grieder et al. (2020) find that informing subjects on their employer's donation to an environmental charity increases subjects' donations for environmental conservation.[30]

We conclude that the literature on moral balancing in environmental behavior is inconclusive and does not precisely predict whether moral balancing occurs in repeated carbon-offsetting decision-making.

2.2 Moderators of moral balancing

The inconclusive results regarding moral balancing (see Section 2.1) suggest that the occurrence might be sensitive to experimental conditions as well as individual attitudes (e.g., Alt and Gallier, 2022; Blanken et al., 2015; Mullen and Monin, 2016). There is a growing literature on factors moderating moral balancing, e.g., the cost of the initial action (Gneezy et al., 2012), the time between both decisions (Schmitz, 2019), feeling responsible for one's behavior (Engel and Szech, 2020), or the similarity of both tasks (Chatelain et al., 2018; Maki et al., 2019; Truelove et al., 2014). For a literature overview, see Blanken et al. (2015); Mullen and Monin (2016). In the following, we focus on (i) the role of the magnitude of the initial action and (ii) environmental values on moral balancing.

To the best of our knowledge, only two studies vary the magnitude of the initial actions. In Gholamzadehmir et al. (2019), participants were reminded of past frequent or infrequent pro-environmental actions. Recalling past (in)frequent actions leads to moral licensing (cleansing), and participants were less (more) likely to seek information about calculating their carbon footprint. Grieder et al. (2020) find that an employer's donation to an environmental charity affects participants' donations, but the magnitude of the donation rate (10% vs. 40%) does not matter. These divergent findings regarding the magnitude emphasize the importance of additional research.

The current literature makes it difficult to form a clear conclusion regarding the moderating effect of environmental values on moral balancing since various terms are used rather interchangeably (such as environmental consciousness (Garvey and Bolton, 2017), environmental self-identity (van der Werff et al., 2014), environmental attitudes (Lacasse, 2016), ecological motivation (Hahnel et al., 2015), or pro-environmental values (Thøgersen and Crompton, 2009)). Moreover, there is no standardized way to measure environmental values, but each paper uses different definitions and questionnaires. Therefore, the following overview covers environmental values more broadly and, in particular, includes findings regarding environmental self-identity as it gets the most attention in the current literature.

In general, the literature suggests that building an identity increases charitable giving (e.g., Charness and Holder, 2019; Kessler and Milkman, 2018). Whereas Bénabou and Tirole (2011) predicts that a threat to identity induces moral behavior.

Regarding moral balancing in the environmental domain, most papers argue that when climate-friendly actions are performed out of extrinsic (e.g., financial incentives, regulation)

[30] In contrast, List and Momeni (2021) find companies' Corporate Social Responsibility activities to increase employees' misbehavior.

instead of intrinsic motivation (e.g., self-identity, concerns, values), it is more likely to generate moral balancing (e.g., Clot et al., 2016, 2022; Lacasse, 2016; Miller and Effron, 2010; Nilsson et al., 2017; Noblet and McCoy, 2018; Thøgersen and Ölander, 2003; Thøgersen and Crompton, 2009; Truelove et al., 2014; van der Werff et al., 2014). Individuals who self-identify as environmentally friendly will engage less in moral balancing and are more likely to perform a subsequent environmental action than individuals with lower environmental self-identity (Garvey and Bolton, 2017; Geng et al., 2016; Truelove et al., 2014; van der Werff et al., 2014). These experimental findings align with the self-perception theory by Bem (1967) that predicts consistent behavior in domains where an individual has integrated past actions in their self-image (Lalot et al., 2022).

In contrast, few papers find environmental identity or attitudes not to mediate dynamics in environmental behavior (e.g., Gholamzadehmir et al., 2019; Gleue et al., 2022). Moreover, Hahnel et al. (2015) find moral cleansing only for individuals high in ecological motivation.

Little research exists regarding environmental concern as a moderator of moral balancing (Truelove et al., 2014). Truelove et al. (2016) find no general moderating effect of environmental concern on moral balancing. Whereas other papers find that individuals with severe concerns regarding climate protection (Burger et al., 2022) or climate change (Hartmann et al., 2023) engage to a larger degree in moral balancing.

We conclude that most papers find high environmental values to decrease moral balancing. However, there is little research regarding environmental concerns.

3 Experimental design

To test our research question, we set up an economic online experiment with oTree (Chen et al., 2016). The experiment was conducted on Prolific (www.prolific.co) between the 14th and 21st of June 2021. We obtained informed consent from all participants, and the study follows the relevant guidelines and regulations. Before data collection, the experiment was preregistered on AsPredicted.org (Nr. 70521, see https://aspredicted.org/JTY_GDD). Participation is only restricted to British individuals (country of birth) that live in the United Kingdom (country of residence) to ensure that subjects are equally familiar with the measurement units and understand the English instructions well. Participants are randomly assigned to one of three treatment conditions, each having 300 observations. The median completion time for the study was 10 minutes, with average earnings of £1.70 ($2.40).[31]

The experiment design adheres to the standard two-part structure commonly seen in moral balancing studies. The first part is designed to impact an individual's self-image and initiate a moral balancing effect, which is then measured in the second task.[32]. The two parts are followed by a questionnaire. The complete introductions can be found in Appendix B.2.

In Part 1, subjects can work for two minutes on a real effort slider task (Gill and Prowse, 2012). A slider is solved when the participant sets it to a given value using the computer mouse.

[31] Exchange rate £ to €: 1.4109 on the 16th of June 2021 (XE.com Inc., 2021).

[32] Some studies deviated from the two parts design, e.g., Brañas-Garza et al. (2013) uses multiple periods to identify a dynamic pattern of moral balancing.

Participants had to solve at least 26 sliders to succeed in the task. This threshold was determined in a pilot study as it resulted in a roughly equal number of successful and unsuccessful participants. Depending on the treatment condition, succeeding in the task leads either to a carbon offset of 10 kg (LOW), 100 kg (HIGH), or a payment of £0.20 to the participant (SELF). We chose the payment in SELF to be similar to the donation required to offset 10 kg of CO_2 to make it comparable to LOW. The slider task has several advantages compared to other real-effort tasks. It is easy to understand, can be conducted online, does not require prior knowledge from participants, and performance is not improved through guessing (Gill and Prowse, 2012). Additionally, the performance in the slider task is relatively insensitive to the size of incentives, as demonstrated in a between-subject design study by Araujo et al. (2016). As a result, the number of sliders that must be solved to complete the task successfully can be kept constant across all treatment conditions, making it an ideal choice for our study.

In Part 2, subjects play a dictator game (Kahneman et al., 1986). They receive £2 and split it between themselves and carbon offset, with the current market price of a selected carbon-offsetting charity as the exchange rate. Next, we elicit incentivized beliefs by asking participants to make two guesses: regarding the success rate of other participants in the real-effort task in Part 1 and the average offset of other participants in Part 2. For every guess that is within 5% of the actual value, a participant receives a bonus payment of £0.20.

Participants are informed that the study consists of two parts, but they only receive details about each part as they proceed. Previous research has shown that when individuals are made aware of the possibility of donating to a charity in the future, they behave less ethically in the present (Cojoc and Stoian, 2014). Moreover, the instructions state that decisions made in one part of the study will not impact the other part. At the end of the study, one part is randomly selected for payment. Restricting payment to one part is essential to rule out wealth effects that would otherwise bias our measurement of moral balancing (Azrieli et al., 2018; Charness et al., 2016).

To ensure that subjects understand the instructions, we include multiple control questions in both parts and an attention check in the questionnaire. Carbon emissions are paired with the corresponding distance in miles traveled with a typical new car that would result in an equivalent amount of carbon emissions. By providing a reference value, we aim to assist participants in evaluating the extent of the carbon offsets. Thereby, we align with the existing research on behavior and emissions (e.g., Falk et al., 2021; Imai et al., 2022; Pace and van der Weele, 2020).

It is essential that participants believe the instructions, e.g., that the carbon offset is implemented and that they are not subjected to any form of misguidance. We use two approaches to ensure that the carbon offset implementation is credible to participants. Firstly, we emphasize that the study does not involve deception, specifically that their decisions have real-world consequences, and the carbon offsets are implemented as stated. Additionally, we include a control question to verify participants' understanding of this aspect. Secondly, the participants are informed that they will receive a private message through Prolific that comprises the computation of the total carbon offset amount and an official donation receipt.

The post-experimental questionnaire elicits (1) demographic characteristics; (2) altruism and patience (Falk et al., 2018); (3) climate change concerns based on the Six Americas Super Short Survey (SASSY)[33] (Chryst et al., 2018); (4) opinion on whether climate change is human-caused (Howe et al., 2015); (5) beliefs regarding the importance of own actions to fight climate change; (6) environmental behavior; (7) previous offsetting and opinion on the effectiveness of offsetting. To determine if participants accurately keep track of the offset they accumulate throughout the experiment, we ask them to report the offset they received in both stages combined. Additionally, to control for the perceived difficulty of the task, we ask participants about the effort they exerted in Part 1.

4 Hypotheses

First, we hypothesize that moral balancing is prevalent in the decision to offset carbon emissions. In our experimental setup, we expect that succeeding in the slider task and acquiring carbon offset will lead to a lower donation for carbon offset than failing to solve the slider task. In line with the theory of moral behavior (Bénabou and Tirole, 2011), we assume that acquiring carbon offset boosts moral self-image while missing the chance to offset leads to a decline in moral self-image. If moral licensing (cleansing) is prevalent, the higher (lower) moral self-image will lead to less (more) moral behavior in the subsequent offsetting decision. More precisely, we expect a lower donation rate in the LOW and HIGH treatment for individuals who succeeded in the slider task than those who failed. In the SELF treatment, no offset is acquired. Hence we assume the moral self-image to stay constant, and we expect similar offsetting rates in Part 2 independent of the outcome of the slider task.

Our first set of hypotheses therefore states:

$$Donation_{\text{LOW}}^{Success} < Donation_{\text{LOW}}^{Failure},$$

$$Donation_{\text{HIGH}}^{Success} < Donation_{\text{HIGH}}^{Failure},$$

and accordingly

$$Donation_{\text{SELF}}^{Success} = Donation_{\text{SELF}}^{Failure}.$$

Second, we test whether moral balancing depends on the initial actions' magnitude. The study exogenously varies the magnitude of carbon offset that can be acquired by solving the slider task in Part 1. In LOW, participants can offset 10 kg of CO_2, whereas in HIGH, participants can work towards offsetting 100 kg. Our primary focus lies on the gap in offsetting rates in Part 2 between participants who successfully completed the slider task in Part 1 and those who did not. We compare this gap between LOW and HIGH. If the gap is similar in LOW and HIGH, we conclude moral balancing to be independent of the magnitude of the initial action. If

[33] Based on their level of concern regarding global warming, the SASSY categorizes individuals into six distinct segments: dismissive (least concerned), doubtful, disengaged, cautious, concerned, alarmed (most concerned)

we observe a significantly larger gap in HIGH than in LOW, we infer that the magnitude of the initial action matters for the size of the moral balancing effect. We expect that the initial action does not significantly impact the size of the subsequent donation. If this assumption holds, we would observe a similar rate of moral balancing in the LOW and HIGH treatment.

$$Donation_{\text{LOW}}^{Failure} - Donation_{\text{LOW}}^{Success} = Donation_{\text{HIGH}}^{Failure} - Donation_{\text{HIGH}}^{Success}.$$

In addition, we exploratively investigate potential moderators of moral balancing. In particular, we investigate the moderating effect of environmental values (for an overview of empirical and theoretical evidence regarding environmental values as a moderator for moral balancing, see Section 2.2).

5 Results

We present our findings using the following abbreviations: Chi-squared test (χ^2), Kruskal-Wallis test (KW), two-tailed Mann-Whitney U test (MWU), two-sample Kolmogorov-Smirnov (KS). We summarize multiple p–values by $p's$.

5.1 Summary statistics

We recruited 900 participants on Prolific, each sorted into one of the three treatments by arrival time. Table 4 confirms that the randomization into treatments was successful. The participants in all three treatments are generally comparable concerning demographics, altruism, patience, and environmental awareness ($\chi^2(2)$/KW, $p > 0.095$). A significant difference between the three treatments is only found regarding *Actions Matter*[34] ($\chi^2(2) = 9.891$, $p = 0.007$). When comparing effects between treatments, we account for this difference by adding a specification that controls for the variable *Actions Matter* in our regression analyses (hereafter denoted by Control A). Since participants filled out the questionnaire as a last step of the experiment, the previous tasks may have influenced responses, particularly regarding environmental awareness.

The success rate in the slider task (SELF: 0.500; LOW: 0.510; HIGH: 0.487; $\chi^2(2) = 0.329$, $p = 0.848$) and the number of solved sliders (SELF: 26.230, LOW: 26.413, HIGH: 25.253; KW, $\chi^2(2)$ with ties $= 3.819$, $p = 0.148$) is similar across treatments. The similarity across treatments implies that participants' behavior in Part 1 is comparable even though the incentives to succeed in the real effort task vary.

The following analyses test for moral balancing by comparing participants that succeeded in the slider task with those who failed. Participants that succeeded in the slider task by solving at least 26 sliders are subsumed under *Success*, and those who failed under *Failure*. Table 5 presents summary statistics separately for *Success* and *Failure*. Significant differences exist concerning gender, age, parenthood, education, political orientation, altruism, beliefs regarding the primary causes of global warming, and the efficacy of personal actions in combating climate change (MWU, $p's < 0.043$). We account for these differences by controlling for these variables (hereafter denoted by Controls B) in the regression specifications.

[34] Participants are asked in the post-experimental questionnaire whether they agree that their personal actions matter to fight climate change.

Table 4: Descriptive statistic by treatment.

	Mean/ Median	SELF	LOW	HIGH	Test Statistic (p-value)
		(N=300)	(N=300)	(N=300)	
Demographics					
Female	Mean	0.623 (0.485)	0.640 (0.481)	0.613 (0.488)	0.465 (0.793)
Male	Mean	0.353 (0.479)	0.357 (0.480)	0.377 (0.485)	0.414 (0.813)
Non-binary	Mean	0.020 (0.140)	0.003 (0.058)	0.007 (0.082)	4.714 (0.095)
Age	Mean	32.197 (12.376)	30.343 (11.115)	32.350 (12.849)	4.073 (0.131)
Has children	Mean	0.337 (0.473)	0.287 (0.453)	0.287 (0.453)	2.366 (0.306)
Education	Median	Undergraduate degree (BA/BSc/other)			1.120 (0.571)
Political orientation: Left	Mean	0.580 (0.494)	0.630 (0.484)	0.547 (0.499)	4.350 (0.114)
Political orientation: Right	Mean	0.213 (0.410)	0.177 (0.382)	0.243 (0.430)	4.016 (0.134)
Income	Median	10,000 - 29,999	10,000 - 29,999	10,000 - 29,999	0.031 (0.861)
Behavioral Preferences					
Altruism	Mean	-0.042 (0.834)	0.031 (0.841)	0.011 (0.799)	0.279 (0.870)
Patience	Mean	0.062 (1.032)	-0.026 (0.962)	-0.036 (1.005)	3.340 (0.188)
Environmental Awareness					
SASSY segment	Median	Concerned	Concerned	Concerned	0.298 (0.862)
Global warming caused by humans	Mean	0.633 (0.483)	0.663 (0.473)	0.680 (0.467)	1.493 (0.474)
Actions matter to fight climate change	Mean	0.773 (0.419)	0.803 (0.398)	0.697 (0.460)	9.891 (0.007)
Pro-environmental behavior	Mean	7.350 (2.114)	7.523 (1.875)	7.370 (2.202)	0.480 (0.787)
Has offset in past	Mean	0.200 (0.401)	0.200 (0.401)	0.217 (0.413)	0.340 (0.844)
Carbon offset effective	Mean	0.677 (0.469)	0.683 (0.466)	0.653 (0.477)	0.675 (0.714)

Note: Standard deviation in parentheses for variables with means. For categorical variables, we use χ^2-tests; for numerical variables, we use Kruskal-Wallis tests. SELF, LOW, and HIGH denote the treatments. Pro-environmental behavior is measured with respect to its frequency on a scale from 0 (Never) to 10 (Very often).

Table 5: Descriptive statistic by success in Part 1.

	Mean/ Median	*Success*	*Failure*	Test Statistic (p-value)
		(N=449)	(N=451)	
Demographics				
Female	Mean	0.512 (0.500)	0.738 (0.440)	49.110 (0.000)
Male	Mean	0.477 (0.500)	0.248 (0.433)	50.753 (0.000)
Non-binary	Mean	0.011 (0.105)	0.009 (0.094)	0.117 (0.733)
Age	Mean	28.192 (8.858)	35.053 (13.912)	48.792 (0.000)
Has children	Mean	0.183 (0.387)	0.424 (0.495)	61.776 (0.000)
Education	Median	Undergraduate degree	Technical/community college	8.387 (0.004)
Political orientation: Left	Mean	0.630 (0.483)	0.541 (0.499)	7.388 (0.007)
Political orientation: Right	Mean	0.196 (0.397)	0.226 (0.419)	1.230 (0.267)
Income	Median	10,000 - 29,999	10,000 - 29,999	0.031 (0.861)
Behavioral Preferences				
Altruism	Mean	-0.067 (0.859)	0.067 (0.784)	4.092 (0.043)
Patience	Mean	-0.032 (1.015)	0.032 (0.985)	0.560 (0.454)
Environmental Awareness				
SASSY segment	Median	Concerned	Concerned	1.804 (0.179)
Global warming caused by humans	Mean	0.697 (0.460)	0.621 (0.486)	5.822 (0.016)
Actions matter to fight climate change	Mean	0.711 (0.454)	0.805 (0.397)	10.926 (0.001)
Pro-environmental behavior	Mean	7.372 (2.079)	7.457 (2.057)	0.610 (0.435)
Has offset in past	Mean	0.207 (0.406)	0.204 (0.403)	0.014 (0.907)
Carbon offset effective	Mean	0.695 (0.461)	0.648 (0.478)	2.293 (0.130)

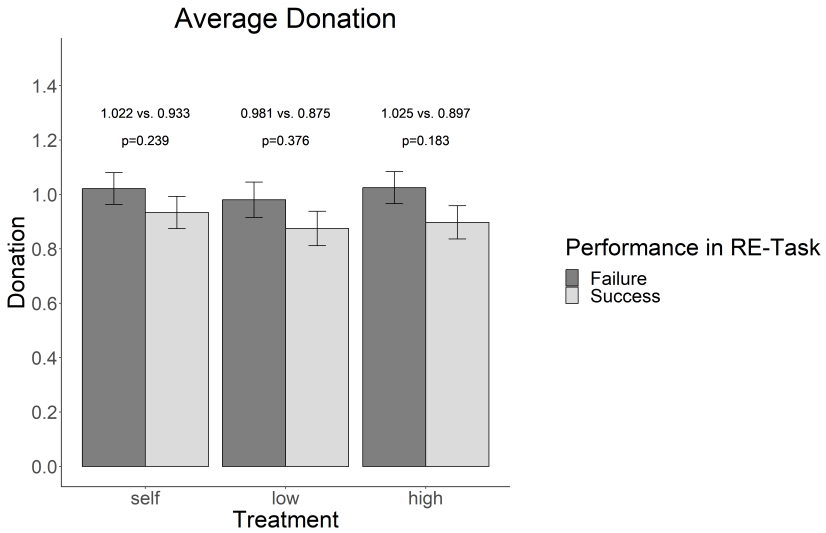
Note: Standard deviation in parentheses for variables with means. For categorical variables, we use χ^2-tests; for numerical variables, we use Kruskal-Wallis tests. *Success* and *Failure* denote performance in the slider task of Part 1. Pro-environmental behavior is measured with respect to its frequency on a scale from 0 (Never) to 10 (Very often).

Summary statistics by success in Part 1 separately for each treatment are shown in Table B.1, Table B.2, and Table B.3. Within a treatment, significant differences exist, particularly regarding gender, age, and having children.

5.2 Prevalence of moral balancing

To test our first set of hypotheses on whether moral balancing is prevalent in carbon offsetting, we separate the data into *Success* (carbon offset is acquired) and *Failure* (no carbon offset is acquired) in the slider task. According to the theory of moral balancing, participants in LOW and HIGH who acquire a positive moral self-image by succeeding in the slider task will donate less to carbon offsetting in Part 2 than participants that acquire a negative moral self-image by failing in the slider task.

Figure 5 displays the average donation for carbon offset by treatment and whether the slider task was completed successfully. In line with our first set of hypotheses, we find in LOW and HIGH that participants who succeeded in the slider task donate on average less than those who failed. While these differences are not significant (LOW: MWU, $z = 1.178$ $p = 0.239$; HIGH: MWU, $z = 1.332$ $p = 0.183$), pooling the LOW and HIGH groups reveals a significant difference at the 10%-level between those who acquired a carbon offset through the slider task and those who did not (MWU, $z = 1.802$ $p = 0.072$). Successful participants donated, on average, about 12% less than unsuccessful ones. In SELF participant's environmental self-image is not targeted, and the difference between those who succeeded and those who failed is not significant, in line with our hypothesis (MWU, $z = 0.886$ $p = 0.376$). The difference in offsetting between successful and unsuccessful participants indicates that moral balancing plays a role in carbon offsetting decisions.



Note: Donations for carbon offset in Part 2 by treatment and success in the slider task of Part 1 (n = 900). The bars show the average donation and the error bars represent the standard errors of the means. The figure also shows the respective means and p-values of MWU tests for differences between *Success* and *Failure* within a treatment.

Figure 5: Average donation by treatment and success in Part 1.

Besides non-parametric tests, we conduct regression analyses in Table 6 to test for the existence of moral balancing. In columns (1) - (3), we split the sample by treatment and regress the amount donated in Part 2 on a dummy capturing success in the slider task of Part 1. We observe negative but insignificant coefficients for all three treatments ($p's > 0.134$). In line

with moral balancing, the coefficients are greater in magnitude in HIGH and LOW compared to SELF.

In columns (4) - (6), we add controls to account for variables that are significantly different between participants who succeeded in the slider task and those who failed (for the selection of the controls, see Table 5). The coefficient for *Success* remains insignificant in all three treatments ($p's > 0.288$).

In column (7), we pool the observations from the LOW and HIGH treatment in which participants could obtain a moral license. The dummy capturing success in the slider task displays the expected negative sign and is significant on the 10%-level (coef = -0.118; 95% CI = -0.239, 0.004; $p = 0.058$). As a robustness check defined in the pre-registration, we excluded in column (8) participants that did not pass the attention check in the post-experiment questionnaire (0 observations in SELF, 4 in LOW, and 2 in HIGH). Similar to the regression specification in (7), the coefficient takes the expected negative sign, even increases in magnitude, and becomes significant at the 5%-level (coef = -0.124; 95% CI = -0.246, -0.001; $p = 0.048$). The results in (7) and (8) suggest that moral balancing influences carbon-offsetting decisions.

In column (9), we add controls for variables that are significantly different between *Success* and *Failure*. In addition, we control for participants' beliefs regarding how much others donate in Part 2 (*Belief Donation Others*) and the proportion of participants that succeeded in the slider task of Part 1 (*Belief Success Others*). The coefficient for *Success* is negative but insignificant (coef = -0.013; 95% CI = -0.148, 0.122; $p = 0.847$). We conclude that the differences in donation rates can not only be explained by moral balancing but are also driven by individuals' characteristics and beliefs regarding others' pro-environmental behavior.

Table 6: OLS regressions for moral balancing.

	Dependent variable: Donation								
	SELF	LOW	HIGH	SELF	LOW	HIGH	LOW + HIGH Pooled	LOW + HIGH Pooled Attentive	LOW + HIGH Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Success	-0.088	-0.106	-0.128	0.079	0.053	-0.096	-0.118*	-0.124**	-0.013
	(0.083)	(0.091)	(0.085)	(0.082)	(0.091)	(0.091)	(0.062)	(0.062)	(0.069)
Belief Donation Others									0.690***
									(0.047)
Belief Success Others									-0.0001
									(0.001)
Constant	1.022***	0.981***	1.025***	0.422	-0.353	0.824***	1.003***	1.009***	0.287*
	(0.058)	(0.065)	(0.059)	(0.324)	(0.253)	(0.245)	(0.044)	(0.044)	(0.163)
Controls B	No	No	No	Yes	Yes	Yes	No	No	Yes
N	300	300	300	300	300	300	600	594	600
R^2	0.004	0.005	0.008	0.302	0.279	0.174	0.006	0.007	0.401

Note: *Success* takes the value 1 if a participant succeeded in the slider task in Part 1, 0 otherwise. "Belief Donation Others" is a participant's guess on the average donation of other participants in Part 2, and "Belief Success Others" refers to a participant's guess on the proportion of participants that solved the slider task in Part 1. In Columns (4) - (6) and (9), we add controls for gender, age, having children, education, political orientation, altruism, opinion of whether climate change is predominantly human-caused, and whether participants agree that their actions matter to fight climate change. For columns (7) - (9), we pool the LOW and HIGH treatment. In Column (8), we exclude participants who did not pass our questionnaire's attention check. Robust standard errors are reported in parentheses; *p<0.1; **p<0.05; ***p<0.01.

In Figure 6, we analyze the distribution of our measurement for moral balancing - the donation for carbon offset in Part 2. Most participants donate 0, £1, or £2. In line with moral

balancing, we find over the whole distribution function that participants that acquired a moral license (*Success*) donate less than participants who did not (*Failure*). However, the distribution functions of donation are not significantly different for *Success* and *Failure* (KS, $p = 0.184$). The findings are similar when analyzing LOW and HIGH separately (KS, p's > 0.447), see also Figure B.1.

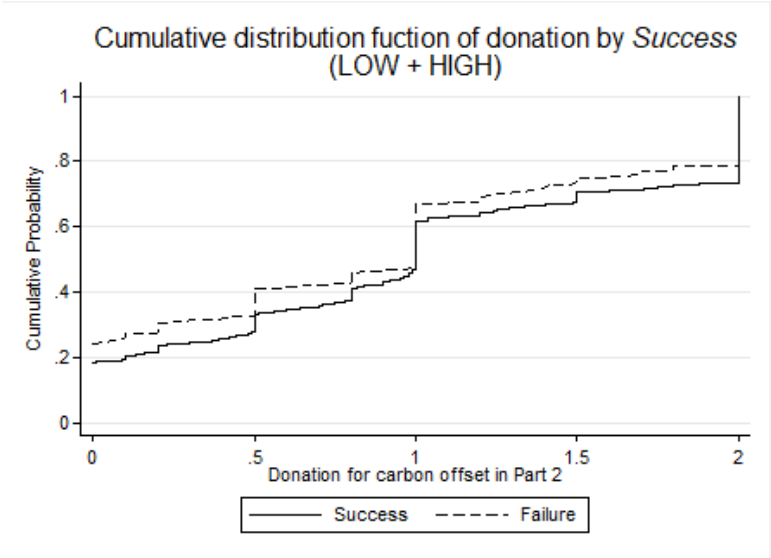


Figure 6: Cumulative distribution function of donation in Part 2 by success in Part 1.

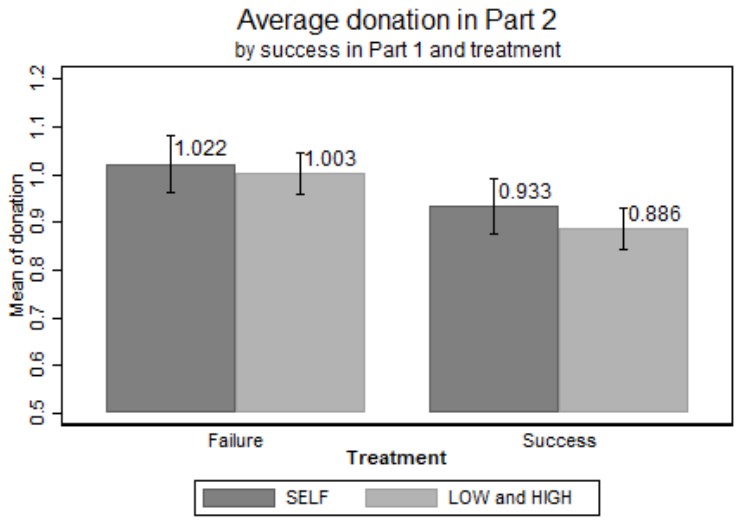
5.3 Disentangling moral licensing and cleansing

In the following, we aim to disentangle the moral balancing effect into moral licensing and cleansing.

To test for moral licensing, we restrict ourselves to successful participants. Figure 7 depicts that successful participants in SELF donate on average more than those in the other two treatments. The lower average donation rate in treatments where a license was acquired (LOW and HIGH) aligns with moral licensing. However, the difference lacks statistical significance (MWU, $z = 0.890$, $p = 0.374$).

To test for moral cleansing, we focus on unsuccessful participants and compare the average donation rate in SELF, where no carbon could be offset, with the average in LOW and HIGH, where participants failed to offset carbon (see Figure 7). Moral cleansing would predict a higher donation rate in LOW and HIGH since participants will try to restore their moral self-worth by donating more in Part 2. However, in our setup, the average donation rates in SELF are similar in magnitude to the average donations in LOW and HIGH (MWU, $z = 0.312$, $p = 0.755$).

Regression analyses in Table B.4 confirm the findings from the non-parametric tests. The lack of significant results when comparing SELF with LOW and HIGH is probably driven by the fact that there is also a gap in donation between successful and unsuccessful participants in the SELF treatment (compare Figure 5), which was designed to serve as a baseline treatment not affecting participants moral self-image. Despite not being significant, this gap makes it hard to



Note: Donations for carbon offset in Part 2 by success in the slider task of Part 1 and treatment (n = 900). The bars show the average donation and the error bars represent the standard errors of the means.

Figure 7: Average donation by success in Part 1 and treatment.

disentangle moral licensing and moral cleansing. We can only speculate about possible reasons. Maybe participants see their success in the slider task as a good deed towards the experimenter (de Quidt et al., 2018). Consequently, succeeding in the task increases a participant's moral self-image, whereas failing decreases it, resulting in moral balancing. In addition, succeeding (failing) in the task might lead to positive (negative) emotions (e.g., happiness, sadness, pride, or guilt) that have been shown to affect giving in donation games (e.g., Ibanez and Roussel, 2021; Tan and Forgas, 2010). E.g., Ibanez and Roussel (2021) find that participants donate less to an environmental charity when negative emotions were previously induced.

To subsume, we find weak but insignificant evidence for moral licensing and no evidence for moral cleansing in our setup.

5.4 Magnitude of the initial pro-environmental action

Regarding our second set of hypotheses, we explore if the magnitude of the moral action affects subsequent pro-environmental behavior. Figure 5 shows that the difference between the average donation of successful and unsuccessful participants is £0.11 in LOW and £0.13 in HIGH. This difference in average donation between the two treatments of £0.02 is low, especially considering that the monetary equivalent of the difference between the carbon offset in HIGH and LOW is £1.8. Moreover, we compare the difference in donation rates between LOW and HIGH separately for successful and unsuccessful participants. The differences are not significant (*Success*: MWU, $z = 0.297$ $p = 0.767$; *Failure*: MWU, $z = 0.370$ $p = 0.712$), which supports our second set of hypotheses and indicates that the magnitude of the initial action does not significantly impact the subsequent donation decision.

Regression analyses in Table 7 confirm the non-parametric results. We find neither for the successful nor the unsuccessful subjects that the donation rates differ between having ac-

quired a substantial (HIGH) or a negligible (LOW) carbon offset (see columns (1) and (3)) ($p's > 0.613$). The results are similar when controlling for *Actions Matter* (see columns (2) and (4)) ($p's > 0.321$).

We conclude that the magnitude of the initial action does not significantly impact participants' subsequent pro-environmental decision-making.

Table 7: OLS regressions separately by *Success* and *Failure*.

	Dependent variable: Donation			
	Failure		*Success*	
	(1)	(2)	(3)	(4)
HIGH	0.044	0.086	0.022	0.061
	(0.088)	(0.087)	(0.088)	(0.085)
Constant	0.981***	0.704***	0.875***	0.540***
	(0.065)	(0.102)	(0.063)	(0.087)
Control A	No	Yes	No	Yes
N	301	301	299	299
R^2	0.001	0.030	0.0002	0.071

Note: The sample is restricted to the LOW and HIGH treatment and split by *Failure* (columns (1) and (2)) and *Success* (columns (3) and (4)) in Part 1. HIGH takes the value 1 if a participant is in the HIGH treatment and 0 if she is in the LOW treatment. In Columns (2) and (4), we control for whether participants agree that their actions matter to fight climate change; as for this variable, we find significant differences between treatments (see Table 4). Robust standard errors are reported in parentheses; *p<0.1; **p<0.05; ***p<0.01.

5.5 Heterogeneous effects

We preregistered to explore heterogeneous effects to identify moderators of moral balancing. In Table 8, we test for heterogeneous effects concerning gender (column (1)), age (column (2)), altruism (column (3)), education (column (4)), political orientation (column (5)), previous offsetting actions (column (6)), participants' pro-environmental behavior (column (7)), and concerns regarding global warming elicited via the SASSY (column (8)). We define a dummy for each variable that splits our sample into two categories.

In general, we do not find that individual characteristics moderate moral balancing (see columns (1) - (7)) ($p's > 0.113$). However, we find a significant heterogeneous effect regarding climate change concerns measured by the SASSY ($p's = 0.042$) (see column (8). We conclude that environmental concerns moderate moral balancing, which is in line with the literature (e.g., Effron et al., 2009; Meijers et al., 2019) (see also Section 2.2). In the subsequent section, we provide a more detailed analysis of the impact of environmental concerns.

5.6 Environmental concern

Figure 8 shows the average donation rates by treatment and success in the slider task, separately for *Alarmed* and *Not Alarmed* participants.[35]

[35]The SASSY categorizes 44% of participants as *Alarmed*, 41% as *Concerned*, 12% as *Cautious* and 3% as either *Disengaged*, *Doubtful*, or *Dismissive*. Since the number of observations in the three latter categories is low, we drop

Table 8: OLS regressions for heterogeneous moral balancing effects.

	Dependent variable: Donation							
	(1) *Female*	(2) *Older*	(3) *More Altruistic*	(4) *Higher Education*	(5) *Right Party*	(6) *Never Offset*	(7) *Pro-env.*	(8) *SASSY Alarmed*
Success	0.023	-0.017	-0.083	-0.050	0.064	-0.041	0.074	-0.135
	(0.109)	(0.086)	(0.089)	(0.091)	(0.077)	(0.071)	(0.089)	(0.085)
Interacted Variable	0.148	0.240**	0.365***	0.015	0.226**	-0.068	0.158*	0.081
	(0.099)	(0.098)	(0.086)	(0.083)	(0.107)	(0.110)	(0.087)	(0.086)
Success ×	-0.063	-0.063	0.083	0.028	-0.210	0.079	-0.183	0.241**
Interacted Variable	(0.128)	(0.121)	(0.117)	(0.117)	(0.144)	(0.143)	(0.115)	(0.118)
Constant	0.147	0.464	-0.084	0.454***	0.231	0.226	0.133	-0.097
	(0.254)	(0.232)	(0.272)	(0.152)	(0.218)	(0.243)	(0.226)	(0.172)
Controls B	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	596	600	600	599	479	600	600	582
R^2	0.195	0.194	0.185	0.193	0.226	0.200	0.204	0.214

Note: The sample is restricted to the LOW and HIGH treatments. *Female* takes the value 1 if the participant is female and 0 if male. Participants answering *Non-binary* or *Rather not say* are excluded. *Is Older* takes the value 1 if age ≥ 28 (median), and 0 otherwise. *More Altruistic* takes the value 1 if Altruism ≥ 0, and 0 otherwise. *Higher Education* takes the value 1 if Education ∈ {undergraduate degree (Ba/Bsc/other), graduate degree (MA/MSc/MPhil/other), doctoral degree (PhD/other)}, and 0 if in {no formal qualifications, secondary education, high school diploma/A-levels, technical/community college}. We excluded those who answered {don't know/not applicable}. *Right Party* takes the value 1 if a participant identifies with ∈ {Conservative, Liberal Democrats}, and 0 if ∈ {Labour, SNP, Green Party}. We excluded those who answered {other, rather not say}. *Never Offset* takes the value 1 if participant had never offset CO_2 before the experiment, and 0 otherwise. *Pro-env. behavior* takes the value 1 if a participant answers an 8 (median) or higher on a scale from 0 (Never) to 10 on the frequency s/he is taking environmentally friendly actions. *SASSY Alarmed* takes the value 1 if a participant is in the sassy segment ∈ {Alarmed}, and 0 if ∈ {Cautious, Concerned}. The following control variables are included: gender, age, having children, education, political orientation, altruism, opinion on whether climate change is predominantly human-caused, and whether participants agree that their actions matter in fighting climate change. The respective variable is not included in the set of controls if it is considered in the specification. Robust standard errors in parentheses; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

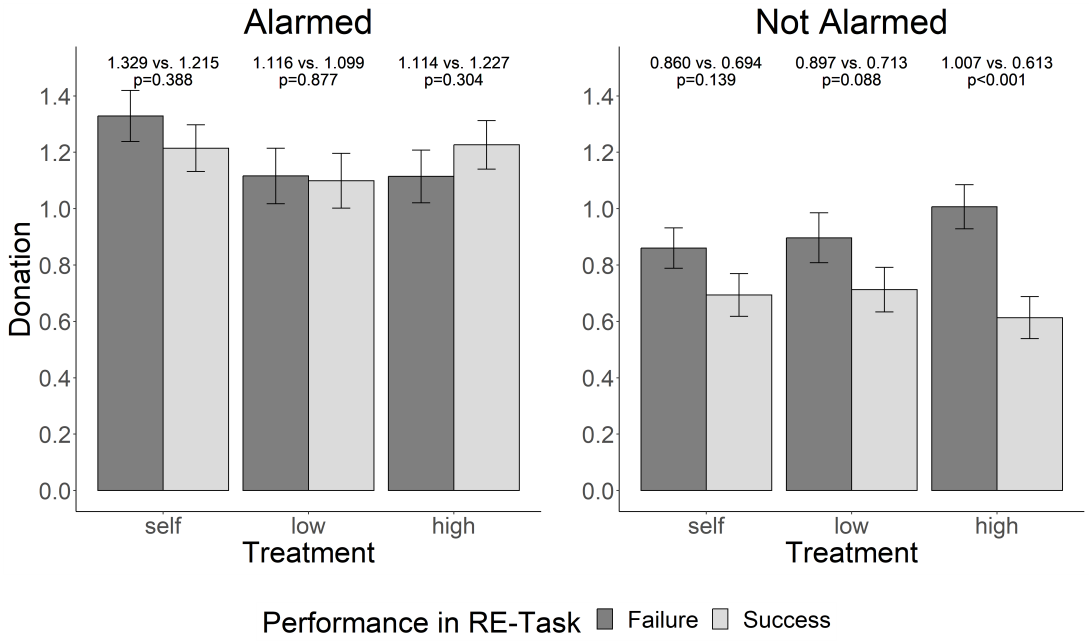
Participants in *Alarmed* do not show different donation rates depending on whether they acquired a moral license by offsetting carbon in Part 1 (LOW: MWU, $z = 0.157$, $p = 0.876$; HIGH: MWU, $z = -1.030$, $p = 0.303$). For participants in *Not Alarmed*, we find a significant difference in donations between those who succeeded versus those who failed in HIGH (MWU, $z = 3.333$, $p = 0.001$) as well as LOW (MWU, $z = 1.710$, $p = 0.088$). Hence, participants with the greatest environmental concerns do not engage in moral balancing, whereas participants who show less concern base their donation decisions on previous offsets.

In Table 9, we report regression analyses that support our non-parametric findings. We do not find moral balancing for the *Alarmed* sub-sample (W, $F(1)$, $p's > 0.377$). Whereas participants in *Not Alarmed* and HIGH engage in moral balancing (W, $F(1)$, $p's < 0.001$). Successful individuals donate on average about £0.39 (39%) less than those who failed to acquire a moral licensing.

In column (2), we add controls to account for significant differences in individual characteristics between *Success* and *Failure*. We find a *higher* donation rate for successful participants for the *Alarmed* and HIGH group, which opposes moral balancing. However, the difference is only significant at the 10%-level (W, $F(1)$, $p's = 0.077$). Participants in HIGH and *Not Alarmed* donate on average less after successfully offsetting carbon in Part 1 (W, $F(1)$, $p's < 0.033$), which is in line with moral balancing.

We conclude that participants with somewhat lower environmental concerns engage in moral balancing after (not) acquiring the high carbon offset. Highly concerned individuals

observations classified as either *Disengaged*, *Doubtful*, or *Dismissive* from our analyses. We pool *Concerned* and *Cautious* and refer to it as *Not Alarmed*.



Note: Donations for carbon offset in Part 2 by SASSY segment, treatment, and success in the slider task of Part 1 (n = 900). *Not Alarmed* denotes the SASSY segment $\in$ {Cautious, Concerned}. The SASSY segments *Doubtful*, *Disengaged*, and *Dismissive* are excluded due to the small number of observations. The bars show the average donation and the error bars represent the standard errors of the means. The figure also shows the respective means and p-values of MWU tests for differences within treatment between *Success* and *Failure*.

Figure 8: Average donations by SASSY segment, treatment, and *Success*.

do not base their offset decisions on past acquired offsets. One explanation could be that less concerned individuals use their previous behavior to justify less environmental behavior later, whereas individuals with a high environmental identity always strive to act environmentally friendly.

5.7 Beliefs regarding others' behavior

To analyze the impact participants' beliefs about others' behavior have on moral balancing, we elicit incentivized guesses on the percentage of participants that succeeded in Part 1 and the average offset in Part 2.[36] Participants expected a higher success rate in Part 1 (belief: 58.61%, actual success rate: 49.89%) and that others donate less to carbon offset (belief: £0.84, actual donation rate: £0.96).

In Table 10, column (1), we regress donation in Part 2 on *Success*, a participant's belief on how much other participants donated, and the interaction of both variables. We restrict the sample to the treatments in which a license could be acquired (LOW and HIGH). We find a significant positive correlation between an individual's donation and his or her beliefs about others' donations (coef = 0.719; 95% CI = 0.585, 0.853; $p < 0.001$). The interaction term is positive and significant on the 10%-level (coef = 0.158; 95% CI = -0.0149, 0.331; $p = 0.073$), which suggests that moral balancing correlates with a belief that others donated little. However,

[36]For every guess that was less than 5% away from the actual value, participants received a bonus payment of £0.20.

Table 9: OLS regressions for heterogeneous balancing effects with respect to SASSY and treatment.

	Dependent variable: Donation	
	(1)	(2)
β_1: LOW	-0.213	-0.116
	(0.134)	(0.122)
β_2: HIGH	-0.214*	-0.171
	(0.130)	(0.120)
β_3: *Not Alarmed*	-0.469***	-0.284**
	(0.115)	(0.111)
β_4: *Success*	-0.114	0.030
	(0.122)	(0.116)
β_5: LOW × *Not Alarmed*	0.249	0.130
	(0.176)	(0.162)
β_6: HIGH × *Not Alarmed*	0.361**	0.280*
	(0.168)	(0.160)
β_7: LOW × *Success*	0.097	-0.008
	(0.185)	(0.169)
β_8: HIGH × *Success*	0.226	0.189
	(0.176)	(0.165)
β_9: *Not Alarmed* × *Success*	-0.052	-0.018
	(0.161)	(0.148)
β_{10}: LOW × *Not Alarmed* × *Success*	-0.115	-0.029
	(0.243)	(0.222)
β_{11}: HIGH × *Not Alarmed* × *Success*	-0.454*	-0.440**
	(0.232)	(0.220)
β_0: Constant	1.329***	0.022
	(0.090)	(0.181)
Controls B	No	Yes
N	874	874
R^2	0.086	0.234
H_0: *Alarmed*-SELF: $\beta_4 = 0$	0.351	0.798
H_0: *Alarmed*-LOW: $\beta_4 + \beta_7 = 0$	0.902	0.766
H_0: *Alarmed*-HIGH: $\beta_4 + \beta_8 = 0$	0.377	0.077*
H_0: *Not Alarmed*-SELF: $\beta_4 + \beta_9 = 0$	0.112	0.907
H_0: *Not Alarmed*-LOW: $\beta_4 + \beta_7 + \beta_9 + \beta_{10} = 0$	0.122	0.929
H_0: *Not Alarmed*-HIGH: $\beta_4 + \beta_8 + \beta_9 + \beta_{11} = 0$	0.0003***	0.033**

Note: *SASSY Alarmed* takes the value 1 if a participant is in the sassy segment $\in$ {Alarmed}, and 0 if $\in$ {Cautious, Concerned}. We exclude observations $\in$ {Disengaged, Dismissive, Doubtful} due to the low number of observations (n = 26). The following control variables are included: gender, age, having children, education, political orientation, altruism, opinion on whether climate change is predominantly human-caused, and whether participants agree that their actions matter in fighting climate change. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. Standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

when controlling for variables for which we observe significant differences between success and failure in Part 1 the interaction gets insignificant (coef = 0.100; 95% CI = -0.065, 0.265; $p = 0.235$).

We see in columns (3) and (4) that a participant's belief regarding the proportion of successful participants does not predict donation in Part 2 ($p > 0.475$). Also, the interaction between *Success* and belief is not significant $p > 0.110$), indicating that the belief regarding others' success rate in Part 1 does not moderate moral balancing.

Table 10: OLS regressions for beliefs on moral balancing.

	Dependent variable: Donation			
	(1)	(2)	(3)	(4)
Success	-0.193**	-0.101	0.149	0.294
	(0.091)	(0.086)	(0.207)	(0.201)
Belief Donation Others	0.719***	0.644***		
	(0.068)	(0.066)		
Success × Belief Donation Others	0.158*	0.100		
	(0.088)	(0.084)		
Belief Success Others			0.0004	0.001
			(0.002)	(0.002)
Success × Belief Success Others			-0.004	-0.005
			(0.003)	(0.003)
Constant	0.366***	0.365**	0.985***	0.064
	(0.073)	(0.164)	(0.093)	(0.270)
Controls B	No	Yes	No	Yes
N	600	600	600	600
R^2	0.306	0.402	0.009	0.693

Note: The sample is restricted to the LOW and HIGH treatment. *Success* takes the value 1 if a participant succeeded in the slider task in Part 1, 0 otherwise. "Belief Donation Others" is a participant's guess on the average donation of other participants in Part 2, and "Belief Success Others" refers to a participant's guess on the proportion of participants that solved the slider task in Part 1. In Columns (2) and (4), we add controls for gender, age, having children, education, political orientation, altruism, opinion whether climate change is predominantly human-caused, and whether participants agree that their actions matter to fight climate change. Robust standard errors are reported in parentheses; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

6 Discussion and conclusion

Motivated by the increasing awareness for environmentally-friendly behavior, we conduct an incentivized online experiment to explore how moral balancing affects pro-environmental decision-making. We add to the literature by testing whether the magnitude of the initial action influences the size of the moral balancing effect. Participants first work on a real-effort slider task and - depending on the treatment - acquire either a 10 kg CO_2 offset, a 100 kg, or a pay-off of £0.2 for themselves. Next, participants receive a windfall endowment and decide how much to donate to carbon offset. We investigate if participants engage in moral balancing, such that they base their offsetting decisions in the second part on previous success or failure in the real-effort task. Since the treatments vary the effectiveness of the real-effort task, we evaluate if moral balancing depends on the magnitude of the initial action.

We find evidence for moral balancing in offsetting decisions. Pooling the treatments in which participants can acquire carbon offset, participants who succeeded in the slider task donate on average less than those who failed. The difference is significant at the 10%-level. Regarding the magnitude of the initial pro-environmental action that varies exogenously across treatments, we do not find consistent evidence that it affects the size of the moral balancing effect. In addition, we find that environmental concerns moderate moral balancing. Individuals with the highest level of environmental concerns do not base their offset decision on success in the previous task, whereas somewhat less concerned individuals engage in moral balancing, particularly in the treatment where they could offset 100 kg of carbon emissions.

Our findings have implications for the design of environmental marketing campaigns and

the welfare evaluation of the voluntary carbon offsetting market. For example, campaigns targeting individuals with a lower level of environmental consciousness need to consider the influence of moral balancing, as they might result in reduced environmental behavior following the campaign. Moreover, environmental campaigns that stress that every act counts could backfire as moral balancing, in general, does not systematically depend on the magnitude of the initial action. Environmental messaging should prioritize promoting actions with a substantial positive environmental impact to mitigate potential adverse spillover effects. Evaluating the benefits of the voluntary carbon offset market solely based on the total amount of carbon offset might lead to overestimating its positive impact. If individuals engage in moral balancing and behave less environmentally friendly after acquiring the license, it diminishes the positive impact of the initial offset. In conclusion, our findings emphasize the importance of considering moral balancing effects when targeting pro-environmental behavior.

Chapter IV

On the robustness of gender differences in economic behavior[37]

1 Motivation

Every day, people all around the world make economic decisions that impact various aspects of their lives: Should I apply for a new job in a highly competitive environment? Should I invest in a high-risk asset or not? How much money should I donate to charities? A vast literature tries to determine the factors that affect decisions in domains such as competitiveness (Villeval, 2012), risk-taking (Thöni and Volk, 2021), and altruism (Bilén et al., 2021). Researchers have looked, among other things, into the role of institutional or market-related features (Balafoutas and Sutter, 2012; Balafoutas et al., 2018; Cassar et al., 2016; Cassar and Rigdon, 2021; Fornwagner et al., 2023; He et al., 2021; Niederle and Vesterlund, 2007; Sisco and Weber, 2019), individual characteristics (Almås et al., 2016*b*; Buser et al., 2018; Guiso and Paiella, 2008; Gutiérrez-Roig et al., 2014; Sutter and Glätzle-Rützler, 2015; von Gaudecker et al., 2011), cultural background (Cárdenas et al., 2015; Croson and Gneezy, 2009; Gneezy et al., 2009; Gong and Yang, 2012; Liu and Zuo, 2019; Wu et al., 2015), hormonal (Boksem et al., 2013; Ranehill et al., 2018; Sapienza et al., 2009; van Anders et al., 2015; Zak et al., 2009; Zethraeus et al., 2009), or other biological factors, such as genetics, and neurological factors (Anderson et al., 2015; Cesarini et al., 2012; Grubb et al., 2016; Moll et al., 2006; Reuter et al., 2011). Among those factors, gender has received a lot of attention. Over the last few decades, extensive research in economics has looked at whether gender is a significant driver of how women and men behave in the domains of competitiveness (Beblo and Markowsky, 2022), risk-taking (Nelson, 2017; Thöni and Volk, 2021), and altruism (Bilén et al., 2021). We refer to Section 2 for a detailed literature review.

But is it really gender that influences behavior? Or, instead, are sex differences causing these observed differences? Or is it a mix of gender and sex? Importantly, sex and gender are two distinct concepts. Whereas sex is defined as "either of the two main categories (male and female) into which humans" are categorized based on their reproductive functions (Oxford En-

[37] This chapter is based on Fornwagner et al. (2022) published in Scientific Reports.

glish Dictionary, n.d.), gender usually refers to the psychological, behavioral, social, and cultural aspects of being male or female (i.e., masculinity or femininity) (Kessler and McKenna, 1985, 2000; VandenBos, 2007). For cisgender individuals, the internal gender identity matches and presents itself by the externally determined cultural expectations of the behavior and roles considered appropriate for one's sex (VandenBos, 2007). However, the gender identity of transmen and transwomen and their gender roles are typically not the same as what is associated with their sex assigned at birth (American Psychological Association, 2015). So the question arises: How much of the differences between men and women often found in the economic literature can really be associated with gender as opposed to an individual's sex?

We investigate this question by using well-known behavioral economic experiments in the domain of *competitiveness*, *risky choices*, and *altruism*. As stated, for these three behavioral traits, gender differences are a common finding. However, existing studies identify gender effects without controlling for sex. Distinguishing gender from sex effects is practically impossible when only investigating cisgender participants. As a novel approach, we run our experimental study with transmen and transwomen in addition to cismen and ciswomen. We do not use the gender that is attributed to a person by others (Federici et al., 2022; Kessler and McKenna, 1985, 2000; Wenzlaff et al., 2018). Instead, this study utilizes the information on the participants' self-identification to a particular gender and sex from self-reported categories and established scaling methods from psychological and medical science. The advantage of having this information is that cisgender and transgender people differ in either their sex or their gender. To illustrate this consider an example: a ciswoman has female sex *and* feminine gender. A transman has female sex *but* masculine gender. So differences in the behavior of those two subject groups might be associated with gender instead of sex.

The experimental method is excellent for studying the economic choices we are interested in because of its standardized and validated measures. We have information on the participants' gender and sex from self-reported categories and established scaling methods from psychological and medical science. Moreover, instead of just analyzing gender and sex effects correlationally, we elicit the causal impact of gender by exogenously varying gender identities with a priming method.

First, we test how gender correlates with the mentioned choices. By contrasting the behavior of the four different subject groups of cismen, ciswomen, transmen, and transwomen, we obtain insights into how far biology (sex) or the cultural and sociological construct of gender explains differences in economic behavior. Our study is the first to investigate competitiveness, risk-taking, and altruism of transmen and transwomen. We hypothesize that if gender is the driving factor, individuals of the same gender (and different sex) make similar decisions, and decisions significantly differ when gender differs (and sex is the same). Second, we concentrate on the causal effect of gender on behavior – an analysis rarely done in the literature. The traditional experimental method of randomizing over the variable of interest is not possible with gender. Hence, we need a different approach to elicit causal effects. As our method to test a directional impact of gender, we employ a gender prime: either a masculine or feminine gender identity is subconsciously activated. Priming is an easy-to-implement intervention that

has shown to influence individual decision-making in various dimensions. Amongst others, it has been used to activate gender identities or change gender stereotypes (Meier-Pesti and Penz, 2008; Steele and Ambady, 2006). Those studies' results are mixed, depending on the objective of the prime (e.g., risk preferences, competitiveness, altruism) and the method of priming (eliciting gender at the beginning of the study or showing pictures).

In our study, we use a word priming method that has shown to be powerful in other contexts (Bargh et al., 2001; Mussweiler and Förster, 2000; Pichon et al., 2007), and has the advantage that we can easily include a gender-neutral condition by using gender-neutral words. In general, it seems to be the case that different genders react differently to gender priming. Importantly, none of the existing priming studies has recruited transgender subjects as researchers usually rely only on self-reported (binary) gender identities. If cisgender and transgender individuals change their behavior when being primed, this would indicate a causal effect of gender on individual economic decisions. To be more specific, our hypotheses are as follows. First, since our priming affects individuals' gender identity and not their sex, we anticipate participants with the same gender to react similarly to the respective prime. Put differently, cismen and transmen (ciswomen and transwomen) should adjust their behavior similarly when being primed. Second, we expect reactions to priming to be different when the gender is not the same among the participants. Lastly, the results should be different when participants are primed with their own gender identity instead of their respective other gender identity.

Based on 780 observations from experiments conducted online, our results generally show no correlational or causal effect of gender or sex for competitiveness, risk-taking, and altruism. The only exceptions are that cismen have a higher rate of entering the competition than all other subject groups when primed masculine. They also risk more when primed with a masculine identity compared to the neutral priming condition. In addition, we find that subjects of male sex (i.e., cismen and transwomen) risk more than their female counterparts (ciswomen and transmen). However, these behavioral differences that sometimes point towards gender and sometimes towards sex as explanatory variables do not replicate if we apply different robustness tests, including correcting for multiple hypothesis testing. Thus, we conclude that neither gender nor sex is a consistent main factor influencing the economic decisions measured in this article.

2 Related literature

2.1 Competitiveness

Differences in competitiveness have become an essential explanation for labor market outcomes like variations in wages (Card et al., 2016), and different demands in wage negotiations (Leibbrandt and List, 2015). Pinning down the causes and consequences of the willingness to compete is important as it correlates with several relevant choices and characteristics for education and labor market outcomes (Shurchkov and Eckel, 2018). For example, subjects who are more competitive have been found to be more likely to choose competitive educational programs (Almås et al., 2016*a*; Buser et al., 2014, 2021; Reuben et al., 2017), to have a higher

income (Buser et al., 2018; Kamas and Preston, 2015; Reuben et al., 2015) and to become entrepreneurs (Berge et al., 2015). But what role does one of the main human characteristics - being a man or a woman - play for competitiveness?

During the last decades, an impressive amount of scientific evidence showed that women are generally less competitive than men (Almås et al., 2016*b*; Balafoutas and Sutter, 2019; Datta Gupta et al., 2013; Niederle and Vesterlund, 2007; Niederle, 2017; Saccardo et al., 2018; Sutter and Glätzle-Rützler, 2015). This gender gap in competitiveness (henceforth GGC) is robust when using different scientific methods. Studies report that men are more likely to compete when using classical lab (Niederle and Vesterlund, 2007), lab-in-the-field (Gneezy et al., 2009), field (Hogarth et al., 2012), and online experiments (Buser et al., 2021). The findings also replicate when using subjects from different age groups like children (Sutter and Glätzle-Rützler, 2015), students (Niederle and Vesterlund, 2007), and non-students (Andersen et al., 2013).

Recently some evidence has been collected on the lack of a GGC in certain circumstances. For example, for the matriarchy of Masai in Kenya, adult women are reported to be even more competitive than men (Gneezy et al., 2009). Similarly, children living in the Khasi matrilineal society in northeast India are equally competitive (Andersen et al., 2013). Without the need to go afar, it has been shown that the type of school children attend influences competitiveness, with female students from girl's schools being as competitive as boys (Booth and Nolen, 2012). Moreover, for children from families with lower socioeconomic backgrounds, no GGC is reported (Almås et al., 2016*b*). Also, cultural differences play a role in competitiveness, as shown by Cárdenas et al. (2015). They found that children are equally competitive in Columbia, but boys in Sweden are more competitive than girls. These mentioned studies suggest that women's lower willingness to compete is not something they are born with but rather a behavioral preference that can be influenced by different factors and can thus be addressed to nurture rather than nature.

Support for this perspective is provided by research showing that the GGC can be closed or reversed when using interventions that do not influence participants' biological makeup. For example, some studies change the institutional environment to resemble different affirmative action policies and obtain gender balance in competitive environments (Balafoutas and Sutter, 2012; Baldiga and Coffman, 2018; Leibbrandt et al., 2018; Niederle and Vesterlund, 2007). Others use the easy-to-implement intervention of priming (Balafoutas et al. (2018) and Cadsby et al. (2013)), which encourages women to enter competitions more often. Moreover, giving feedback about relative performance (Wozniak et al., 2016) and the earnings implications related to competition avoidance (Kessel et al., 2021) successfully increases women's entry rates, as well as when more experienced people advise strong-performing women to compete (Brandts et al., 2015). Besides, when the price of the competition benefits not the participants themselves but their offspring, again, no GGC has been observed (Cassar et al., 2016).

However, it is also plausible that biological factors like genes and hormones may lead to different decisions of women and men and are also a primary driver of behavior. Thus, a new and still developing field of research focuses on competitiveness from a more elementary

perspective by taking hormones into account. Up to now, there is only one study by Ranehill et al. (2018), which causally analyses the effect of estrogen and progestin (by administrating oral contraceptives) on competitiveness. The authors find no impact of the two hormones on the willingness to compete. All other studies use self-reported hormonal measures by asking female participants about their menstrual cycle day and taking hormonal contraceptives to infer their hormonal levels. Using self-reports is noisy (for a detailed discussion of why this is the case, see, Dreber and Johannesson (2018)) and leads to mixed findings on whether hormones play a role for competitiveness or not (Buser et al., 2018; Wozniak et al., 2014).

The existing evidence already provides results on what factors correlate with competitive behavior and how differences in competitiveness between men and women can be closed. However, this paper will be the first to test the robustness of the GGC when priming subjects with a specific gender identity. Moreover, we contribute to the literature by investigating the willingness to compete of transgender subjects. To the best of our knowledge, no economic experiments have been done using transgender participants. According to our review of the literature, considering the behavior of LGBTQ+ individuals is extremely rare in experimental economics. We only found one paper on homosexuality and competitiveness by Buser et al. (2018). These aspects point out our study's potential to expand the knowledge in the domain of competitive behavior.

2.2 Risk

Risk-taking is considered a fundamental determinant of individual behavior in different domains like health (Anderson and Mellor, 2008; Barsky et al., 1997), stock market participation (Almenberg and Dreber, 2015), saving decisions (Sutter et al., 2013), occupational and self-employment choices (Bonin et al., 2007), personal and household finance (Bucciol and Miniaci, 2011; Guiso and Paiella, 2008), education (von Gaudecker et al., 2011) and environmental decision making (Gollier, 2001). One strand of the literature in Behavioral Economics reports seemingly strong evidence for women preferring to take less risk compared to men (Charness and Gneezy, 2012). This difference in risk-taking is robust when using different experimental methods to measure risk, such as lotteries (Holt and Laury, 2002), investment games (Gneezy and Potters, 1997) or card games (Czibor et al., 2019). It is also reported for subjects varying from children (Cárdenas et al., 2015), to students (Croson and Gneezy, 2009), to non-students (Hardies et al., 2013). Moreover, the difference is not influenced by conducting the experiment in the lab or in other environments like on online platforms (Hardies et al., 2013).

Another strand of the literature does not support that risk-taking differs by gender. Those papers mainly concentrate on different underlying methodologies than those used by the studies mentioned above. First, they claim that it is important to clearly distinguish between differences on the individual level (categorical differences between men and women) and patterns that appear only at the aggregate level (such as, e.g., statistically detectable different means) (Nelson, 2016). Second, using quantitative measures of substantive differences that are not yet that common in economic studies (such as Cohen's *d*) or measures of substantive overlap (like, e.g., the *Index of Similarity*) also results in not having a substantially large gender gap

in risk-taking (Nelson, 2015, 2016). Nelson (2015), e.g., claims that standardized differences in means across gender mostly amount to less than one standard deviation and that the degree of overlap in distributions of risk-taking behavior of men and women is generally exceeding 80%. On average differences between (cis-) men and women in behavior are smaller than sex differences in, e.g., height or throwing ability (Hyde, 2005) and pale next to the effects of aspects such as cultural manipulations or gender priming (e.g., Nelson, 2015). These papers align with the so-called gender similarities hypothesis from the psychological literature, which argues that males and females are similar on most, but not all, psychological dimensions (Hyde, 2005). Nelson (2014) claims that one explanation for gender differences in risk-taking still being a prominently repeated finding is that science is biased towards these results because of, e.g., exiting stereotypes or confirmation bias for existing publications.

Several studies analyze gender differences in risk preferences for sub-populations of managers (Adams and Funk, 2012; Atkinson et al., 2003; Croson and Gneezy, 2009) and find that females are similar or even less risk-averse than men. The reasons could be a selection or social learning and adaptive behavior to the job demands. To disentangle these different factors, Drupp et al. (2020) uses an online experiment with scientists. They vary the salience of either the private or the professional identity of the subjects. They report that priming the professional identity reduces the gender gap in risk-taking. Besides, the gender gap decreases with increasing age as female senior scientists choose riskier options in the treatment where the profession is made salient.

Also, attempts to explore the connection between biological factors and risk-taking are taken for the domain of risky behavior. First, studies are exploring the causal effect of hormones on behavior.[38] For example, Zethraeus et al. (2009) test for administered testosterone or estrogen affecting women's risk-taking. No effect of either testosterone or estrogen on risk-taking could be detected. In line, the study by Boksem et al. (2013) and Buskens et al. (2016) find no effect of testosterone on risk aversion. Ranehill et al. (2018) take a comparative approach and administer an oral contraceptive or not. Again, no connection between hormones and behavior is reported. Second, studies test for the correlation between the variation in risk-taking and genes. On the one hand, for example, Anderson et al. (2015) find no relationship between the dopamine and the serotonin gene and risk-taking. On the other hand, studies using, for example, the twin methodology and genome-wide association techniques (GWAS) report genetic foundations for the willingness to take risk (Cesarini et al., 2009, 2010, 2012). Third, a recent study by Keaveney et al. (2020) showed that taking a small dose of Acetaminophen, a popular pain killer, increases risk-taking.

Several researchers prime subjects and study the effect on risk-taking (Erb et al., 2002; Gilad and Kliger, 2008; Guiso et al., 2018; König-Kersting and Trautmann, 2018; Newell and Shaw, 2017). The study closest to our research is Benjamin et al. (2010), which finds that making the subject's gender salient with a short questionnaire does not impact risk preferences. Also, Meier-Pesti and Penz (2008) report an effect of gender priming through questions and stereotypical pictures only on male risk preferences. Cohn et al. (2017) prime financial profes-

[38] For our literature review, we summarize only studies that concentrate on pharmacological testosterone administration with double-blind placebo-controlled designs, which allows us to interpret results causally.

sionals with their professional salience, which leads to a decrease in risk-taking in a high-stakes investment game. With a similar subject pool, Cohn et al. (2015) find that individuals primed with a bust scenario are more risk-averse compared to those primed with a boom scenario. Alempaki et al. (2019) test the robustness of the results of Cohn et al. (2015) with an Amazon Mechanical Turk subject pool. They report no evidence of priming influencing risk-taking. Callen et al. (2014) primed individuals who were exposed to violence by asking them to either recall happy, fearful or neutral moments. They find that remembering frightening experiences leads to a higher preference for certainty.

The only related study we are aware of that investigates the risk-taking behavior of LGBTQ+ individuals is Buser et al. (2018). It analyzes risk preferences by asking the subjects about their risk perception (survey question). It finds no significant differences between homosexual and heterosexual men and homosexual and heterosexual women.

2.3 Altruism

To what extent someone is pro-social, i.e., altruistic, is argued to explain behavior in the labor market, how individuals vote, whether they take up volunteer work, and how willing someone is to give to a charity (Bilén et al., 2021). Altruistic behavior is typically measured with a dictator game, where participants are asked how much they want to transfer to an anonymous other participant (Forsythe et al., 1994; Kahneman et al., 1986), or how much they wish to donate to a charity (Eckel and Grossman, 1996). It is a robust finding that experiment participants transfer quite a substantial part of their endowment in dictator games, thus acting altruistically (Carpenter et al., 2008). The literature reports mixed findings on the external validity of those experiments. One strand of the literature finds that individuals behave in donation experiments similar as in naturally occurring decision situations on charitable giving (Benz and Meier, 2008; Franzen and Pointner, 2013). Other research contradicts these findings, as recently summarized by Galizzi and Navarro-Martinez (2019).

Concerning the level of altruism exhibited by men and women, a wide range of studies shows that women are generally more generous in dictator games. See, e.g., Bilén et al. (2021) for an up-to-date meta-analysis of the existing literature on gender differences in charitable giving. These authors report that the magnitude of the gender differences in altruism is sensible to the experimental context. For example, the difference is more prominent if the dictator decides to donate to a charity than giving to an anonymous recipient. However, the difference is more minor if the dictator chooses between giving all or nothing compared to deciding on a continuous scale.

Turning to studies attempting to link hormones to altruism causally, Buskens et al. (2016) and Zak et al. (2009) found no impact of administered testosterone on dictators' giving. Zethraeus et al. (2009) used another approach and administered testosterone, estrogen, or a placebo to the experimental participants. Again, no connection between either hormone and altruism is reported. Moreover, administering an oral contraceptive containing synthetic progesterone as the main ingredient suggests no hormonal impact on altruism levels. However,

there is evidence that the underlying genes influence altruism. See, for example, Reuter et al. (2011) who used twins for their study.

Multiple studies explore if different priming influences altruistic behavior. For example, subsequent donations are affected by religious primes (Ahmed and Salas, 2011; Benjamin et al., 2016; McKay et al., 2011; Shariff and Norenzayan, 2007), by reminding subjects of secular, moral institutions (Shariff and Norenzayan, 2007), and by priming with subtle cues of observability (Bateson et al., 2006; Haley and Fessler, 2005; Rigdon et al., 2009). Boschini et al. (2018) report an increased gender gap in altruism when making gender more salient by requiring participants to specify their gender before the dictator game and informing them about the gender of the recipient. Again, we have found no published studies analyzing altruism of LGBTQ+ individuals in economics.

3 Experimental design

To test our research questions, we set up an economic online experiment. This experiment received ethical approval from the UEBS Research Ethics Committee of the University of Exeter (Ethics application - eUEBS004241; 26.05.2021) and the Ethics Committee of the University of Regensburg (28.04.2021). All research was performed in accordance with the relevant guidelines and regulations. We have obtained informed consent from all participants. The study was preregistered on AsPredicted.org (Nr. 68888) before data collection (see https://aspredicted.org/rc9vn.pdf). We conduct our study (tasks and questionnaires) with oTree (Chen et al., 2016) on Prolific (www.prolific.co). To recruit the different subject groups, we used specific filters provided by Prolific. Prolific was especially well suited to host our study as they have a pool of subjects who registered as being either a transman or a transwoman. We used the Prolific filters on gender identity to recruit our subjects. However, our classification into the subject groups cismen, ciswomen, transmen, and transwomen is based on the self-reported information we elicited with the experimental questionnaire.

Each participant completes six parts and several questionnaires. One part is randomly selected for payment at the end of the experiment. In Part 1, a participant is randomly assigned to either the baseline treatment (NEUTRAL) or a treatment condition that refers to one of the gender priming interventions: FEMININE (primes a feminine gender identity) or MASCULINE (primes a masculine gender identity). Participants are primed by a word search task where different words are used depending on the underlying treatment (Bargh et al., 2001). The words in FEMININE are: female, woman, she, women, her, girl, hers, lady; in MASCULINE, they are: male, man, he, men, him, boy, his, gentleman. In the baseline condition NEUTRAL, participants also solve the word search task with the following (neutral) words: person, it, people, its, child, theirs, individual, neuter. Participants are shown the words and have two minutes to mark these words in a 10×10 grid. In case they find all words, they receive £5.

After the word search task, each participant enters the next parts of the experiments, which are the respective economic decision–making parts. As our first decision dimension, we employ monetary incentives to measure competitiveness (Buser et al., 2021). We measure the

performance in a real effort math task, where the participants are instructed to solve puzzles by finding two two-digit numbers that add up to 100 in 3×3 matrices for two minutes. In Part 2, they complete the math task under piece-rate incentives, which means they receive £0.50 for every solved puzzle. In Part 3, the same math task is performed under tournament incentives. The participants are divided into groups of four and receive £2 for every solved puzzle, but only if they solve more puzzles than every other group member. In Part 4, the participants have to choose, before performing, whether their performance in this part will be paid based on the piece-rate incentives (like Part 2) or according to the tournament rules (like Part 3). Whenever a participant decides on the tournament incentives in Part 4, s/he is classified as competitive and competes against the group member's performance in the previous part, Part 3. In all parts, the participants do not receive feedback on how well they perform compared to the other group members until the end of the experiment and have no information on the other group members' identity or characteristics. Additionally, we measure the participants' confidence in Part 2 (how well they think they performed compared to the other participants in the session) and Part 3 (how well they think they performed compared to the other group members) with incentivized questions.

Our second decision dimension is the willingness to take risks in Part 5. It is measured using a simple lottery task (Gneezy and Potters, 1997). Participants receive £4 and can invest in a lottery with a 50% chance of success. The invested amount is multiplied by 2.5 in case of success. In case of no success, the invested amount is lost. The participants keep the amount not invested. Risk preferences are measured as the amount a participant invests, where higher investments indicate a higher willingness to take risks. The third decision dimension is altruism in Part 6. We investigate the participants' altruistic preferences with a dictator game (Kahneman et al., 1986). Participants receive £5 and split this amount between themselves and up to five different charities. Altruism is quantified as the sum donated by a participant.

The post-experimental questionnaire contains (1) a 30-items version of the Bem Sex Role Inventory (BEM) that explores a person's masculine and feminine self-identification on a continuous scale (Geldenhuys and Bosch, 2020); (2) the Transgender Congruence Scale (TCS) (Kozee et al., 2012) which evaluates if and how much someone identifies as transgender; (3) demographic questions, as well as questions on the biological sex, gender, sexual orientation, and whether one self-identifies as transgender; and (4) the Steps to Transition (STT) questionnaire that describes typical steps transgender people undertake in their transition (Kozee et al., 2012). This questionnaire controls for aspects like legally changing a name, undergoing hormone replacement therapy, having surgery to alter genitalia, or a non-genital surgery like breast removal. In addition, we include debriefing questions to check if the participants are aware of the study topic and the priming intervention (Chartrand and Bargh, 1996).

Section C.14.3 in the Appendix provides a detailed description of all instructions and questionnaires (including the screenshots) of our experiment.

4 Results

Presenting our results, we use the following abbreviations: Brown-Forsythe test (BF), Chi-squared test (χ^2), Kruskal-Wallis test (KW), two-tailed Kendall's rank correlation coefficients test (KTAU), two-tailed Mann-Whitney U test (MWU), Robust Wald test (W), two-tailed Variance Ratio test (VR), Cohen's d (d), and standard deviation (SD). The significance levels are defined as follows: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), where a significant result must have at least $p < 0.05$. We summarize multiple p–values by $p's$.

4.1 Summary statistics

We collected a total of $n = 780$ observations, out of which 425 are cisgender (214 cismen and 211 ciswomen) and 355 transgender (215 transmen and 140 transwomen; see the subsection C.14.1 for more details). The questionnaire is used to classify one subject into one of the four groups, which asks about their current gender, sex, and whether they self-identify as transgender. We generally find support for the classification into groups according to the guidelines of the American Psychological Association (2015), as the data indicate that only 5.07% of transgender individuals in our sample report that their sex changed since birth.

We did a pre-experimental power analysis to calculate the needed sample sizes based on existing work by Balafoutas et al. (2018). We used their neutral priming condition to inform our power calculations. Based on their effect size delta of -0.264, the needed observations for $\alpha = 0.05$ and a power of 0.80 are 44 for one subject group in one treatment. Following, it would be enough to have in total $n = 528$. To be more conservative, we preregistered having 72 observations for each subject group in each treatment, resulting in a power of 0.95 (see https://aspredicted.org/rc9vn.pdf for further information). In our particular case of having a non-usual subject group of transgender individuals, we already mentioned in the preregistration that having 72 observations in each treatment is very ambitious for transgender individuals, also because of the number of registered transgender individuals on Prolific. We ended up in NEUTRAL with the minimum needed amount of 44 transwomen. Consequently, we had a priori, based on the ex-ante preregistered power calculations and depending on the underlying comparison, at least a power of 0.82. The power increases up to 0.95 for the subject groups with $n = 72$ in one treatment.

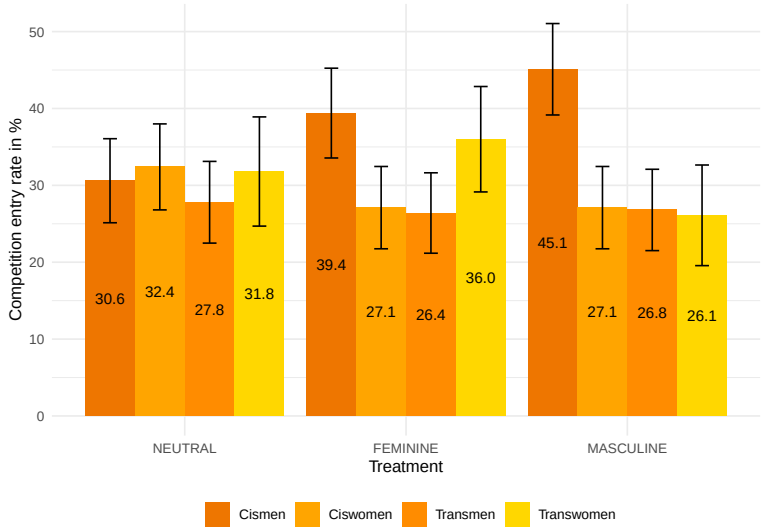
As summarized in Table C.1, the participants are on average 24.4 years old (SD = 6.60), have an average height in centimeters of 170 (SD = 10.8), and approximately half of them are students (47.2%). Around one-third holds a university degree, 69.4% have an income lower than £20,000, and 25.8% report being religious. Our sample consists mostly of participants from the United States, followed by Continental Europe and the United Kingdom. Less than 10% live outside these three mentioned regions. Responses to the BEM classify 28.5% as feminine, 19.4% as masculine, 24.1% as androgynous, and 28.1% as undifferentiated. On the TCS scale ranging from 1 to 5, participants show an average score of 3.67 (SD = 1.1). The average score on the STT, which ranges from 0 to 16, is 4.35 (SD = 4.6). The various subject groups are comparable in several characteristics as indicated by the statistical tests added in

Table C.1. Descriptive statistics broken down by subject groups are presented in Table C.2 and Table C.3 (cisgender) as well as Table C.4 and Table C.5 (transgender).

For the outcomes of Part 1, the section C.2 includes the detailed summarizing descriptives on the participants' priming. On average, the participants marked 7.45 out of 8 words (SD = 1.53), and 83.97% (i.e., $n = 655$) marked all words from the list within the given time of two minutes.

4.2 Competitiveness

Figure 9 and Table C.14 summarize the tournament entry rates in Part 4. In order to investigate whether gender and competitiveness are correlated, we focus on the baseline treatment NEUTRAL. No significant variation is reported across the four subject groups ($\chi^2(3) = 0.408$, $p = 0.939$). Similarly, when pooling the results by gender (Figure C.2; cismen + transmen vs. ciswomen + transwomen), tournament entry rates do not differ for feminine and masculine subjects ($\chi^2(1) = 0.273$, $p = 0.601$) and also no difference is found for male and female subjects when pooling the data by sex (Figure C.3; cismen + transwomen vs. ciswomen + transmen; $\chi^2(1) = 0.028$, $p = 0.867$). We compare the differences between the priming conditions (FEMININE and MASCULINE) and the baseline treatment (NEUTRAL) for the causal analysis. Priming does not influence the competition entry rates for any subject group ($\chi^2(1)$, $p's > 0.265$), including for cismen when comparing the MASCULINE treatment to the NEUTRAL treatment ($\chi^2(1)$, $p = 0.073$). We shall see in the regression analysis that when adding further controls, the impact of MASCULINE priming on cismen becomes significant. Looking at the MASCULINE priming condition only, where the entry rates look very similar for all subject groups except for cismen, the competition entry rate is around 20 percentage points higher for cismen than for all other subject groups ($\chi^2(3) = 7.991$, $p = 0.046$).



Note: The bars show the percentage of participants (between 0 and 100) who chose to compete rather than to perform under piece–rate incentives. The error bars represent the standard errors of the means.

Figure 9: Tournament entry rates in Part 4 by treatment and subject groups in alphabetical order ($n = 780$).

In Table C.15, we run Probit regressions for the baseline treatment (NEUTRAL) to disentangle the effects of gender and sex. As our basic regression framework, we have in column (1) just the subject groups and in (2) additionally controls for the performance measures in the real effort task. In column (3), we further take into account the participant's confidence and willingness to take risks. In column (4), we add the variables age, height, student status, income, religion, and residence, whereas in (5), we control for the outcomes in the TCS and STT. The TCS is interesting in our setting as it accounts for how much individuals feel genuine, authentic, and comfortable with their gender identity and external appearance. Similarly important, the STT measures details about the transition process, especially biological aspects like whether one has had surgery to alter genitalia, a non-genital surgery (like breast removal), or is undergoing hormone replacement therapy. Using joint coefficient tests (see Table C.15), we find neither gender (W, $\chi^2(1)$, $p's > 0.437$) nor sex (W, $\chi^2(1)$, $p's > 0.214$) to have a significant effect on competitiveness. We thus conclude that there is no correlation between neither gender nor sex and competitiveness in our study.

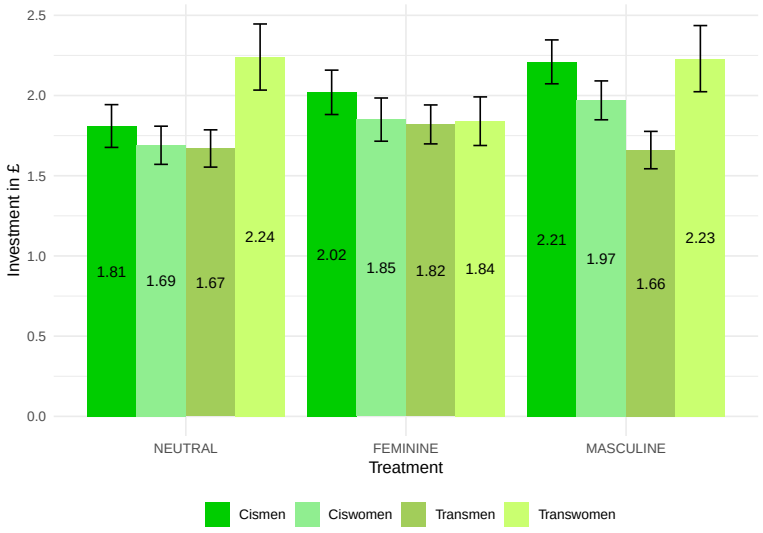
To analyze a potential causal effect of gender, we run Probit regressions in Table C.16. The non–parametrized analyses are confirmed for ciswomen, transmen, and transwomen. For cismen, we find that the gender prime with MASCULINE has a significant impact increasing the competition entry rates in specification (2) (coef = 0.473; 95% CI = 0.036, 0.909; $p = 0.034$; controlling for performance) and (4) (coef = 0.544; 95% CI = 0.076, 1.012; $p = 0.021$; controlling for beliefs, risk attitude, and other person-specific covariates). Summing up, only cismen's competition entry rates seem to be influenced (positively) when priming them with their own gender identity. We do not find a significant impact of gender priming for all other subject groups and priming combinations. We will interpret those results in Section 6.

Our experimental design does not only allow us to look into the choice to enter a tournament but also into participants' confidence (i.e., how well they believe they performed in the real effort task when competing, see Table C.11). In NEUTRAL, there is no evidence that subjects of masculine gender have higher performance beliefs than subjects of feminine gender (MWU, $z = -0.912$, $p = 0.362$). However, we do find differences between subjects of female and male sex (MWU, $z = -3.470$, $p = 0.001$). For priming, no subject group increases or decreases their beliefs when being primed (MWU, $p's > 0.177$). Regressions in Table C.12 confirm that beliefs depend on the participants' sex: male subjects generally have higher confidence in their performance than female subjects (W, $F(1)$, $p's < 0.001$). And again, confidence does not differ across gender (W, $F(1)$, $p's > 0.259$). That gender does not play a role in this setting is further confirmed when looking at the causal impact of gender priming on the participants' confidence. For none of the subject groups, we do find any effect of gender priming on the beliefs when using regression analyses (see Table C.13, W, $F(1)$, $p's > 0.178$).

Another interesting aspect is to see in how far behavior pays off in the competitiveness task. We provide details and various analyses of the performances in the real effort task and the related payoffs of Part 2 to 4 in Section C.3.

4.3 Risk

Investment rates in the lottery are depicted in Figure 10 and stated in Table C.20. When applying non–parametric tests, we do not find any differences between the various subject groups within the baseline treatment NEUTRAL (KW, $\chi^2(3)$ with ties = 4.712, $p = 0.194$). If anything, transwomen seem to be more risk-taking than transmen in a pairwise comparison (MWU, $z = -1.979$, $p = 0.048$). This, however, does not point towards a systematic impact of gender and/or sex when pooling data (Figure C.4 and Figure C.5; gender: cismen + transmen vs. ciswomen + transwomen, sex: cismen + transwomen vs. ciswomen + transmen; MWU, $p's > 0.130$). Turning to the causal impact of priming, again, we see MASCULINE priming increases the risk attitude for cismen only (MWU, $z = 2.075$ $p = 0.038$), bringing the level of cismen to the one of transwomen in the MASCULINE priming (MWU, $z = 0.156$, $p = 0.876$). For every other subject group, we do not find any significant impact of gender priming (MWU, $p > 0.206$).



Note: The bars show the average investment rate, and the error bars represent the standard errors of the means.

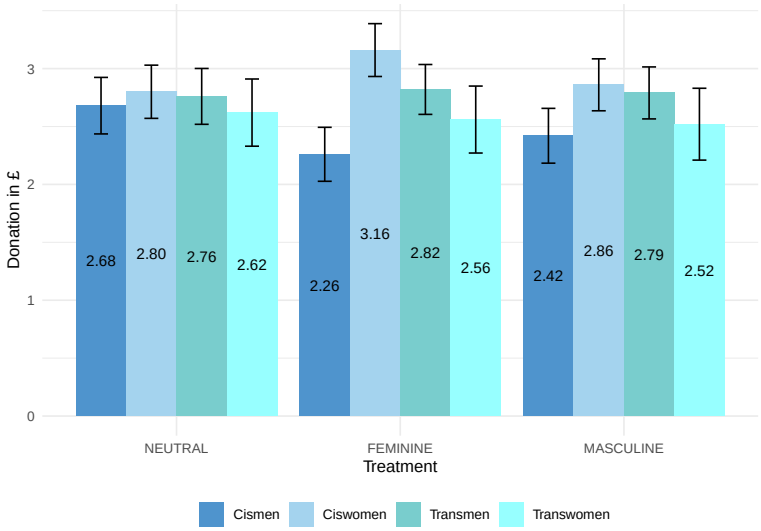
Figure 10: Investments into the risky lottery in Part 5 by treatment and subject groups in alphabetical order ($n = 780$).

Joint coefficient tests for the regressions (with and without control variables) in Table C.21 show the correlational results for our baseline condition. We find no differences in risk-taking of subjects of feminine and masculine gender (W, $F(1)$, $p's > 0.132$). However, we find a sex effect: male subjects risk more than female subjects (W, $F(1)$, $p's < 0.042$).

Turning to priming, we have significant differences in risk-taking of cismen when being primed MASCULINE (W, $F(1)$, $p's < 0.046$; see Table C.22). We find no difference in risk-taking for all other subject groups when primed with a gender (W, $F(1)$, $p's > 0.092$). The findings are independent of what other control variables are taken into account. The regression analysis for risk attitudes is thus similar to what we found for competition entry rates. When being primed with their own gender, only cismen significantly increase their risk-taking behavior.

4.4 Altruism

Last, we test for differences in the donation task (see Figure 11 and Table C.24). Donations in NEUTRAL are not distinguishable across subject group (KW, $\chi^2(1)$ with ties = 0.434, $p = 0.933$). Neither pooled results for gender nor for sex yield a difference in donation rates (Figure C.6 and Figure C.7; MWU, $p's > 0.564$). Concerning the causal impact of gender priming, we do not find significant effects for any subject group or priming condition (MWU, $p's > 0.260$).



Note: The average donations are indicated by the bars, and the error bars represent the standard errors of the means.

Figure 11: Donation in Part 6 by treatment and subject groups in alphabetical order ($n = 780$).

The regression analyses in Table C.25 and Table C.26 confirm these findings. Joint coefficient tests for gender or sex do not show significant correlations in the baseline condition (W $F(1)$, $p's > 0.580$). Moreover, the impact of all priming conditions on all subject groups remains insignificant, even after controlling for different sets of additional personal covariates (W, $F(1)$, $p's > 0.214$).

To summarize, we find no correlation between gender or sex on altruism and do not detect any causal impact of gender priming on altruistic behavior in our setup.

4.5 Gender and sex differences within priming conditions

As we have shown so far, there is no systematic correlation between gender and behavior in the NEUTRAL treatment. Here we briefly test for gender and sex differences in behavior within the two other priming treatments. Looking at Figure C.2 to Figure C.7 and analyzing the gender differences with non-parametric tests, we see no difference in competition entry rates across subject groups (FEMININE: $\chi^2(1) = 0.124$, $p = 0.725$, MASCULINE: $\chi^2(1) = 2.488$, $p = 0.115$), risk-taking (FEMININE: MWU, $z = 0.584$ $p = 0.560$, MASCULINE: MWU, $z = -0.663$, $p = 0.507$), and altruism (FEMININE: MWU, $z = -1.507$, $p = 0.132$, MASCULINE: MWU, $z = -0.625$, $p = 0.532$). Turning to sex differences, the picture slightly changes. First, we see differences between subjects of male and female sex in both priming

conditions (FEMININE and MASCULINE) for competitiveness. The differences are close to conventional levels of significance (FEMININE: $\chi^2(1) = 3.808$, $p = 0.051$, MASCULINE: $\chi^2(1) = 3.349$, $p = 0.067$). Second, for risk-taking, we find a significant difference in the MASCULINE treatment only, with subjects of male sex taking more risk than subjects of female sex (MWU, $z = 2.558$ $p = 0.011$). Third, for altruism, we find subjects of female sex having significantly higher scores than those of male sex in the FEMININE treatment (MWU, $z = -2.269$, $p = 0.023$). Hence, for risk and altruism, we find that only those sexes show higher scores who are primed with the gender identity that they would cisgender-stereotypically be associated with.

5 Robustness tests

In the remainder of the article, we apply different approaches to test the robustness of our results for comparing behavior across subject groups within NEUTRAL and by subject groups across primings.

5.1 Comparing variances instead of means

Recent literature argues that gender differences, for example, in preferences, often remain undetected because the researchers almost exclusively focus on differences in means (o'Dea et al., 2018; Thöni and Volk, 2021). It is suggested that when comparing variance ratios (i.e., the standard effect size measure for variance differences), one reliably finds evidence for greater male variability in cooperation, time, risk, social preferences, and academic grades. Thus, we rerun our analysis based on variance ratios for risk and altruism only, given that competitiveness is measured on a binary scale.

No significant differences in standard deviations of all subject groups within the baseline treatment NEUTRAL (BF(3,255), $W50 = 2.564$, $p = 0.055$) are found for the lottery investment rates. Pooling the results for gender does again show no differences in the variances (VR(143, 114), $f = 0.805$, $p = 0.219$). Only the investment rate of male subjects has a greater variability compared to females when pooling data based on sex (VR(115, 142), $f = 1.5617$, $p = 0.012$). This result is in line with a recent meta-analysis, finding a significant difference in variances between men and women of 1.25 (Thöni and Volk, 2021). Additionally, no causal impact of gender priming between any priming condition for any subject group (VR, $p's > 0.100$) is reported.

The variances of the donations in NEUTRAL are not distinguishable across subject groups (BF(3, 255), $W50 = 1.100$, $p = 0.350$). The literature reports a variance ratio between men and women of 1.18 (Thöni and Volk, 2021), which is in line with the variance ratio in our sample of 1.144 between cismen and ciswomen. Neither pooled results for gender nor sex show significant differences in the variances of donation rates (VR, $p's > 0.480$). Similarly, the donation rates do not differ based on the variances for any subject group when comparing the different priming conditions (VR, $p's > 0.343$).

5.2 Using Cohen's d

Cohen's d can be used with the p-value from a common t-test to illustrate if an effect size is not only significant but if a significant result is also relevant. One restriction of this approach is that it is only possible to conduct it for pairwise comparisons, which is not fully in line with the main analyses we provide in the Results section. Moreover, t-tests and their p-values are generally presented together with the Cohen's d. The p-values tell if the effect is statistically significant, whereas the Cohen's ds determine the effect size. However, t-tests are usually applied to normally distributed data or in case a dataset is considered to be very large. Nevertheless, we believe that discussing Cohen's ds adds another valuable robustness test for our results. We consider an effect to be (i) small, when the absolute Cohen's d is smaller than 0.2, (ii) medium for absolute Cohen's d between 0.2 and 0.5, and (iii) large if the absolute Cohen's d is larger than 0.5. In the following, we discuss the Cohen's d statistics and add the p-values from respective t-tests only for those that report at least a medium Cohen's d.

Table C.17 summarizes the Cohen's d analyses for competitiveness. When comparing the subject groups in NEUTRAL, we find only small effects ($d \in [0.012, 0.101]$). The same is true when pooling by gender or sex in NEUTRAL ($d \in [0.021, 0.065]$). The effects sizes for comparing all four subject groups separately between NEUTRAL and FEMININE ($d \in [0.031, 0.187]$) and NEUTRAL and MASCULINE ($d \in [0.023, 0.115]$) are again small. The only exception are cismen, where the difference between NEUTRAL and MASCULINE becomes medium ($d = 0.303$) but is insignificant ($p = 0.072$).

The analyses for risk can be found in Table C.23. Within NEUTRAL, the effect sizes of comparing cismen or ciswomen with transwomen is medium ($d = 0.354, 0.484$) and only significant for the latter comparison ($p = 0.011$). Besides, the Cohen's d is getting large and significant for transmen vs. transwomen ($d = 0.504$; $p = 0.008$). For all other comparisons, the effects are small between subject groups ($d \in [0.017, 0.134]$). Pooling by sex reveals a medium, significant effect size ($d = 0.268$; $p = 0.032$), while the effect size for the gender-wise comparison is small ($d = 0.143$). The effects sizes for each subject group when looking at NEUTRAL vs. FEMININE are small ($d \in [0.143, 0.182]$), except for the medium insignificant one of transwomen ($d = 0.335$; $p = 0.107$). For NEUTRAL vs. MASCULINE, cismen show a medium and significant effect size ($d = 0.348$; $p = 0.039$), whereas all other subject groups have small or medium, but insignificant Cohen's ds ($d \in [0.013, 0.282]$; $p > 0.097$).

The effect sizes for the participants' donations are listed in Table C.27. They are small and insignificant within NEUTRAL when comparing by sex, gender, or between subject groups ($d \in [0.004, 0.098]$; $p > 0.600$). Similarly, the effect sizes for all other comparisons considering the different treatments are small and lack significance ($d \in [0.016, 0.188]$; $p > 0.267$). The only slightly medium and insignificant exception ($d = 0.210$: $p = 0.212$) is reported for cismen in NEUTRAL vs. FEMININE.

5.3 Using a continuous instead of a categorical gender measure

With just a handful of exceptions (Brenøe et al., 2022; Kastlunger et al., 2010; Lemaster and Strough, 2014; Meier-Pesti and Penz, 2008), researchers in economics always used a categor-

ical way to measure gender. However, it is more and more discussed that gender might be a continuous characteristic rather than a binary (or categorical) one (Hyde et al., 2019). Techniques accounting for it include asking different questions (Bittner and Goodyear-Grant, 2017) or using identity status concerning adherence to actual gender role beliefs (McDermott et al., 2022). Another method is the BEM sex role inventory (Geldenhuys and Bosch, 2020). It provides a continuous gender scale, and we conducted it in the post-experimental questionnaire. The BEM is a very accurate predictor for gender and is highly correlated with other continuous gender measures and single-item measures (Brenøe et al., 2022).

We rerun all regression analyses and include, instead of the subject groups, the variables $BEMscore : Feminine$ (defined as the score participants reached on the BEM questions measuring femininity) and $BEMscore : Masculine$ (score on masculine questions in the BEM). Results in Table C.28, Table C.30, and Table C.32 show throughout that neither the feminine nor the masculine score significantly influence how the participants decide in NEUTRAL (W, $p's > 0.057$). This is not surprising since the BEM scores and the gender categories are highly correlated (feminine: KTAU, Kendall's score = 21692, $p = 0.001$, masculine: KTAU, Kendall's score = -18485, $p = 0.003$), and we did not find correlational gender differences in the baseline condition for neither of the economic decisions we investigate.

Also, for the causal impact of gender priming, no evidence is found for an effect of the BEM score on behavior. Table C.29, Table C.31, and Table C.33 confirm this with the insignificant variables measuring the two BEM scores ($p's > 0.056$), the insignificant interaction terms of the priming condition with the feminine or masculine BEM score ($p's > 0.054$), and the insignificant respective joint coefficient tests (W, $\chi^2(1)/F(1)$, $p's > 0.108$).

5.4 Controlling for gender congruent upbringing

One limitation of our approach is that the subjects are sorted into distinct gender categories based on their *current* gender identity. This potentially lacks accounting for psychological, behavioral, social, and cultural experiences that shape a gender identity over time, particularly during adolescence. While we can not fully account for this confound, we can analyze if being raised according to one's current gender affects our primary outcomes.

In our post-experimental questionnaire, we asked the participants according to which gender their parents treated them. Based on the answers and the self-reported gender, we create the variable gender congruent upbringing (GCU). GCU is equal to 1 if someone was raised according to their current gender identity (or was raised neutrally) and 0 otherwise. How the participants were raised matches the currently reported gender of 32.09% of transmen and 15.00% of transwomen. For cisgender individuals, the variable CGU equals 1 for 99.76%. Due to the lack of variation of CGU for the cisgender sample, we conducted all analyses for transgender individuals only.

We rerun all main regression and include, instead of the different subject groups, the variable GCU. Results in Table C.43 and Table C.47 show that whether participants were raised according to their current gender does not significantly influence the participants' competitiveness and altruism in NEUTRAL ($p's > 0.473$). For risk, we see in Table C.45 a significantly

negative coefficient in NEUTRAL for two out of the three regression) ($p's < 0.043$). When considering the causal impact of the gender priming, there is again no evidence for an effect of being raised gender-congruent on competitiveness and altruism. Table C.44 and Table C.48 show this based on the insignificant coefficient for GCU ($p's > 0.423$) and the insignificant respective joint coefficient tests (W, $\chi^2(1)/F(1)$, $p's > 0.076$). For risk (see Table C.46), the coefficients are again significantly negative for NEUTRAL only ($p's < 0.022$) because the joint coefficient tests taking the treatments and GCU interactions into account remain insignificant (W, $F(1)$, $p's > 0.055$).

5.5 Controlling for the strength of the priming intervention

To underline the strength of our results concerning the priming, we look at the answers to the survey question "Do you remember any of the words from the word-search puzzle? If not, leave empty.", which was implemented (not incentivized) at the very end of our experiment. We use the outcome of this question to control for the strength of the priming intervention. It can be assumed that the more words a subject remembered, the more they were still primed toward the end of the study. First, 93.08% of all participants remember at least one out of the eight words. The average number of recognized words is 4.33, and 70.00% of all participants reported at least four words. Thus, it can be assumed that the prime was activated for the majority of participants throughout the experiment.

Second, we rerun the regressions in Table C.16, Table C.22, and Table C.26 for the three behavioral outcomes. The dummy variables, accounting for the different primings (i.e., the treatment variables), are replaced by $Rem.\,feminine\,words$, $Rem.\,masculine\,words$, and $Rem.\,neutral\,words$, which measure the number of words remembered in each treatment. The only significant and close to significant results found are that cismen in MASCULINE are investing more into the lottery, the more masculine words they remember (see Table C.35; $p's < 0.014$) and ciswomen in FEMININE are donating more, the more feminine words they remember (see Table C.36; $p's < 0.051$), compared to the NEUTRAL condition. Moreover, we did a subgroup analysis for those who remembered at least the median amount of priming words (i.e., four words) or less (see Table C.37 to Table C.42). Overall, when using the remembered words instead of a simple priming variable, our findings in the main results section replicate when controlling for the number of remembered priming words.

5.6 Correcting for multiple hypothesis testing

Like other scientists, we face the problem of simultaneously evaluating several hypotheses. Conducting multiple comparisons increases the likelihood that a non-negligible proportion of tests are false positives. Thus, drawing valid conclusions requires considering the number of performed statistical tests and adjusting the statistical confidence measures accordingly. We employ the free online tool "Multiple Testing Correction" by Menyhart et al. (2021), available at www.multipletesting.com.

As we perform with our novel pool of transgender individuals a mix of exploratory and confirmatory analysis, the suitable methods for correction are Bonferroni, the Holm (step-

down) approach, and the Hochberg (step-up) correction which allows for calculating False Discovery Rates (FDR). According to the Multiple Testing Correction, the first significant p-value (values over these thresholds are not considered significant) is $p = 0.0015$, independent of the method used. So if we – instead of the significance levels explained in the results section – define a result as significant if it is at least $p < 0.0015$, all results (non-parametric and findings from regressions) turn out to be insignificant.

6 Discussion and conclusion

This paper applies well-known and extensively used experimental techniques to identify the influence of gender and sex on economic decision-making. First, we separate the impact of gender and sex on economic decisions by collecting data from participants whose gender and sex differ, which is new to the literature. We compare the competitive, risk, and altruistic behavior of four different subject groups - cismen, ciswomen, transmen, and transwomen. Second, we induce either a neutral, feminine, or masculine gender identity by having different priming conditions. Thus, with our experimental setup, we go beyond correlating gender and sex with decisions and try to evoke gender identities through a priming manipulation causally.

While this study is pre-registered and carefully designed following existing literature and the state of the art standards in experimental economics, the findings diverge from previous work. Our results do not show conclusive correlational or causal evidence for gender or sex as determinants of economic decision-making. As described in the main results section, we find just a handful of significant results. These results do generally not replicate when applying different robustness tests, including accounting for multiple hypothesis testing. Thus, the pattern is essentially consistent: gender and sex differences in behavior remain statistically indistinguishable. Besides, we see that cis- and transgender participants do not systematically differ from each other in their behavior. Our overall interpretation of the data is that gender and sex might not matter as much as initially thought. But what can explain these findings?

First, one explanation could be that gender effects might depend on the underlying subject pool. The existing literature has treated gender differences in behavior as a well-established and robust finding. However, the vast majority of these papers use standard student subjects (Marianne, 2011). Studies that use other samples (e.g., Charness and Villeval, 2009) or online samples are generally less likely to report gender differences, especially when controlling for a set of participants' characteristics (Flory et al., 2018; Almås et al., 2016*b*). Moreover, differences in sample size are likely to play a role. We pre-registered a sample that would give us enough statistical power based on existing literature. Still, it remains true that small gender differences in behavior may lie below our minimum detectable effect sizes. The total sample size in this experiment has been constrained by the availability of transgender individuals on Prolific. However, we expect the availability of transgender individuals for future studies to increase, hence allowing for replications of our findings with larger sample sizes.

Second, almost two decades have passed since the first studies that looked into competitiveness, risk, and altruism were published and found gender differences in behavior. One can

thus speculate that female empowerment, educational initiatives, and the broader awareness of gender and sex equality in private and professional settings have led to a narrowing of potential behavioral differences in the meantime.

Third, the absence of an effect of gender priming on the behavior of transgender subjects may be rooted in the connotation those subject groups have with gender. For transgender individuals, the concept of gender might be a relatively continuous spectrum, whereas, for cis-individuals, it might be seen as a binary dimension. As such, gender might not be as decisive for transgender as for cisgender individuals. The fact that gender priming seems to work only for cismen but not for ciswomen might hinge on the role gender usually has played for those two subject groups. Whereas for cismen, their gender usually comes with advantages and, as such, has a positive connotation, ciswomen might have negative experiences concerning the way society treats them based on their gender.

Despite the partly unexpected findings, we believe that there are several key "takeaways" from this study. For the first time, we present evidence from a sample of cis- and transgender participants in one framework, which allows for both a correlational and a causal approach and look at how they decide in a competitive context and when making risky or altruistic decisions. Transgender individuals have become a more and more visible part of society. Thus, we think it is crucial to understand their economic behavior. Furthermore, having transgender participants in our sample makes it possible to look deeper into the part that an individual's gender - as opposed to sex - plays in economic decision-making. In our setting, we shed light on the part of gender effects that can be attributed to biological factors (which refer to a participant's sex) and other aspects of one's gender identity. Additionally, we do not measure gender only on a categorical scale; instead, we also apply a continuous gender scale. Our results are qualitatively the same, independent of what gender scale is used. Besides, we use different statistical techniques to analyze our data, which overall point towards the same interpretation of our results. Moreover, we test for the first time if upbringing according to the current gender influences the behavior of transgender individuals. We found that gender-congruent upbringing makes transgender individuals more risk-averse only in the neutral priming condition. For this result, we encourage future research to look into the explanations of this outcome, which would go beyond the original scope of this paper.

Based on our findings, we conclude that the role of gender and sex is not as decisive for economic behavior as previously assumed.

Chapter V

Conclusion

In this dissertation, I presented three online experiments in behavioral economics that explore individual decision-making in the field of pro-environmental behavior and gender differences. The first two projects investigate what encourages individuals to translate their environmentally friendly attitude into action. The third project studies the robustness of gender differences in economic behavior. To increase external validity, all experiments were conducted online on the platform Prolific, which allows to employ a more heterogeneous subject pool than standard lab experiments. Hence, this thesis contributes to research in experimental economics with a non-student subject pool.

Chapter II explores whether environmentally friendly decisions can be encouraged by providing individuals with monetized carbon information. Monetization is gaining relevance as an increasing number of companies use it to report on their societal impact. The experiment provides empirical evidence on the effect of three types of carbon information on participants' consumption decisions. In the online experiment, participants decide on purchasing a virtual product whose carbon emissions are either reported in kilograms, abatement costs, or social costs. The experiment contributes to the literature by differentiating between monetization based on abatement cost and social cost. In contrast to Hummel and Hörisch (2020), our experiment does not find that the type of carbon information significantly affects consumption decisions. Interestingly, participants believe that social cost information is most effective and prefer receiving this type of monetized information as consumers. Irrespective of the type of carbon information presented, participants have difficulties transferring carbon information into everyday action. Regarding the project's research question on whether monetized carbon information increases pro-environmental behavior, the study reveals that neither abatement nor social cost leads to significantly different purchase decisions compared to kilogram information. The finding that all studied labels were equally effective in promoting climate-friendly action have implications for the design of eco-labels and can inform internal management decisions and corporate reporting on societal impacts. While contradicting extant findings, our results stress the need for future research. Investigating the effect of carbon metrics on decision-making in a field setting makes for a promising extension of our research.

Chapter III studies the role of moral balancing in pro-environmental decision-making. In particular, it explores whether the magnitude of the initial action determines the size of the

moral balancing effect. The incentivized online experiment finds that participants who successfully offset carbon and acquired a moral license subsequently offset less carbon than those who failed to obtain a moral license. This result suggests that moral balancing affects pro-environmental decision-making. The amount of carbon offset does not consistently affect subsequent offsetting decisions, indicating that moral balancing is rather insensitive to the absolute environmental impact. Furthermore, environmental concerns moderate moral balancing, and individuals highly concerned about global warming do not base their offsetting decisions on previous carbon offsets. The findings highlight the importance of considering moral balancing whenever targeting pro-environmental behavior, e.g., when designing environmental campaigns or evaluating the impact of voluntary carbon offset markets. To mitigate potential adverse spillover effects, campaigns should focus on promoting actions with substantial environmental effects.

Chapter IV investigates the role gender (compared to sex) plays in economic decision-making. It uses an incentive–compatible economic experiment to study the influence of gender and sex on competitiveness, risk–taking, and altruism with two approaches. First, it compares the behavior of cis- and transgender individuals using correlational analysis and discusses the importance of gender and an individual's sex. Second, it tests causally for the effect of gender by priming participants with either a feminine or masculine gender identity before they make their decisions. The findings suggest that the role of gender (and sex) is not as decisive for economic behavior as commonly found in previous work. The experiment contributes to the literature by being the first to study cis- and transgender individuals within a single experimental framework. Understanding the economic behavior of transgender individuals is essential due to their increasing visibility in society. Moreover, it allows to disentangle the impact of biological factors that refer to an individual's sex and other characteristics of an individual's gender identity on economic decision-making. The study introduces a novel approach, as gender is not solely measured categorically, and the analysis incorporates a continuous gender scale.

Overall, this dissertation explores two domains – voluntary climate actions and gender differences – for which behavioral economics contributes important insights. At the outset of tackling existing challenges, carefully crafted experiments were employed to understand the current situation and advice on possible improvements. All in all, the presented research provides interesting findings and builds a stepping stone for future research and initiatives on climate change mitigation and towards gender equality. Enhancing our knowledge of what determines individual behavior promises to have a valuable impact on designing effective policy interventions and promoting sustainability and gender equality. As the field of behavioral economics continues to evolve, it holds countless intriguing questions yet to be explored. I am excited to witness its application in addressing global challenges and shaping solutions for a better future.

Appendices

Appendix A

CO_2 Labels

A.1 Additional results

Table A.1: Average time participants spend on key sections.

Key sections	Pages	Average time (in seconds)
Instructions on virtual product	5-7	118
Instructions on product's carbon emissions	8-12	167
Purchase decision	13	14
Guessing task	14-15	50
Belief and preference elicitation	16	120
Questionnaire	18-23	132
Other	1-4, 17, 24	114

Table A.2: Logit regressions.

	Dependent variable: Purchase (0 or 1)				
	(1)	(2)	(3)	(4)	(5)
ABATEMENT COSTS	-0.298	-0.341	-0.309	-0.329	-0.368
	(0.215)	(0.232)	(0.252)	(0.253)	(0.258)
SOCIAL COSTS	-0.044	-0.017	-0.009	-0.012	-0.029
	(0.210)	(0.228)	(0.246)	(0.247)	(0.252)
Constant	-0.597***	1.969	1.696	1.774	1.780
	(0.148)	(1.494)	(1.522)	(1.525)	(1.532)
Wald test: ABATEMENT COSTS	2.210	2.671	1.884	2.071	2.454
= SOCIAL COSTS = 0 (p-value)	(0.331)	(0.263)	(0.390)	(0.355)	(0.293)
Demographics	No	Yes	Yes	Yes	Yes
Environmental awareness	No	No	Yes	Yes	Yes
Offset in the past	No	No	No	Yes	Yes
Altruism & patience	No	No	No	No	Yes
N	600	600	600	600	600

Note: The following control variables are included in Demographics: *gender, age, has children, religious, highest education level achieved, income category, job in science, job in business, political view* and in Environmental awareness: *SASSY segment, global warming caused by humans, (strongly) agrees to actions matter, often takes env. friendly action*; *** p<0.01, ** p<0.05, * p<0.1

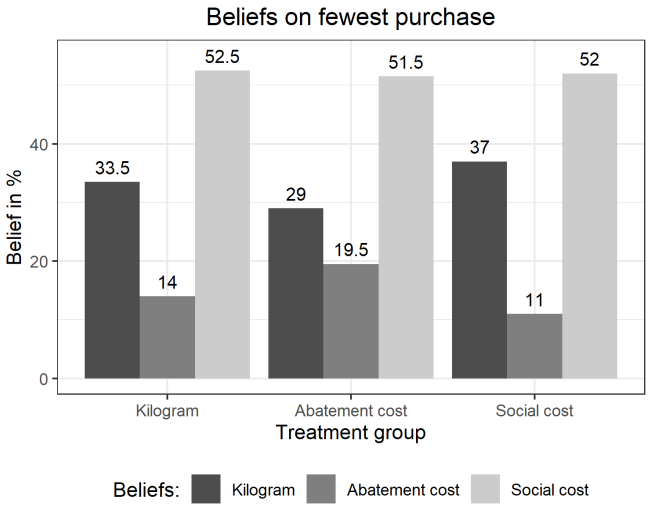


Figure A.1: Beliefs on information that leads to the lowest purchasing rate by treatment.

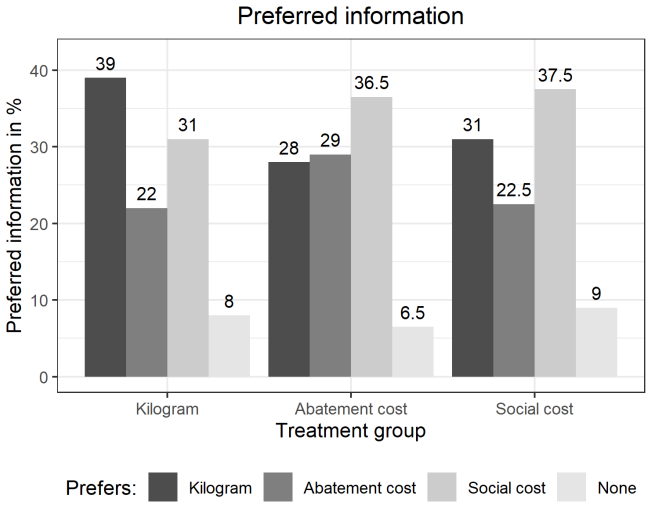


Figure A.2: Preferred information by treatment.

Table A.3: Heterogeneous effects.

	Dependent variable: Purchase (0 or 1)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Female	Left	Right	Green	Alevel	Income	Altruism	Engage	Job science	Job business
Abatement costs	-0.062	-0.076	-0.075	-0.067	-0.182	-0.046	-0.060	-0.116	-0.072	-0.038
	(0.082)	(0.069)	(0.049)	(0.047)	(0.155)	(0.056)	(0.044)	(0.073)	(0.045)	(0.048)
Social costs	-0.007	0.001	-0.038	0.002	0.0001	0.018	-0.004	0.064	-0.022	-0.003
	(0.090)	(0.068)	(0.053)	(0.049)	(0.169)	(0.056)	(0.045)	(0.080)	(0.046)	(0.049)
Interacted variable	-0.110	-0.064	-0.007	-0.021	-0.067	-0.003	-0.090**	-0.043	-0.364***	0.047
	(0.074)	(0.067)	(0.073)	(0.084)	(0.128)	(0.076)	(0.041)	(0.069)	(0.086)	(0.091)
Interacted variable ×	-0.003	0.016	0.038	0.015	0.142	-0.036	0.016	0.091	0.266**	-0.112
Abatement costs	(0.097)	(0.088)	(0.103)	(0.113)	(0.161)	(0.098)	(0.055)	(0.091)	(0.126)	(0.112)
Interacted variable ×	0.009	-0.029	0.095	-0.112	-0.0005	-0.011	0.024	-0.106	0.366**	-0.009
Social costs	(0.103)	(0.089)	(0.102)	(0.105)	(0.175)	(0.104)	(0.054)	(0.097)	(0.161)	(0.119)
Constant	0.796**	0.717**	0.685*	0.669*	0.895***	0.801**	0.799**	0.825**	0.808**	0.784**
	(0.363)	(0.347)	(0.359)	(0.356)	(0.150)	(0.354)	(0.360)	(0.390)	(0.367)	(0.353)
Wald test: Abatement costs	0.363	0.838	1.185	1.421	1.009	0.776	1.242	3.050	1.379	0.393
= Social costs = 0 (p-value)	(0.696)	(0.433)	(0.307)	(0.242)	(0.365)	(0.461)	(0.290)	(0.048)	(0.253)	(0.675)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Environmental awareness	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Offset in the past	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Altruism & patience	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	593	600	600	600	600	538	600	600	600	600
R2	0.254	0.237	0.234	0.235	0.241	0.277	0.255	0.261	0.259	0.256

Note: Robust standard errors in parentheses; *Left, Right and Green* denote dummy variables for political orientation. *Alevel* is a dummy variable that is equal to one if the individual has an a level or higher education status (specification does not include education dummies). *Income* is equal to one if the individual earns minimum £30,000. *Altruism, Job science* and *Job business* are defined as before and *Engage* abbreviates frequent engagement in environmentally friendly behavior. The following control variables are included in Demographics: *gender, age, has children, religious, highest education level achieved, income category, job in science, job in business, political view* and in Environmental awareness: *SASSY segment, global warming caused by humans, (strongly) agrees to actions matter, often takes env. friendly action*. The respective variable is not included in the set of controls if it is considered in the specification; *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Overestimation regressions.

	Dependent variable:					
	Overestimate number of bottles (0 or 1)				*Purchase (0 or 1)*	
	OLS	Logit	OLS	Logit	OLS	Logit
	(1)	(2)	(3)	(4)	(5)	(6)
ABATEMENT COSTS	-0.055**	-0.858**	-0.064**	-1.280***	-0.073*	-0.427
	(0.026)	(0.415)	(0.025)	(0.488)	(0.044)	(0.263)
SOCIAL COSTS	-0.056**	-0.863**	-0.068***	-1.426***	-0.022	-0.090
	(0.026)	(0.415)	(0.026)	(0.497)	(0.046)	(0.255)
Overestimate bottles					-0.185***	-1.672***
					(0.061)	(0.607)
Constant	0.101***	-2.192***	0.572*	1.531	0.891***	2.643
	(0.021)	(0.236)	(0.333)	(1.959)	(0.302)	(1.723)
Wald test (p-value):	2.71	6.50	4.02	11.04		
ABATEMENT COSTS	(0.067)	(0.039)	(0.019)	(0.004)		
= SOCIAL COSTS = 0						
Demographics	No	No	Yes	Yes	Yes	Yes
Environmental awareness	No	No	Yes	Yes	Yes	Yes
Offset in the past	No	No	Yes	Yes	Yes	Yes
Altruism & patience	No	No	Yes	Yes	Yes	Yes
N	598	598	598	578	598	598
R2	0.011		0.111		0.267	

Note: Robust standard errors in parentheses for OLS regressions; *Overestimate number of bottles = 1* if individual overestimated equivalent number of bottles. The following control variables are included in Demographics: *gender, age, has children, religious, highest education level achieved, income category, job in science, job in business, political view* and in Environmental awareness: *SASSY segment, global warming caused by humans, (strongly) agrees to actions matter, often takes env. friendly action*; in column (4) 20 observations are excluded as variable values predict failure perfectly; *** p<0.01, ** p<0.05, * p<0.1

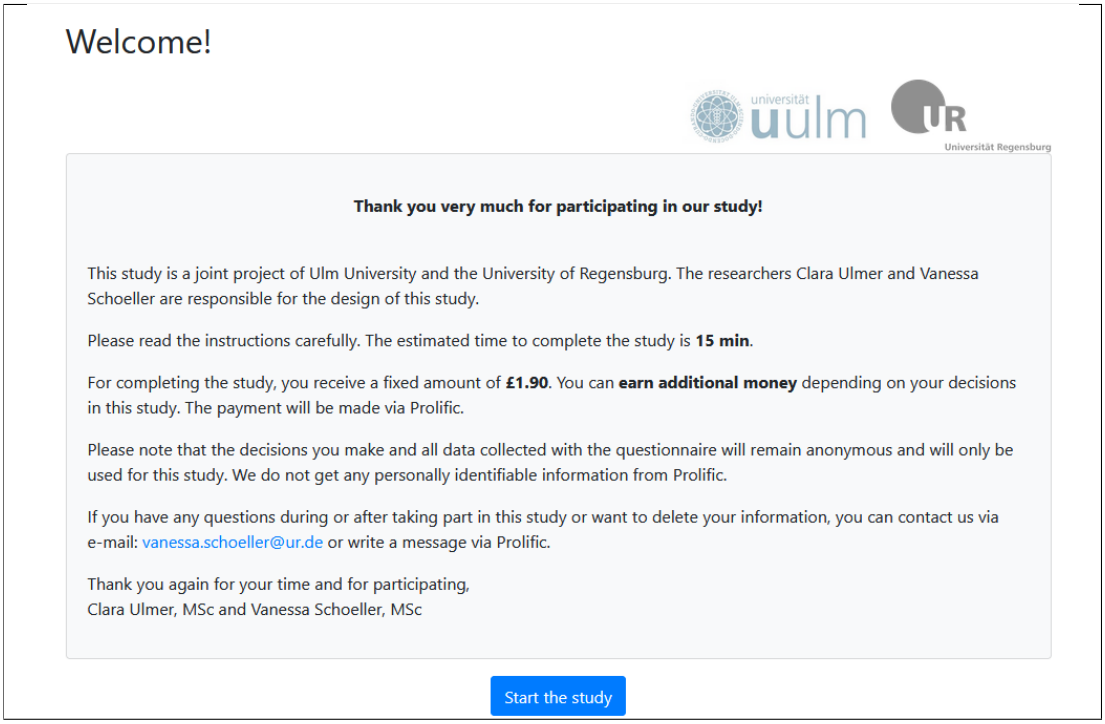
Table A.5: Guess regressions.

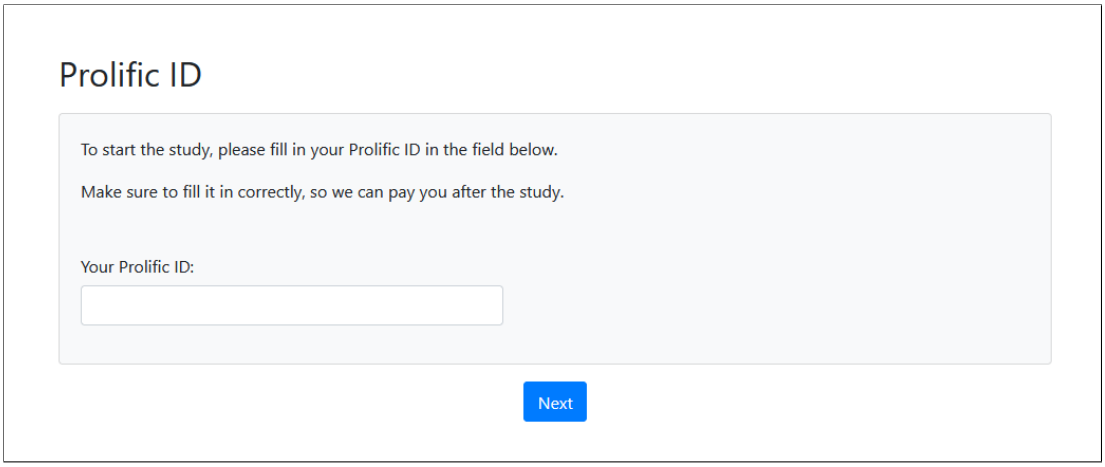
	Dependent variable: Purchase (0 or 1)			
	All		Exclude > 1000	
	OLS	Logit	OLS	Logit
	(1)	(2)	(3)	(4)
Guessed number of bottles	-9.56e-06***	-0.0003	-0.0002*	-0.002
	(2.16e-06)	(0.0003)	(8.71e-05)	(0.001)
ABATEMENT COSTS	-0.061	-0.373	-0.072	-0.466*
	(0.043)	(0.260)	(0.045)	(0.264)
SOCIAL COSTS	-0.0003	-0.026	-0.014	-0.081
	(0.045)	(0.254)	(0.047)	(0.255)
Constant	0.816**	1.966	0.850**	2.214
	(0.363)	(1.536)	(0.350)	(1.567)
Demographics	Yes	Yes	Yes	Yes
Environmental awareness	Yes	Yes	Yes	Yes
Offset in the past	Yes	Yes	Yes	Yes
Altruism & patience	Yes	Yes	Yes	Yes
N	600	600	591	591
R2	0.260		0.260	

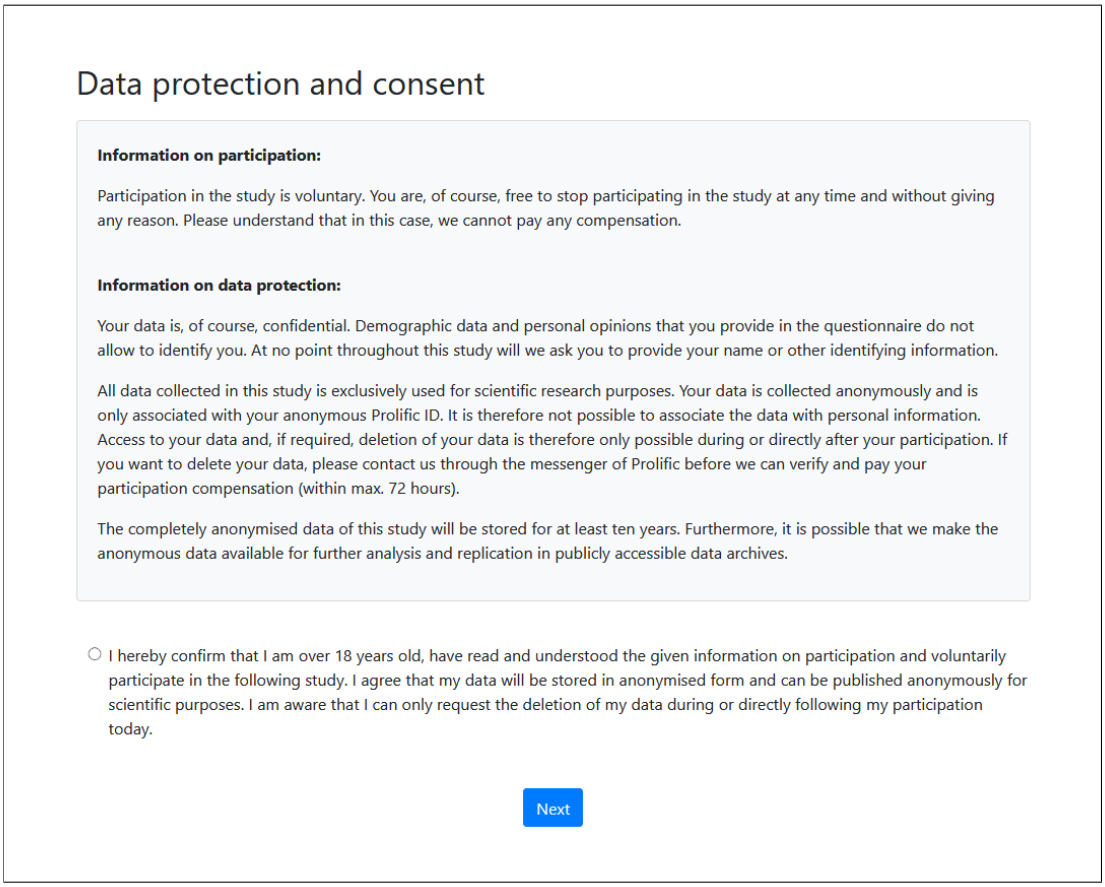
Note: Robust standard errors in parentheses for OLS regressions. In *All*, every observation is included; in *Exclude > 1000*, we exclude guesses that are larger than 1000. The following control variables are included in Demographics: *gender, age, has children, religious, highest education level achieved, income category, job in science, job in business, political view* and in Environmental awareness: *SASSY segment, global warming caused by humans, (strongly) agrees to actions matter, often takes env. friendly action*; *** p<0.01, ** p<0.05, * p<0.1

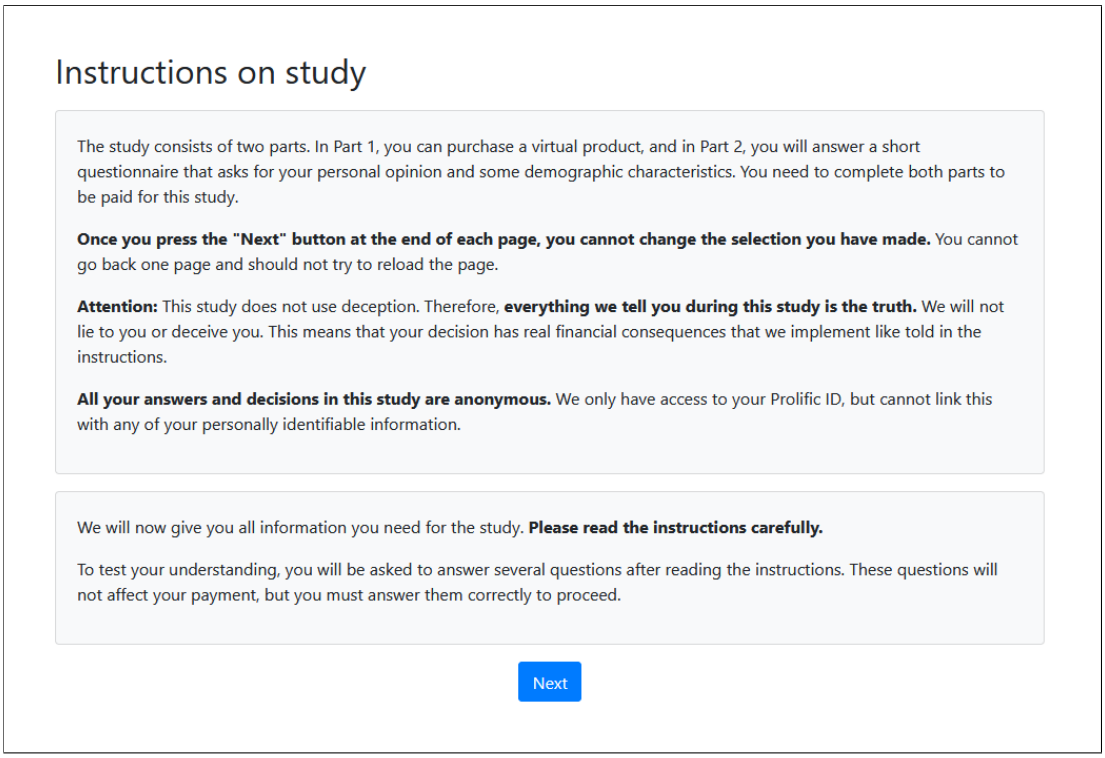
A.2 Instructions

The following pages contain screenshots of the online study conducted via Prolific. The study showed each participant partly different slides, depending on the treatment condition, the participant was allocated to randomly. Headlines stating the specific treatment mark the varying screens. All other pages were identical.






Data protection and consent
Information on participation:
Participation in the study is voluntary. You are, of course, free to stop participating in the study at any time and without giving any reason. Please understand that in this case, we cannot pay any compensation.
Information on data protection:
Your data is, of course, confidential. Demographic data and personal opinions that you provide in the questionnaire do not allow to identify you. At no point throughout this study will we ask you to provide your name or other identifying information.
All data collected in this study is exclusively used for scientific research purposes. Your data is collected anonymously and is only associated with your anonymous Prolific ID. It is therefore not possible to associate the data with personal information. Access to your data and, if required, deletion of your data is therefore only possible during or directly after your participation. If you want to delete your data, please contact us through the messenger of Prolific before we can verify and pay your participation compensation (within max. 72 hours).
The completely anonymised data of this study will be stored for at least ten years. Furthermore, it is possible that we make the anonymous data available for further analysis and replication in publicly accessible data archives.
I hereby confirm that I am over 18 years old, have read and understood the given information on participation and voluntarily participate in the following study. I agree that my data will be stored in anonymised form and can be published anonymously for scientific purposes. I am aware that I can only request the deletion of my data during or directly following my participation today.
Next

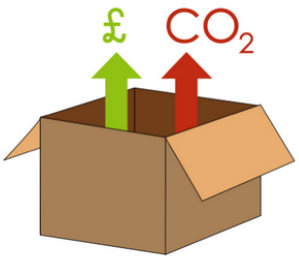

Instructions on study
The study consists of two parts. In Part 1, you can purchase a virtual product, and in Part 2, you will answer a short questionnaire that asks for your personal opinion and some demographic characteristics. You need to complete both parts to be paid for this study.
Once you press the "Next" button at the end of each page, you cannot change the selection you have made. You cannot go back one page and should not try to reload the page.
Attention: This study does not use deception. Therefore, everything we tell you during this study is the truth. We will not lie to you or deceive you. This means that your decision has real financial consequences that we implement like told in the instructions.
All your answers and decisions in this study are anonymous. We only have access to your Prolific ID, but cannot link this with any of your personally identifiable information.
We will now give you all information you need for the study. Please read the instructions carefully.
To test your understanding, you will be asked to answer several questions after reading the instructions. These questions will not affect your payment, but you must answer them correctly to proceed.
Next

Instructions on decision

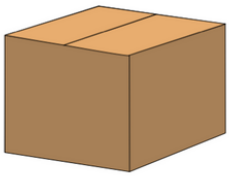
Your progress in the study: 5%

In Part 1 of this study, you have the chance to purchase a virtual product.

Virtual product means that you will not receive a real product by mail, but your purchasing decision affects your total payment and carbon dioxide (CO_2) emissions, the main driver of gobal warming.



If you buy the product, you will earn an **additional payment**, and you will emit **CO_2** into the atmosphere.



If you do not buy the product, you will **not earn an additional payment**, and you will **not emit CO_2** into the atmosphere.

In the following, first the **additional payment** and then the **CO_2 emissions** will be explained to you.

Next

Instructions on benefit, price, and additional payment

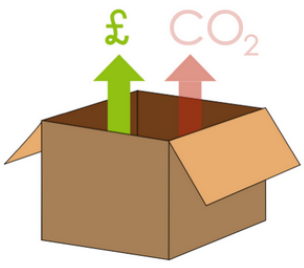
Your progress in the study: 10%

The value of the product is £2.00. The price of the product is £1.00. The additional payment is calculated as:

additional payment = value – price = £1.00

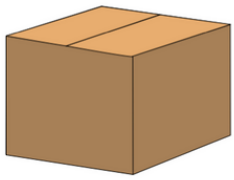
If you decide to buy the product, you earn an additional payment of £1.00.

With the payment of £1.90 for completing the study, you receive a total payment of £2.90 if you decide to buy the product.



If you decide not to buy the product, you will not earn an additional payment.

Therefore, you receive a total payment of £1.90 for completing the study if you decide not to buy the product.



Next

Fill in the missing numbers

Your progress in the study: 15%

To check your understanding, please fill in the missing numbers correctly. The information from the previous slide is given again at the bottom.

For completing the study, I will receive a fixed amount of	£	, independent of my decisions in the study.
If I buy the product, I will receive an additional payment of	£	, and I will emit CO_2 into the atmosphere.
If I do not buy the product, I will receive an additional payment of	£	, and I will not emit CO_2 into the atmosphere.

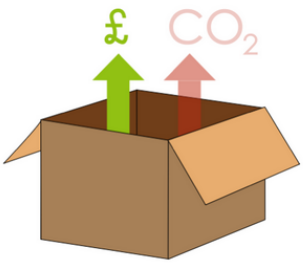
Remember:

The value of the product is £2.00. The price of the product is £1.00. The additional payment is calculated as:

additional payment = value − price = £1.00

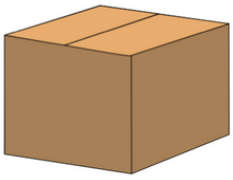
If you decide to buy the product, you earn an additional payment of £1.00.

With the payment of £1.90 for completing the study, you receive a total payment of £2.90 if you decide to buy the product.



If you decide not to buy the product, you will not earn an additional payment.

Therefore, you receive a total payment of £1.90 for completing the study if you decide not to buy the product.



Next

Instructions on carbon emissions

Your progress in the study: 20%

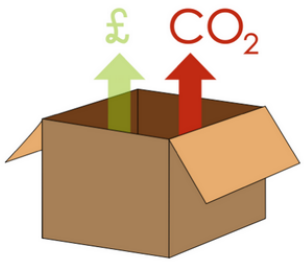
On the previous pages, the additional payment you will receive if you buy the product was explained.

However, note that if you buy the product, you will emit CO_2 into the atmosphere. Scientists agree that carbon emissions are the main driver of global warming, which will probably affect human beings by, e.g., coastal floodings due to rising sea levels.

We will now explain the carbon emissions of the product.

If you buy the product, you will emit CO_2 into the atmosphere.

Scientists agree that carbon emissions are the main driver of global warming.



If you do not buy the product, you will not emit CO_2 into the atmosphere.

Next

Attention!

Your progress in the study: 25%

These carbon emissions are real!

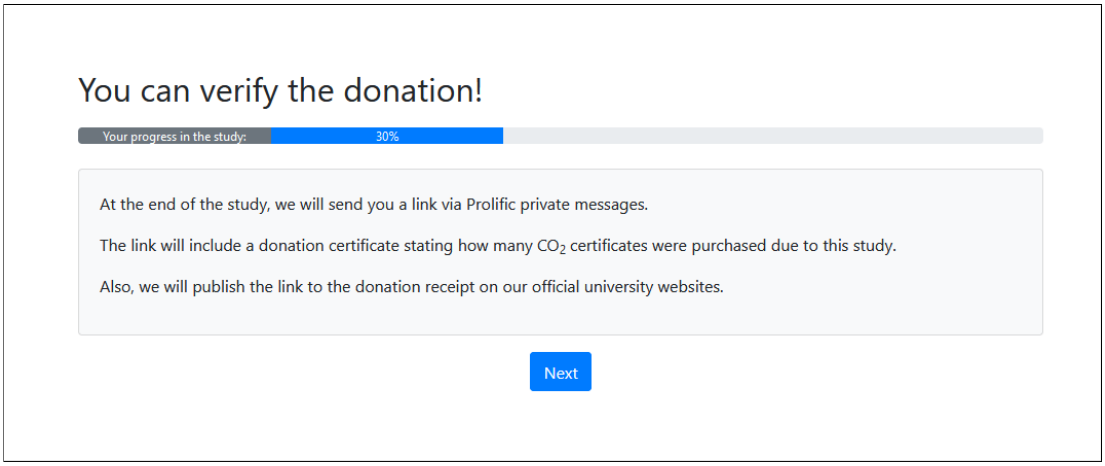
Your decision is real and will have real-world consequences!

For every participant in this study, we have arranged a donation to the nonprofit organisation Compensators*. Compensators* reduces a requested amount of CO_2 emissions by participating in the European Union Emissions Trading System, the largest worldwide. Their strategy is to buy permits that would allow them to emit CO_2 without actually doing so. As the number of permits is fixed, the carbon emissions of the EU are reduced by the requested amount.

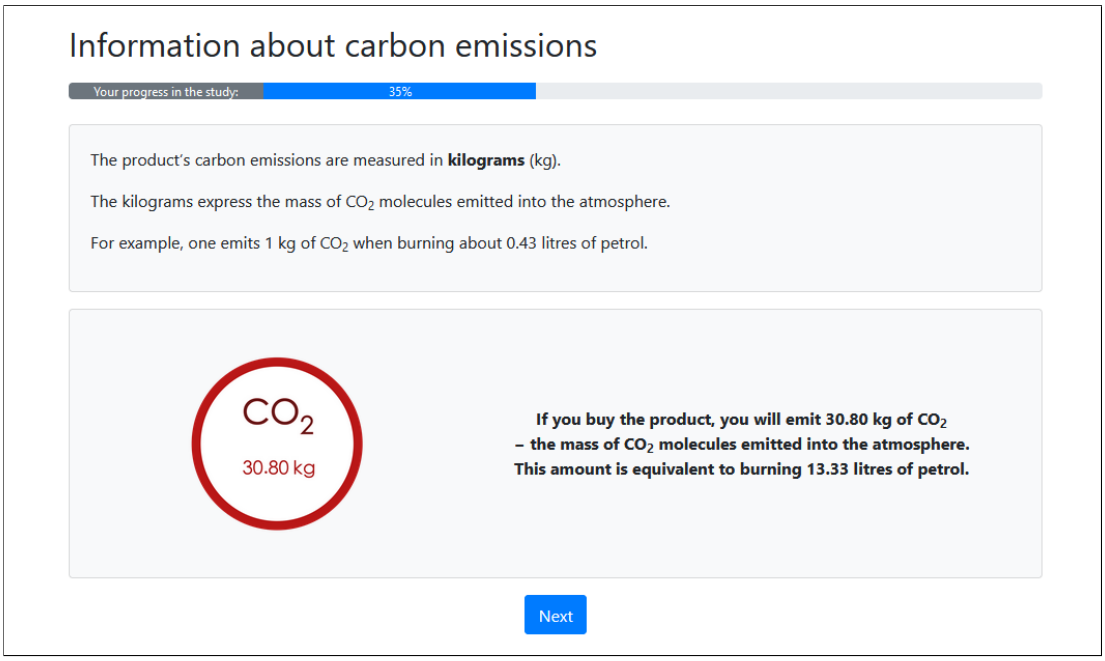
If you want to know more about Compensators*, you can access their website www.compensators.org.

If you buy the virtual product, we will cancel the particular donation to Compensators* that we have arranged for you, which leads to an increase in CO_2 emissions.

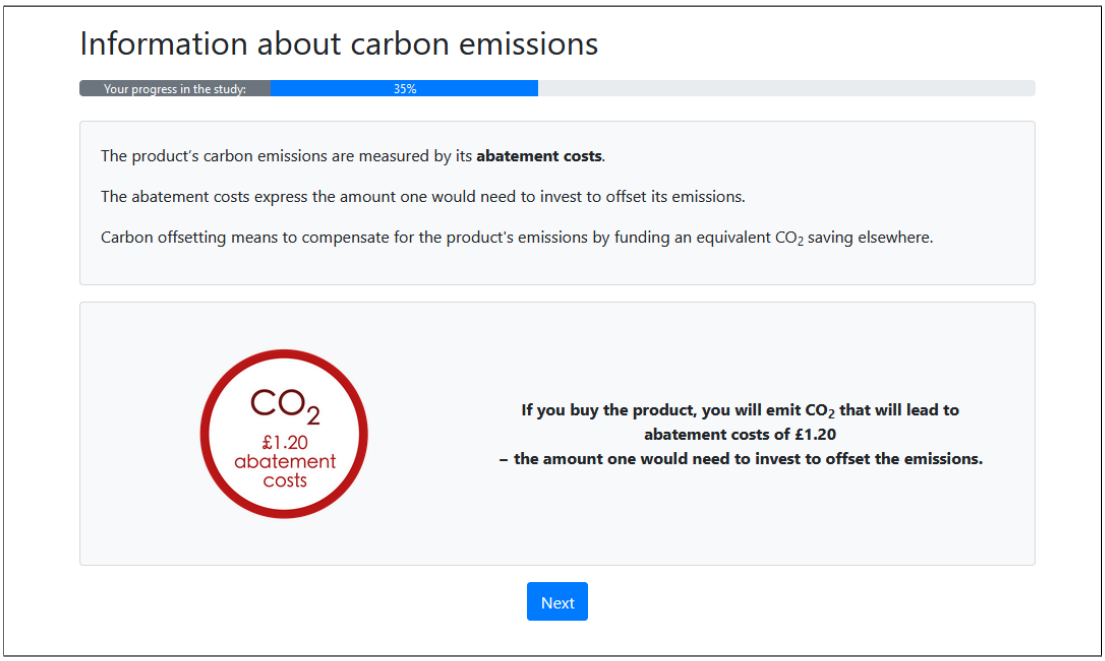
Next



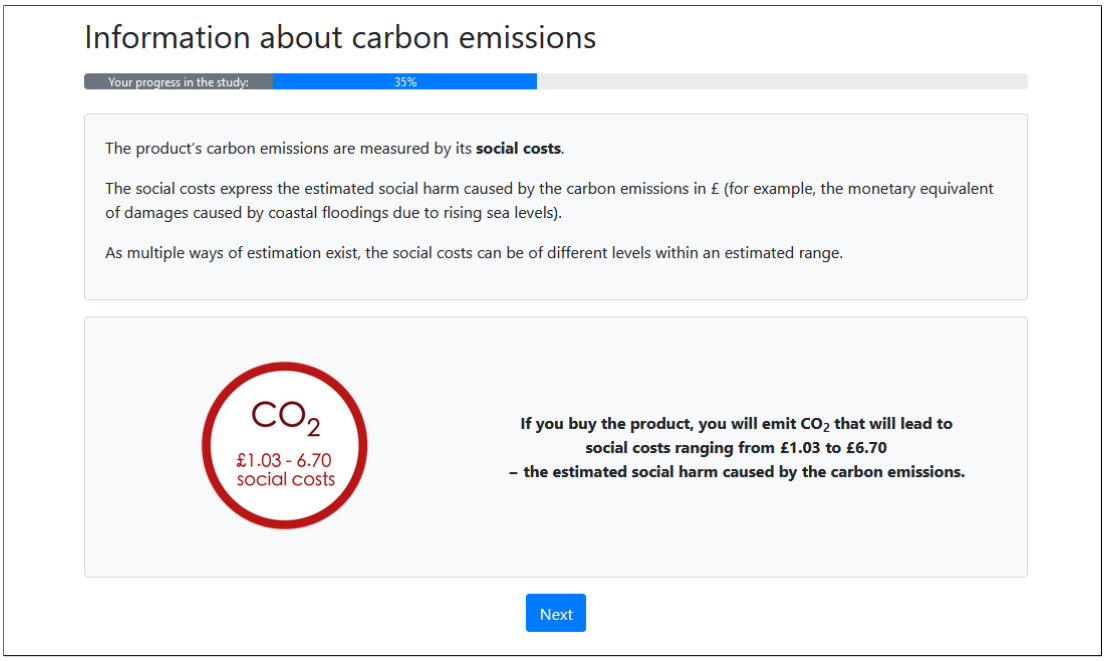
Treatment: KILOGRAM



Treatment: ABATEMENT COSTS



Treatment: SOCIAL COSTS



Treatment: KILOGRAM

Please answer the following questions:

Your progress in the study: 40%

Please answer the following questions:

If I buy the product, I will emit CO_2. These carbon emissions are real because this study does not use deception.

- ○ True
- ○ False

When the study is over, I will receive a private message via Prolific with the link to the donation receipt.

- ○ True
- ○ False

Fill in the missing words/numbers correctly. The information from the previous slide is given again at the bottom.

The product's carbon emissions are measured in		.
The kilograms express the		of CO_2 molecules emitted into the atmosphere.
If I buy the product, I will emit		kg of CO_2.
30.80 kg of CO_2 are equivalent to burning		litres of petrol.

Remember:

The product's carbon emissions are measured in **kilograms** (kg).

The kilograms express the mass of CO_2 molecules emitted into the atmosphere.

For example, one emits 1 kg of CO_2 when burning about 0.43 litres of petrol.

If you buy the product, you will emit 30.80 kg of CO_2 – the mass of CO_2 molecules emitted into the atmosphere. This amount is equivalent to burning 13.33 litres of petrol.

Next

Treatment: ABATEMENT COSTS

Please answer the following questions:

Your progress in the study: 40%

Please answer the following questions:

If I buy the product, I will emit CO_2. These carbon emissions are real because this study does not use deception.

- ○ True
- ○ False

When the study is over, I will receive a private message via Prolific with the link to the donation receipt.

- ○ True
- ○ False

Fill in the missing words/numbers correctly. The information from the previous slide is given again at the bottom.

The product's carbon emissions are measured by its	.
The abatement costs express the amount one would need to invest to	its emissions.
If I buy the product, I will emit CO_2 that will lead to abatement costs of	£ .

Remember:

The product's carbon emissions are measured by its **abatement costs**.

The abatement costs express the amount one would need to invest to offset its emissions.

Carbon offsetting means to compensate for the product's emissions by funding an equivalent CO_2 saving elsewhere.

If you buy the product, you will emit CO_2 that will lead to abatement costs of £1.20 – the amount one would need to invest to offset the emissions.

Next

Treatment: SOCIAL COSTS

Please answer the following questions:

Your progress in the study: 40%

Please answer the following questions:

If I buy the product, I will emit CO_2. These carbon emissions are real because this study does not use deception.

- ○ True
- ○ False

When the study is over, I will receive a private message via Prolific with the link to the donation receipt.

- ○ True
- ○ False

Fill in the missing words/numbers correctly. The information from the previous slide is given again at the bottom.

The product's carbon emissions are measured by its		.		
The social costs express the estimated		caused by the carbon emissions.		
If I buy the product, I will emit CO_2 that will lead to social costs ranging from	£	to	£	

Remember:

The product's carbon emissions are measured by its **social costs**.

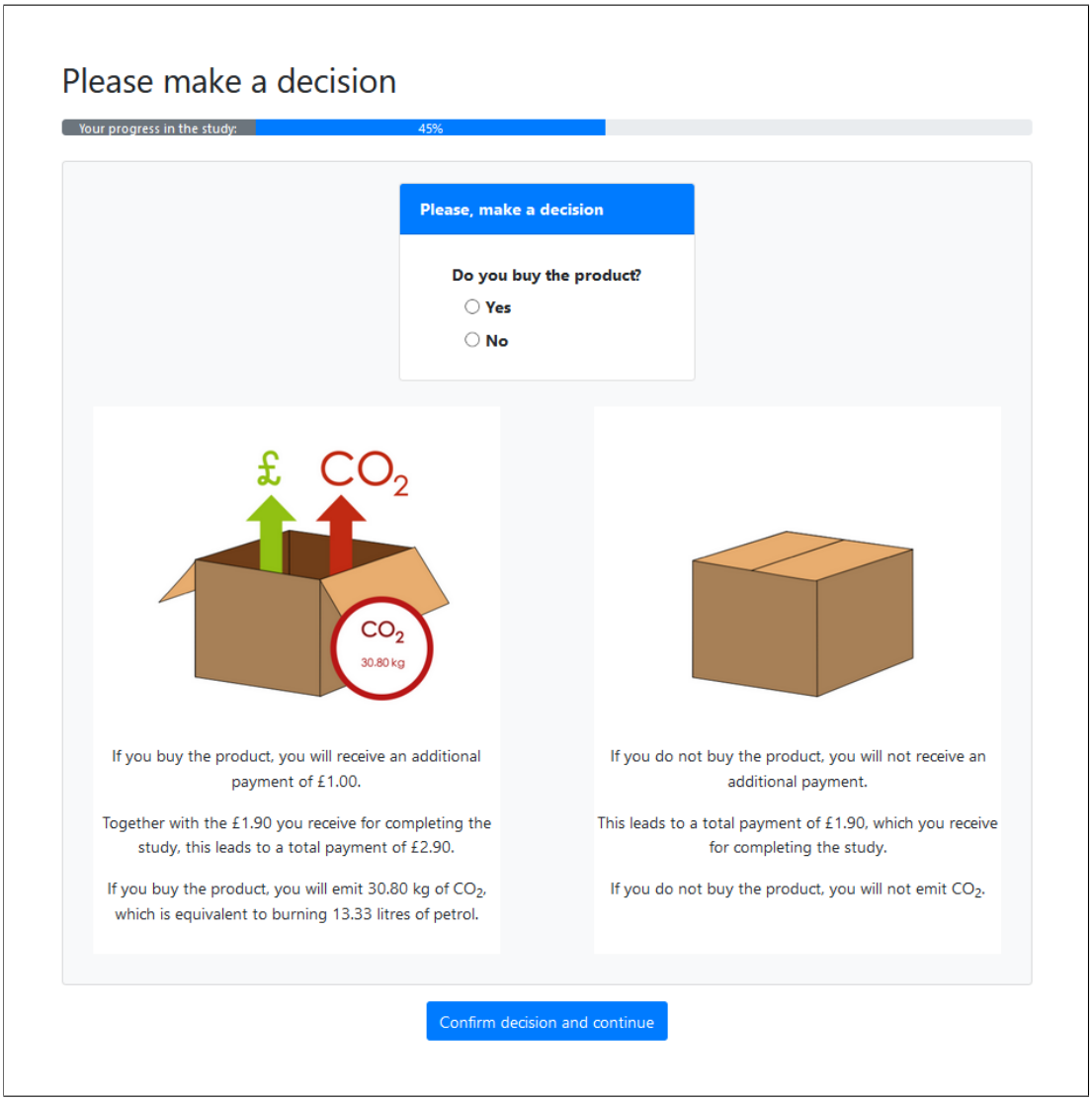
The social costs express the estimated social harm caused by the carbon emissions in £ (for example, the monetary equivalent of damages caused by coastal floodings due to rising sea levels).

As multiple ways of estimation exist, the social costs can be of different levels within an estimated range.

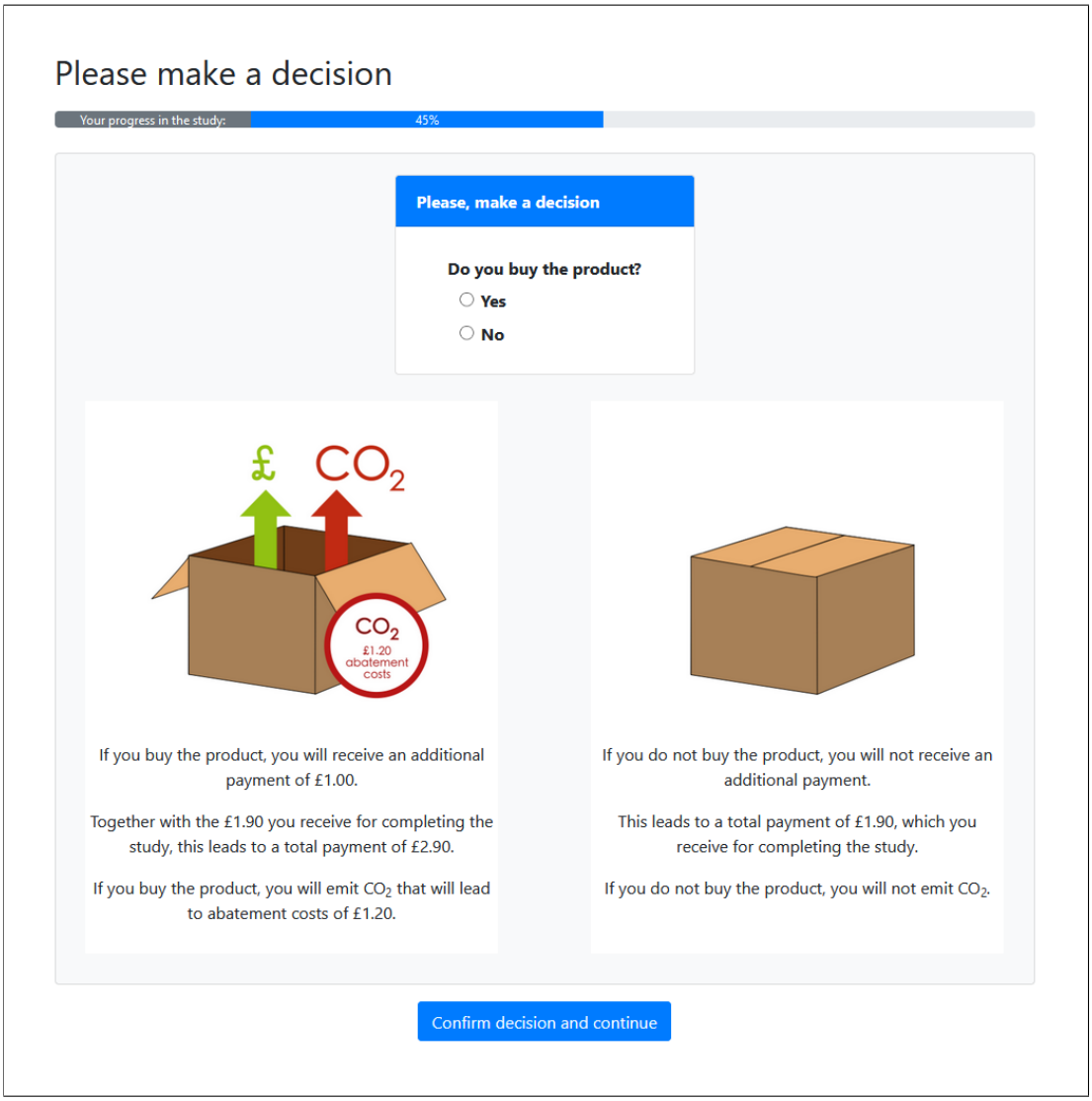
If you buy the product, you will emit CO_2 that will lead to social costs ranging from £1.03 to £6.70 – the estimated social harm caused by the carbon emissions.

Next

Treatment: KILOGRAM



Treatment: ABATEMENT COSTS



Treatment: SOCIAL COSTS

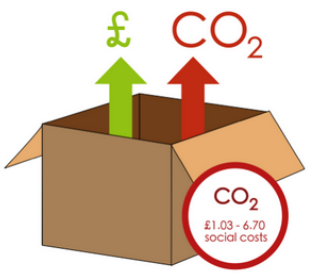
Please make a decision

Your progress in the study: 45%

Please, make a decision

Do you buy the product?

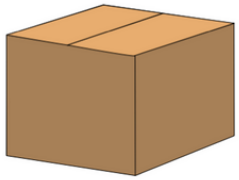
- **Yes**
- **No**



If you buy the product, you will receive an additional payment of £1.00.

Together with the £1.90 you receive for completing the study, this leads to a total payment of £2.90.

If you buy the product, you will emit CO_2 that will lead to social costs ranging from £1.03 to £6.70.



If you do not buy the product, you will not receive an additional payment.

This leads to a total payment of £1.90, which you receive for completing the study.

If you do not buy the product, you will not emit CO_2.

Confirm decision and continue

Instructions on guess

Your progress in the study: 50%

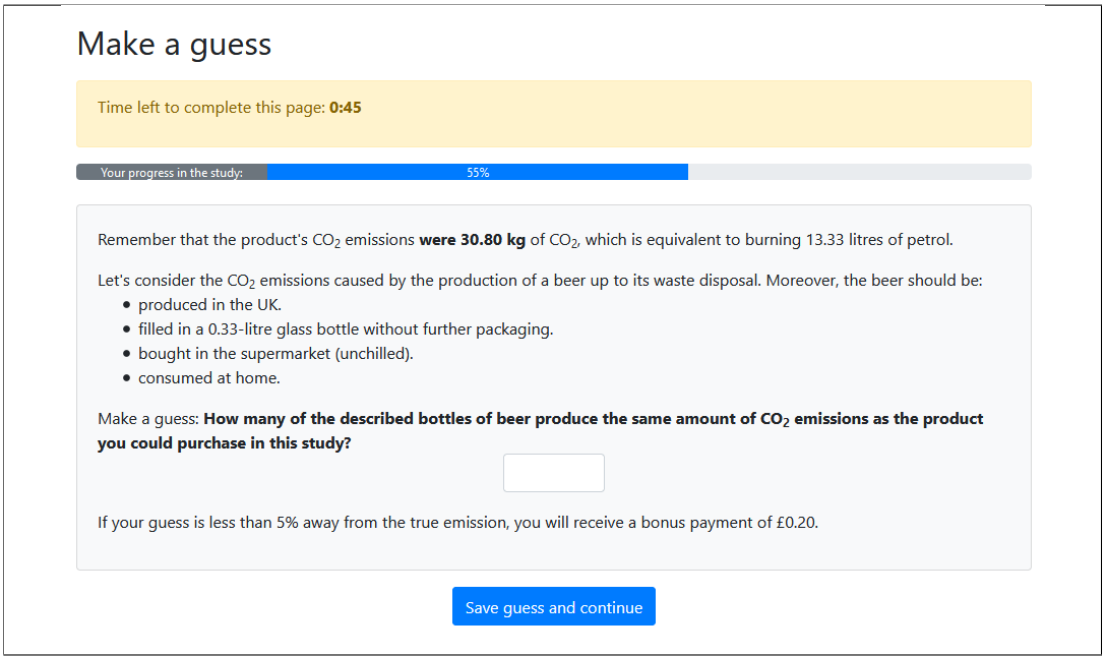
On the next page, you will have to make a guess. If your guess is less than 5% away from the correct value, you will receive a bonus payment of £0.20 at the end of the study, also made via Prolific.

You have 50 seconds to make your guess.

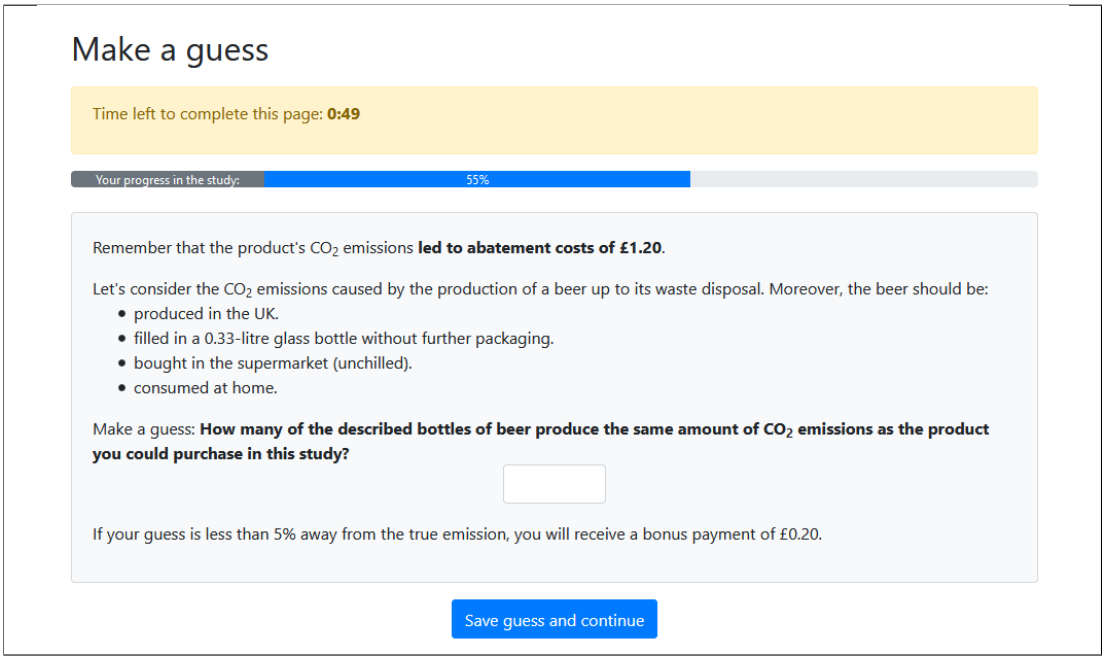
The following page will only be visible for 50 seconds. As soon as you press the button below, the timer starts.

Continue to next page and start timer

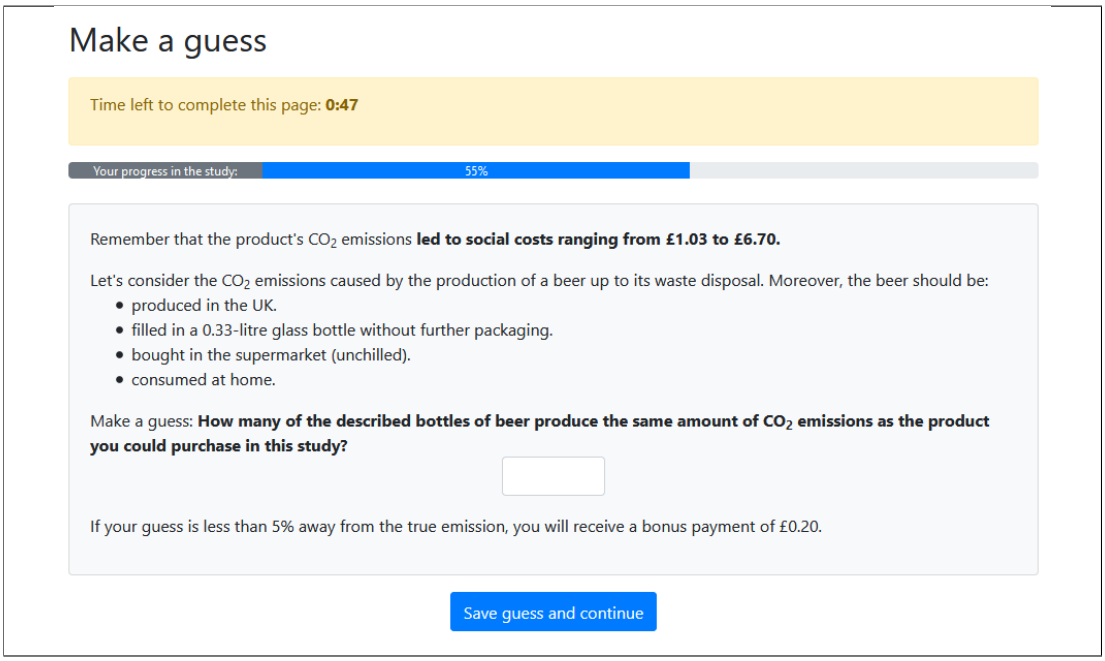
Treatment: KILOGRAM

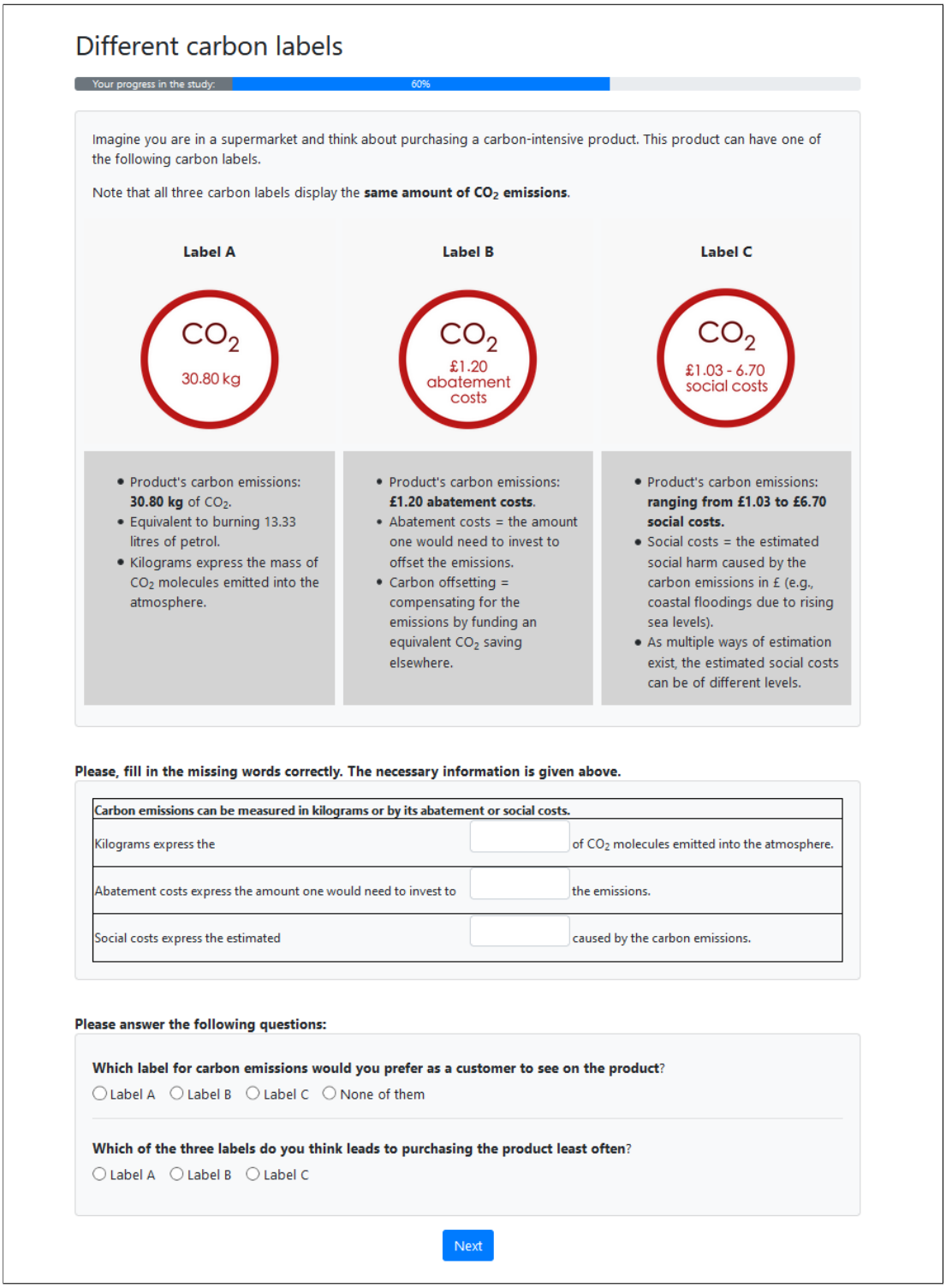


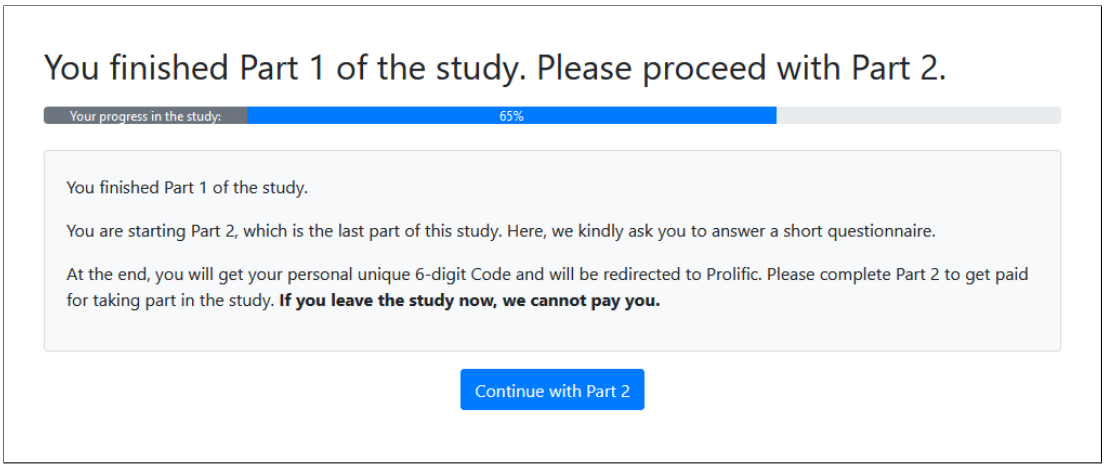
Treatment: ABATEMENT COSTS

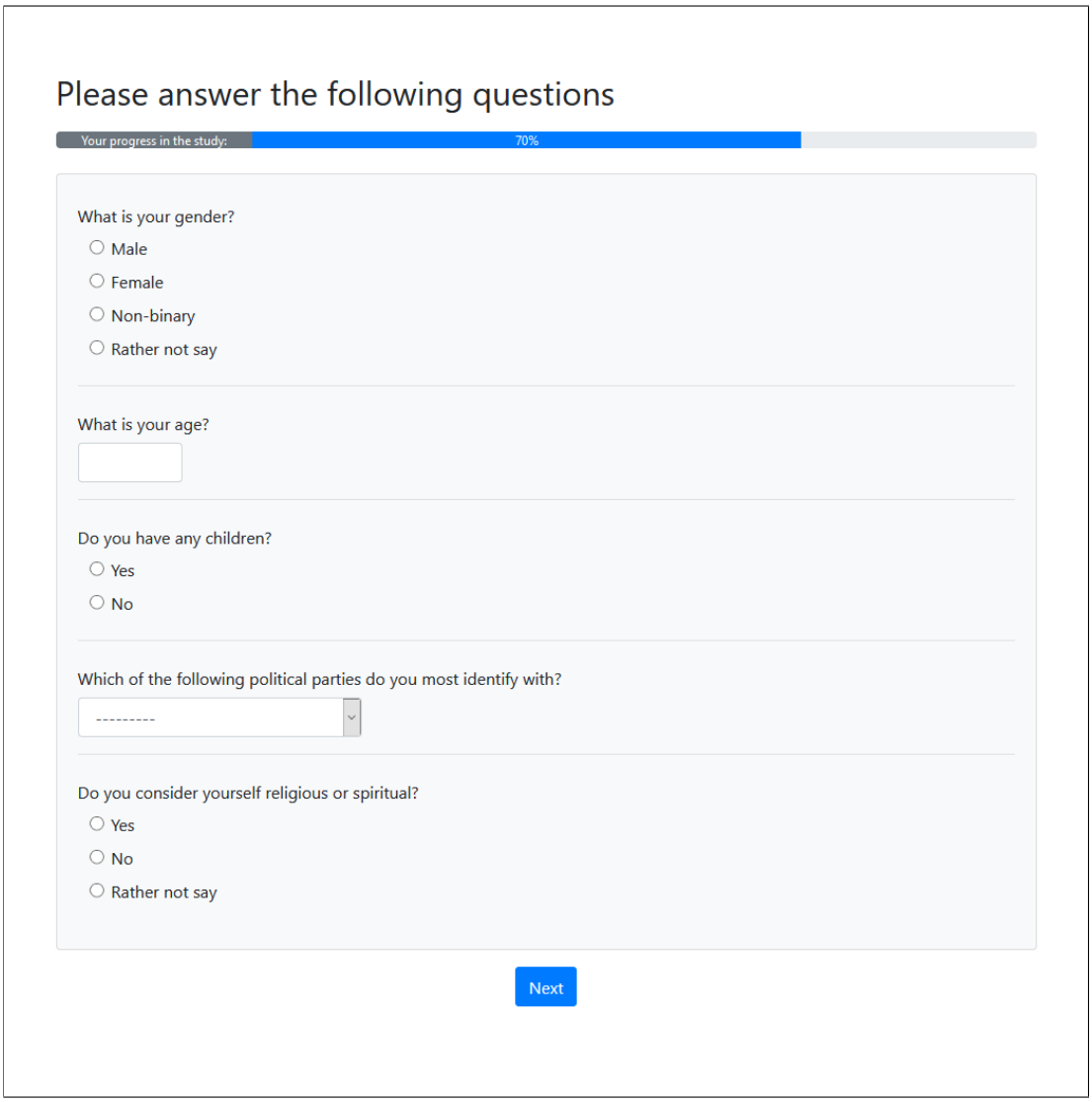


Treatment: SOCIAL COSTS

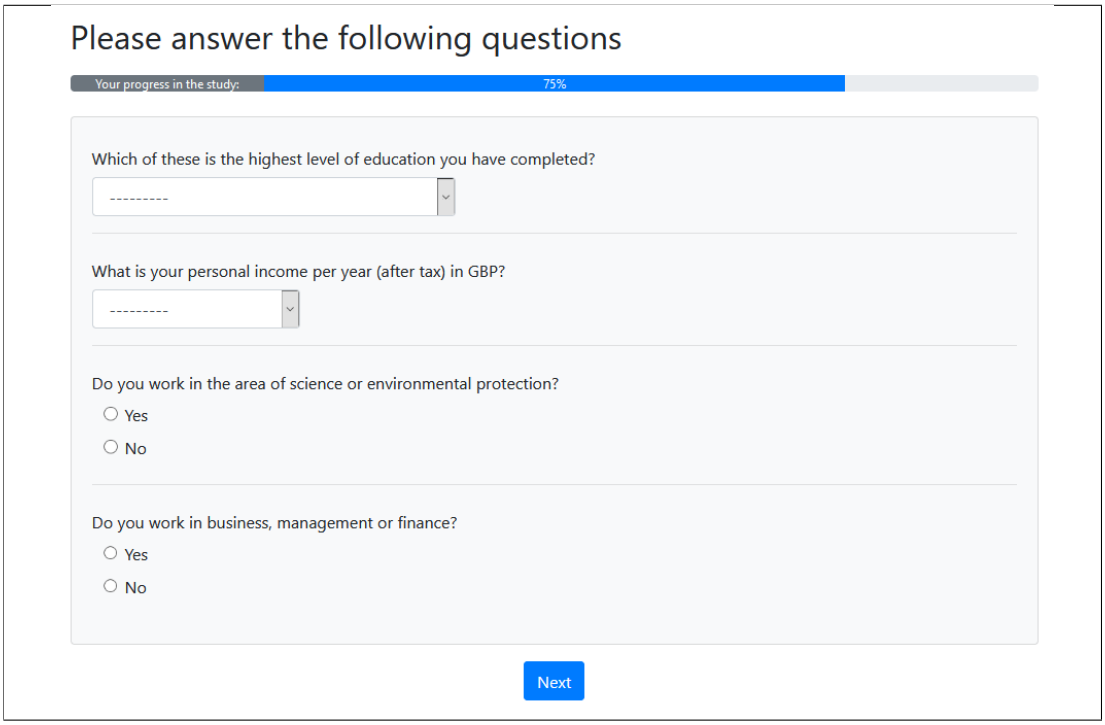



Different carbon labels
Your progress in the study:
60%
Imagine you are in a supermarket and think about purchasing a carbon-intensive product. This product can have one of the following carbon labels.
Note that all three carbon labels display the same amount of CO2 emissions.
Label A
CO2
30.80 kg
Label B
CO2
£1.20 abatement costs
Label C
CO2
£1.03 - 6.70 social costs
• Product's carbon emissions: 30.80 kg of CO2.
• Equivalent to burning 13.33 litres of petrol.
• Kilograms express the mass of CO2 molecules emitted into the atmosphere.
• Product's carbon emissions: £1.20 abatement costs.
• Abatement costs = the amount one would need to invest to offset the emissions.
• Carbon offsetting = compensating for the emissions by funding an equivalent CO2 saving elsewhere.
• Product's carbon emissions: ranging from £1.03 to £6.70 social costs.
• Social costs = the estimated social harm caused by the carbon emissions in £ (e.g., coastal floodings due to rising sea levels).
• As multiple ways of estimation exist, the estimated social costs can be of different levels.
Please, fill in the missing words correctly. The necessary information is given above.
Carbon emissions can be measured in kilograms or by its abatement or social costs.
Kilograms express the
of CO2 molecules emitted into the atmosphere.
Abatement costs express the amount one would need to invest to
the emissions.
Social costs express the estimated
caused by the carbon emissions.
Please answer the following questions:
Which label for carbon emissions would you prefer as a customer to see on the product?
Label A Label B Label C None of them
Which of the three labels do you think leads to purchasing the product least often?
Label A Label B Label C
Next


You finished Part 1 of the study. Please proceed with Part 2.
Your progress in the study:
65%
You finished Part 1 of the study.
You are starting Part 2, which is the last part of this study. Here, we kindly ask you to answer a short questionnaire.
At the end, you will get your personal unique 6-digit Code and will be redirected to Prolific. Please complete Part 2 to get paid for taking part in the study. If you leave the study now, we cannot pay you.
Continue with Part 2

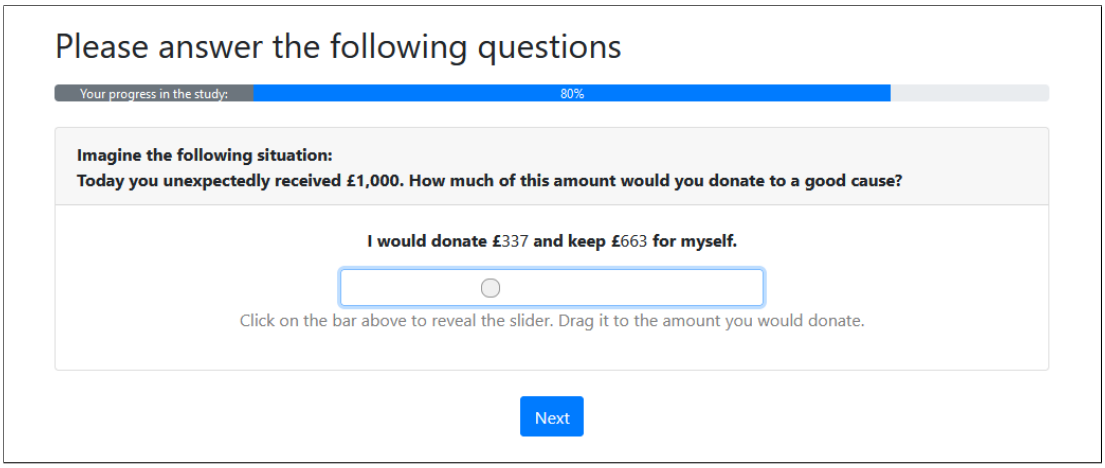

Please answer the following questions
Your progress in the study:
70%
What is your gender?
Male
Female
Non-binary
Rather not say
What is your age?
Do you have any children?
Yes
No
Which of the following political parties do you most identify with?

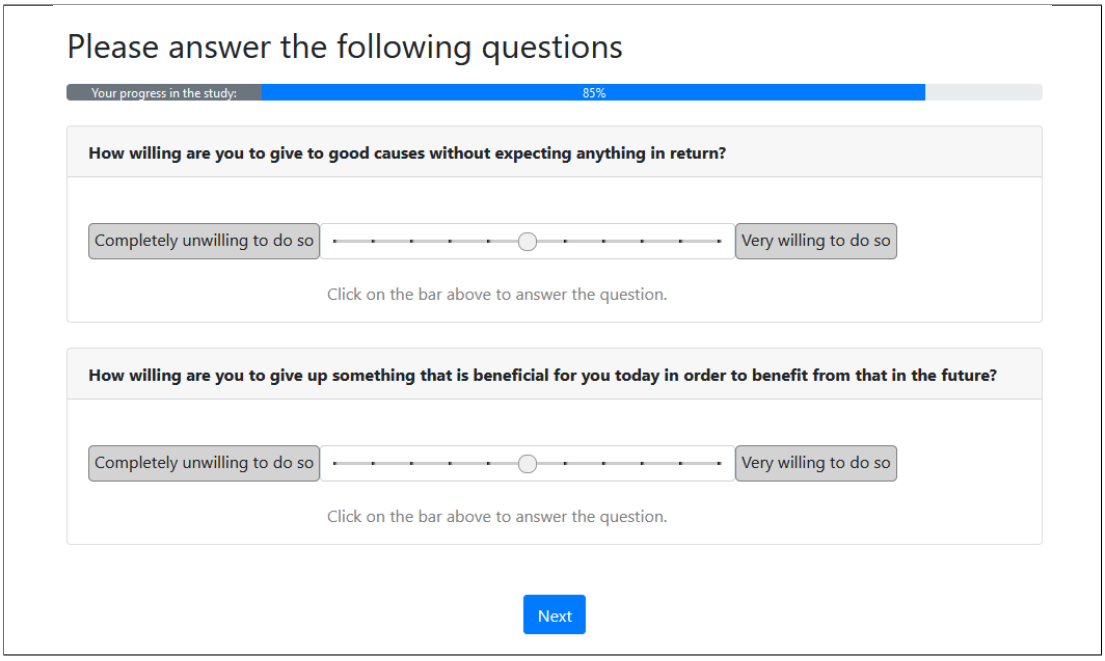
Do you consider yourself religious or spiritual?
Yes
No
Rather not say
Next

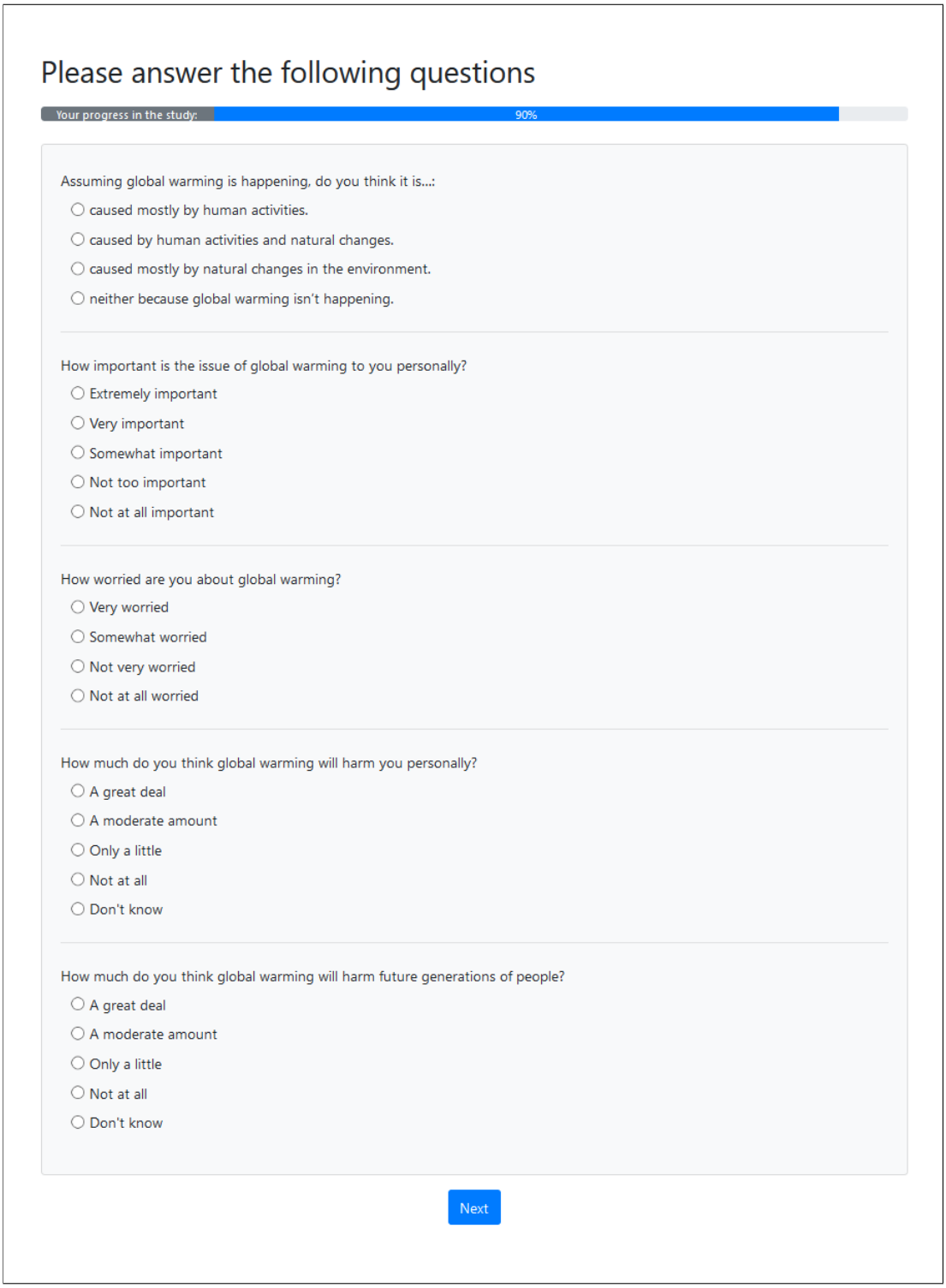

Please answer the following questions
Your progress in the study:
75%
Which of these is the highest level of education you have completed?

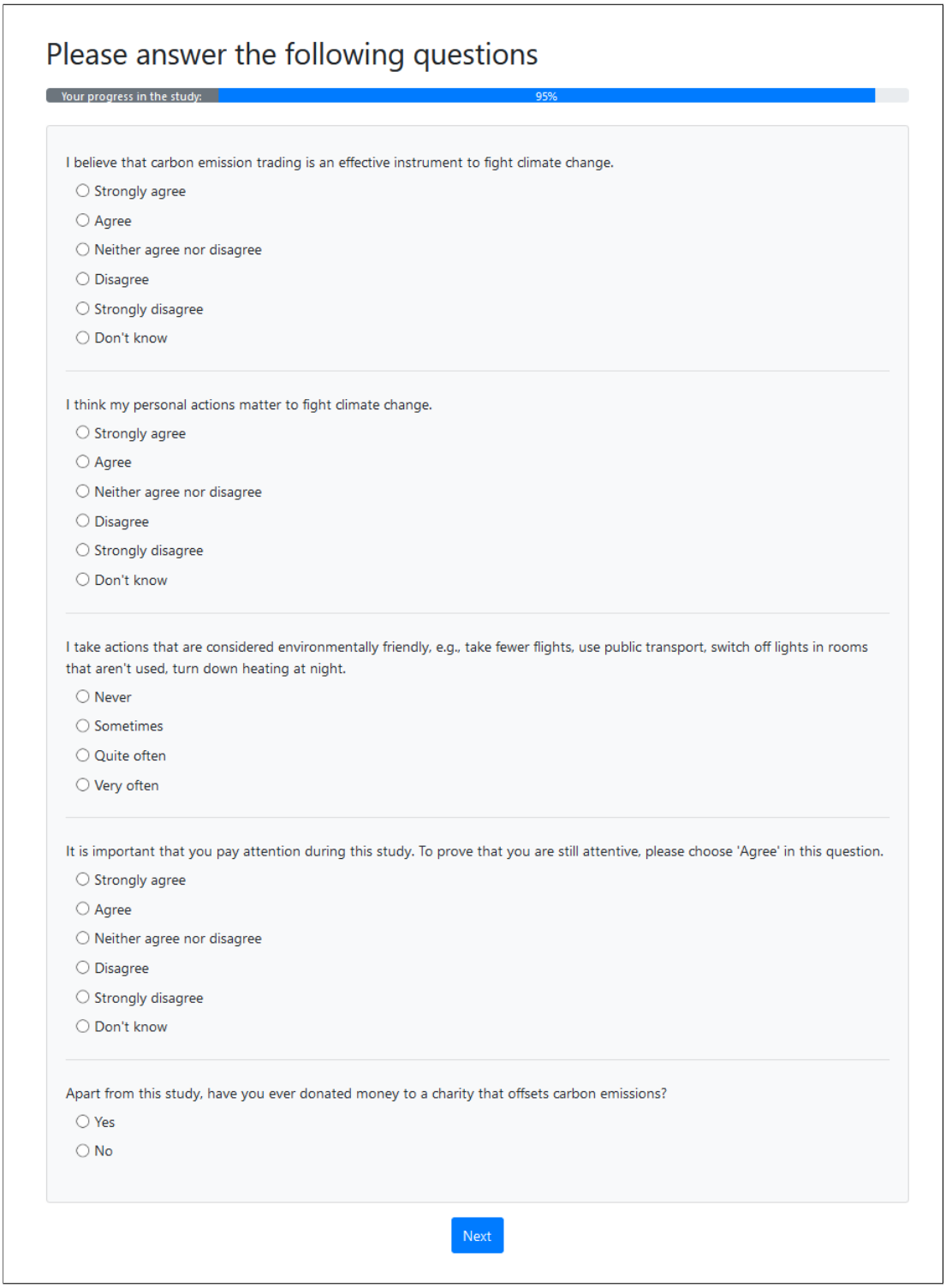
What is your personal income per year (after tax) in GBP?

Do you work in the area of science or environmental protection?
Yes
No
Do you work in business, management or finance?
Yes
No
Next


Please answer the following questions
Your progress in the study:
80%
Imagine the following situation:
Today you unexpectedly received £1,000. How much of this amount would you donate to a good cause?
I would donate £337 and keep £663 for myself.
Click on the bar above to reveal the slider. Drag it to the amount you would donate.
Next


Please answer the following questions
Your progress in the study:
85%
How willing are you to give to good causes without expecting anything in return?
Completely unwilling to do so
Very willing to do so
Click on the bar above to answer the question.
How willing are you to give up something that is beneficial for you today in order to benefit from that in the future?
Completely unwilling to do so
Very willing to do so
Click on the bar above to answer the question.
Next


Please answer the following questions
Your progress in the study:
90%
Assuming global warming is happening, do you think it is...:
caused mostly by human activities.
caused by human activities and natural changes.
caused mostly by natural changes in the environment.
neither because global warming isn't happening.
How important is the issue of global warming to you personally?
Extremely important
Very important
Somewhat important
Not too important
Not at all important
How worried are you about global warming?
Very worried
Somewhat worried
Not very worried
Not at all worried
How much do you think global warming will harm you personally?
A great deal
A moderate amount
Only a little
Not at all
Don't know
How much do you think global warming will harm future generations of people?
A great deal
A moderate amount
Only a little
Not at all
Don't know
Next



Please answer the following questions

Your progress in the study: 95%

I believe that carbon emission trading is an effective instrument to fight climate change.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree
- ○ Don't know

I think my personal actions matter to fight climate change.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree
- ○ Don't know

I take actions that are considered environmentally friendly, e.g., take fewer flights, use public transport, switch off lights in rooms that aren't used, turn down heating at night.

- ○ Never
- ○ Sometimes
- ○ Quite often
- ○ Very often

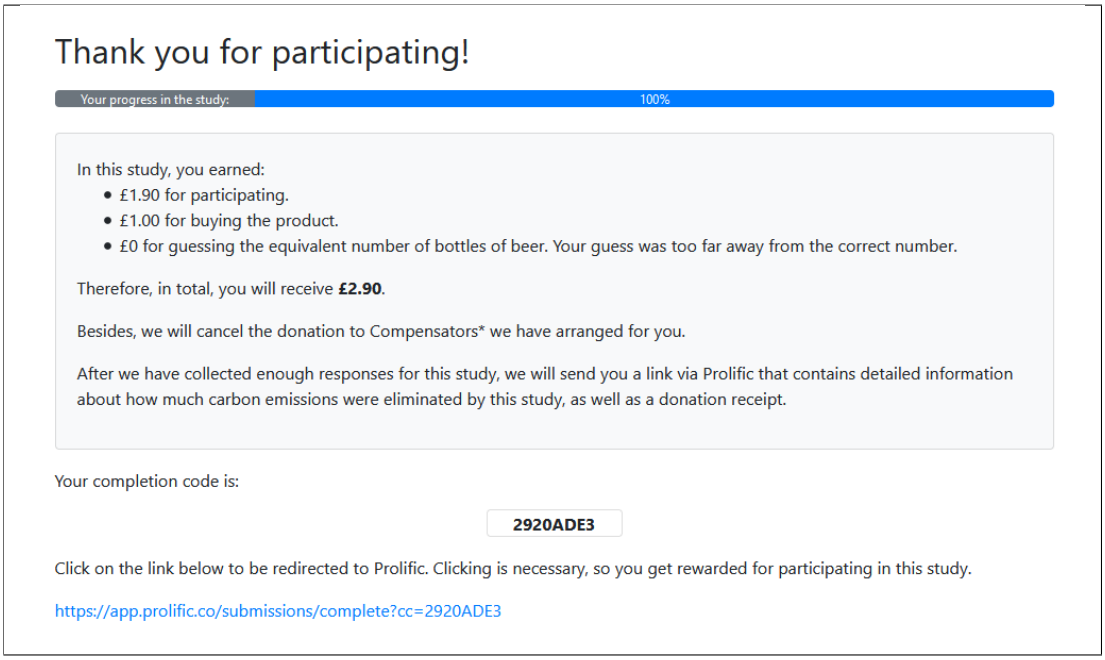
It is important that you pay attention during this study. To prove that you are still attentive, please choose 'Agree' in this question.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree
- ○ Don't know

Apart from this study, have you ever donated money to a charity that offsets carbon emissions?

- ○ Yes
- ○ No

Next


Thank you for participating!
Your progress in the study:
100%
In this study, you earned:
£1.90 for participating.
£1.00 for buying the product.
£0 for guessing the equivalent number of bottles of beer. Your guess was too far away from the correct number.
Therefore, in total, you will receive £2.90.
Besides, we will cancel the donation to Compensators* we have arranged for you.
After we have collected enough responses for this study, we will send you a link via Prolific that contains detailed information about how much carbon emissions were eliminated by this study, as well as a donation receipt.
Your completion code is:
2920ADE3
Click on the link below to be redirected to Prolific. Clicking is necessary, so you get rewarded for participating in this study.
https://app.prolific.co/submissions/complete?cc=2920ADE3

Appendix B

Moral Balancing

B.1 Additional results

Table B.1: Descriptive statistic by success in Part 1 in SELF.

Treatment: SELF	Mean/ Median	*Success*	*Failure*	Test Statistic (p-value)
		(N=150)	(N=150)	
Demographics				
Female	Mean	0.493 (0.502)	0.753 (0.433)	21.594 (0.000)
Male	Mean	0.480 (0.501)	0.227 (0.420)	21.066 (0.000)
Non-binary	Mean	0.013 (0.115)	0.027 (0.162)	0.680 (0.409)
Age	Mean	28.12 (9.043)	36.273 (13.860)	27.510 (0.000)
Has children	Mean	0.213 (0.411)	0.460 (0.500)	20.434 (0.000)
Education	Median	Undergraduate degree (BA/BSc/other)		0.587 (0.444)
Political orientation: Left	Mean	0.653 (0.478)	0.507 (0.502)	6.623 (0.010)
Political orientation: Right	Mean	0.153 (0.362)	0.273 (0.447)	6.435 (0.011)
Income	Median	10,000 - 29,999	10,000 - 29,999	0.015 (0.904)
Behavioral Preferences				
Altruism	Mean	-0.128 (0.917)	0.044 (0.735)	1.420 (0.233)
Patience	Mean	0.117 (0.994)	0.007 (1.069)	0.663 (0.416)
Environmental awareness				
SASSY segment	Median	Concerned	Concerned	2.545 (0.111)
Global warming caused by humans	Mean	0.707 (0.457)	0.560 (0.498)	6.947 (0.008)
Actions matter to fight climate change	Mean	0.713 (0.454)	0.833 (0.374)	6.161 (0.013)
Pro-environmental behavior	Mean	7.373 (2.138)	7.327 (2.097)	0.049 (0.824)
Has offset in past	Mean	0.167 (0.374)	0.233 (0.424)	2.083 (0.149)
Offset effective	Mean	0.667 (0.473)	0.687 (0.465)	0.137 (0.711)

Note: The sample is restricted to the SELF treatment. Standard deviation in parentheses for variables with means. For categorical variables, we use χ^2-tests; for numerical variables, we use Kruskal-Wallis tests. *Success* and *Failure* denote performance in the slider task of Part 1. Pro-environmental behavior is measured with respect to its frequency on a scale from 0 (Never) to 10 (Very often).

Table B.2: Descriptive statistic by success in Part 1 in LOW.

Treatment: LOW	Mean/ Median	*Success*	*Failure*	Test Statistic (p-value)
		(N=153)	(N=147)	
Demographics				
Female	Mean	0.523 (0.501)	0.762 (0.427)	18.591 (0.000)
Male	Mean	0.477 (0.501)	0.231 (0.423)	19.745 (0.000)
Non-binary	Mean	0.000 (0.000)	0.007 (0.082)	1.044 (0.307)
Age	Mean	28.065 (9.052)	32.714 (12.515)	8.518 (0.004)
Has children	Mean	0.203 (0.403)	0.374 (0.486)	10.788 (0.001)
Education	Median	Undergraduate degree (BA/BSc/other)		1.921 (0.166)
Political orientation: Left	Mean	0.654 (0.477)	0.605 (0.490)	0.746 (0.388)
Political orientation: Right	Mean	0.190 (0.363)	0.163 (0.371)	0.356 (0.551)
Income	Median	10,000 - 29,999	10,000 - 29,999	2.425 (0.119)
Behavioral Preferences				
Altruism	Mean	-0.036 (0.869)	0.100 (0.809)	1.236 (0.266)
Patience	Mean	-0.128 (1.009)	0.080 (0.901)	2.736 (0.098)
Environmental awareness				
SASSY segment	Median	Concerned	Concerned	0.001 (0.981)
Global warming caused by humans	Mean	0.699 (0.460)	0.626 (0.486)	1.813 (0.178)
Actions matter to fight climate change	Mean	0.752 (0.433)	0.857 (0.351)	5.283 (0.022)
Pro-environmental behavior	Mean	7.562 (1.747)	7.483 (2.005)	0.065 (0.798)
Has offset in past	Mean	0.203 (0.403)	0.197 (0.399)	0.013 (0.403)
Offset effective	Mean	0.706 (0.457)	0.660 (0.475)	0.734 (0.392)

Note: The sample is restricted to the LOW treatment. Standard deviation in parentheses for variables with means. For categorical variables, we use χ^2-tests; for numerical variables, we use Kruskal-Wallis tests. *Success* and *Failure* denote performance in the slider task of Part 1. Pro-environmental behavior is measured with respect to its frequency on a scale from 0 (Never) to 10 (Very often).

Table B.3: Descriptive statistic by success in Part 1 in HIGH.

Treatment: HIGH	Mean/ Median	*Success*	*Failure*	Test Statistic (p-value)
Number of observations		146	154	
Demographics				
Female	Mean	0.521 (0.501)	0.701 (0.459)	10.325 (0.001)
Male	Mean	0.473 (0.501)	0.286 (0.453)	11.149 (0.001)
Non-binary	Mean	0.007 (0.083)	0.006 (0.081)	0.001 (0.970)
Age	Mean	28.397 (8.512)	36.097 (14.998)	16.488 (0.000)
Has children	Mean	0.130 (0.338)	0.435 (0.497)	34.078 (0.000)
Education	Median	Undergraduate degree	Technical/community college	7.798 (0.005)
Political orientation: Left	Mean	0.582 (0.495)	0.513 (0.501)	1.448 (0.229)
Political orientation: Right	Mean	0.247 (0.433)	0.240 (0.429)	0.016 (0.899)
Income	Median	10,000 - 29,999	10,000 - 29,999	3.816 (0.501)
Behavioral Preferences				
Altruism	Mean	-0.039 (0.787)	0.059 (0.810)	1.754 (0.185)
Patience	Mean	-0.085 (1.032)	0.011 (0.980)	0.313 (0.576)
Environmental awareness				
SASSY segment	Median	Concerned	Concerned	0.523 (0.470)
Global warming caused by humans	Mean	0.685 (0.466)	0.675 (0.470)	0.032 (0.858)
Actions matter to fight climate change	Mean	0.664 (0.474)	0.727 (0.447)	1.403 (0.236)
Pro-environmental behavior	Mean	7.171 (2.320)	7.558 (2.074)	1.832 (0.176)
Has offset in past	Mean	0.253 (0.436)	0.182 (0.387)	2.264 (0.132)
Offset effective	Mean	0.712 (0.454)	0.597 (0.492)	4.371 (0.037)

Note: The sample is restricted to the HIGH treatment. Standard deviation in parentheses for variables with means. For categorical variables, we use χ^2-tests; for numerical variables, we use Kruskal-Wallis tests. *Success* and *Failure* denote performance in the slider task of Part 1. Pro-environmental behavior is measured with respect to its frequency on a scale from 0 (Never) to 10 (Very often).

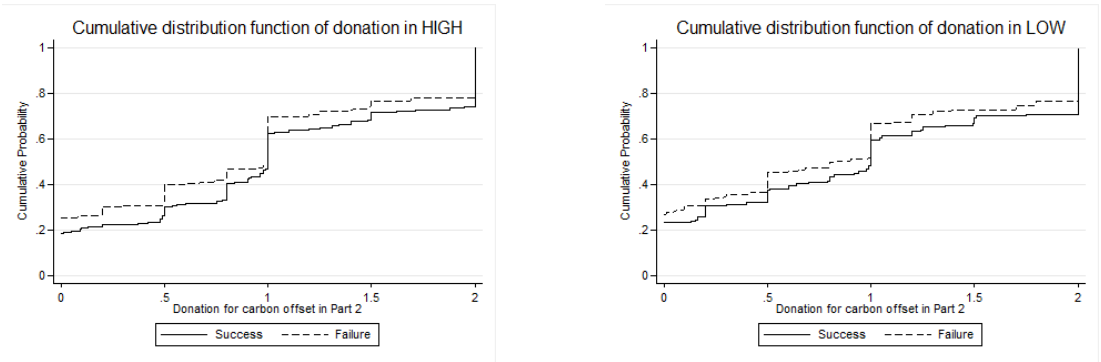


Figure B.1: Cumulative distribution functions of donations for carbon offset in Part 2 by treatment.

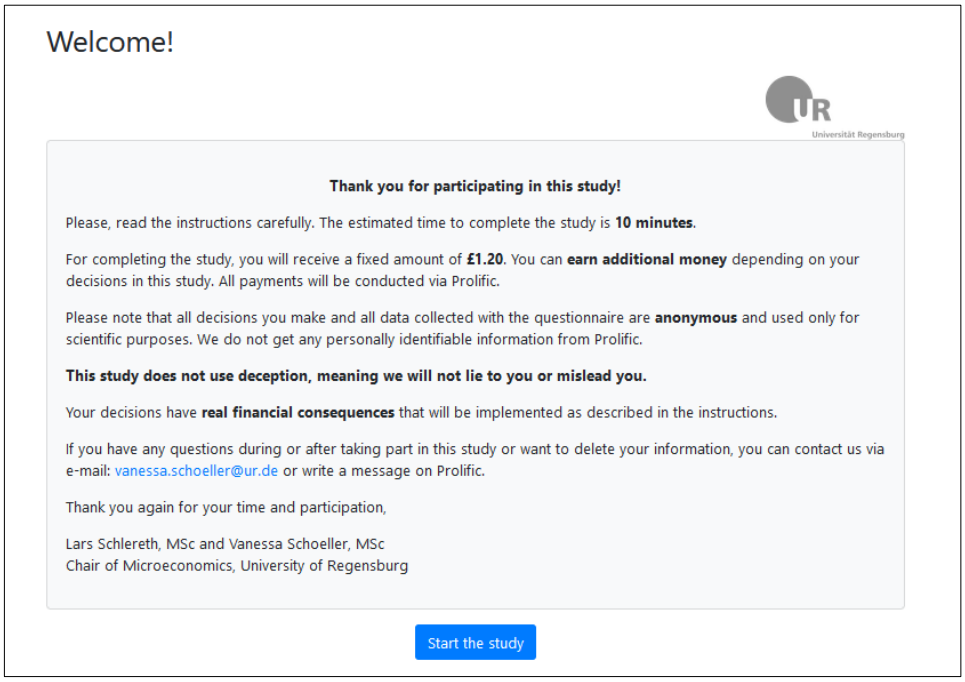
Table B.4: OLS regressions to test for moral balancing between treatments.

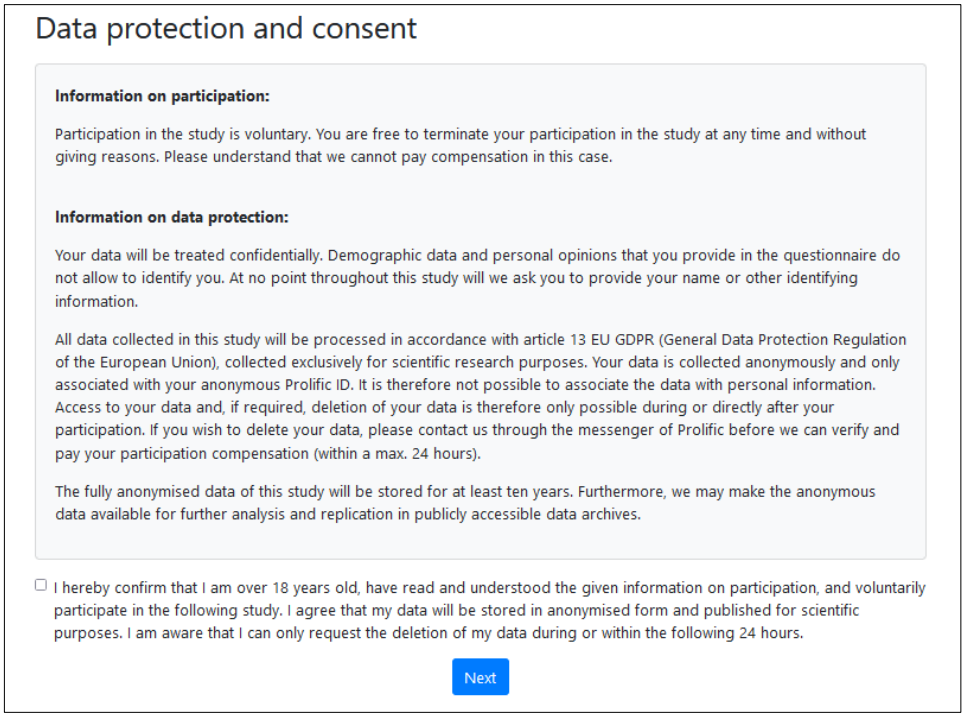
	Dependent variable: Donation			
	Failure		*Success*	
	(1)	(2)	(3)	(4)
LOW/HIGH	-0.018	-0.004	-0.048	-0.046
	(0.073)	(0.072)	(0.074)	(0.073)
Constant	1.022***	0.741***	0.933***	0.682***
	(0.058)	(0.088)	(0.059)	(0.082)
Additional Control Included: Actions Matter	No	Yes	No	Yes
Observations	451	451	449	449
R^2	0.0001	0.032	0.001	0.047

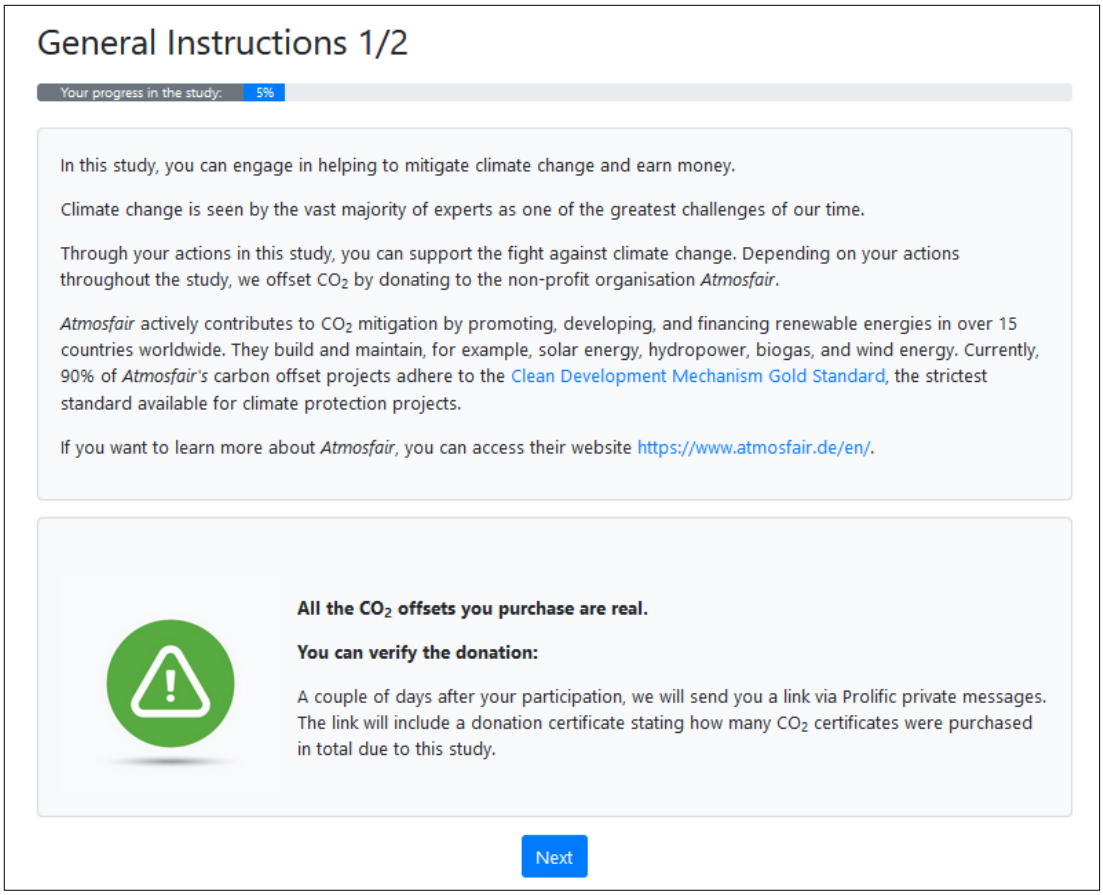
Note: The sample is split by *Failure* (columns (1) and (2)) and *Success* (columns (3) and (4)) in the slider task of Part 1. LOW/HIGH is a dummy, taking the value 1 if a participant is in the LOW or HIGH treatment, and 0 if she is in the SELF treatment. In columns (2) and (4), we control for whether participants agree that their actions matter to fight climate change; as for this variable, we find significant differences between treatments (see Table 4). Robust standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01.

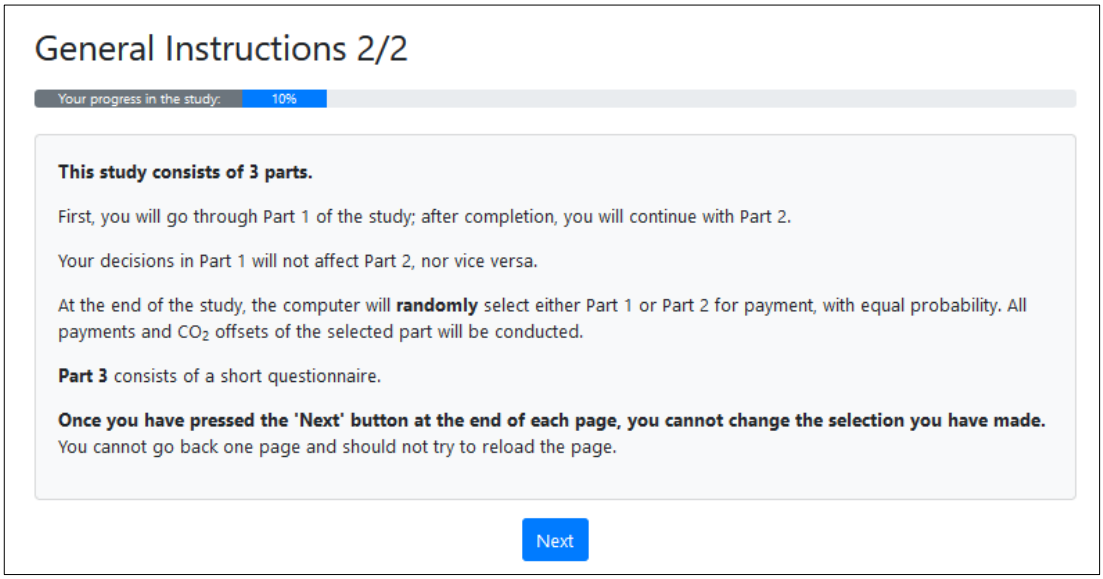
B.2 Instructions

The following pages contain screenshots of the online experiment conducted via Prolific. The study showed each participant partly different slides, depending on the assigned treatment condition and the participant's performance in the real-effort task. Participants were randomly allocated to one of three treatment conditions. Headlines stating the specific treatment as well as failure or success in the real-effort task mark the varying screens. All other pages were identical.

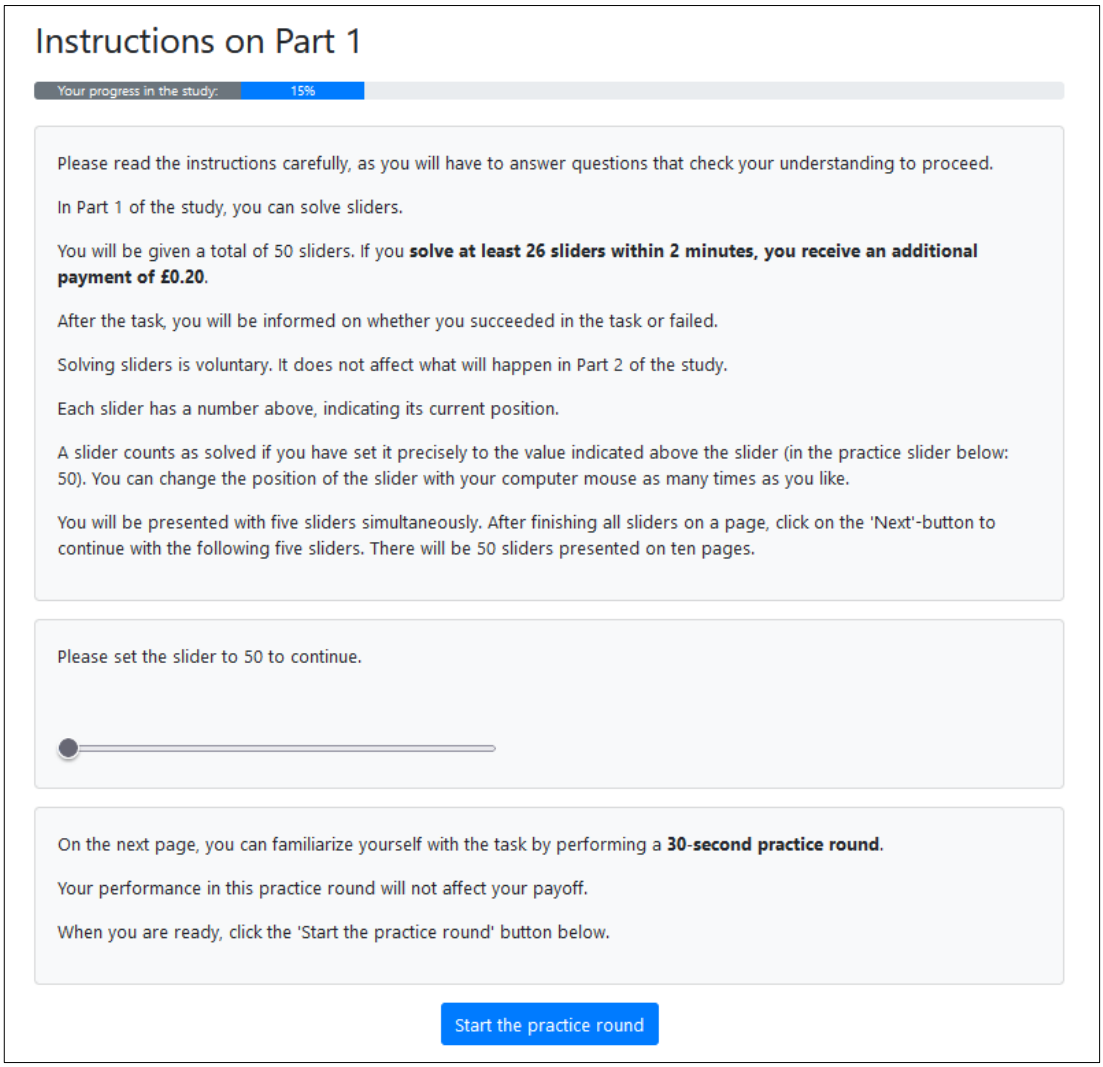




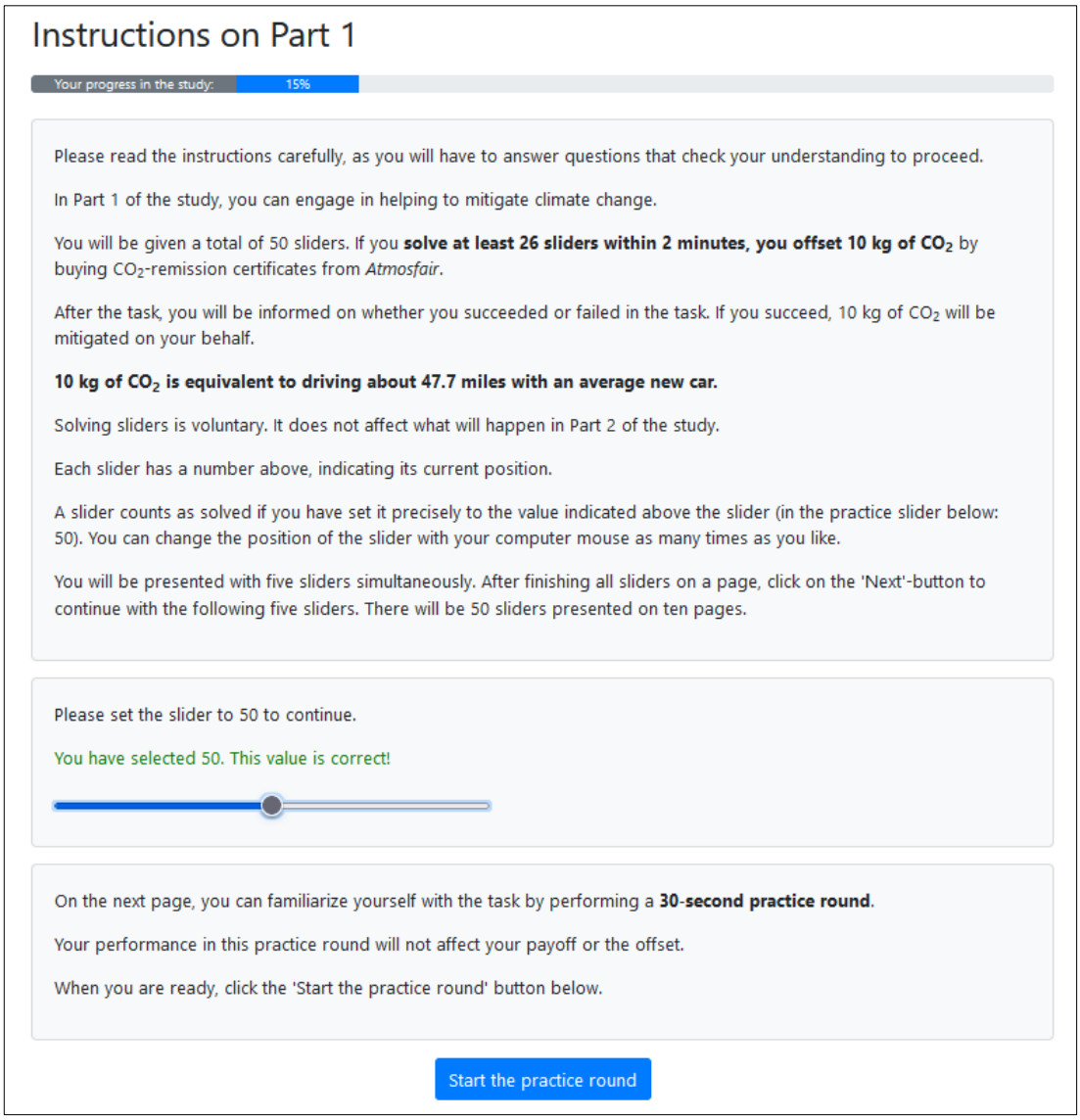

General Instructions 1/2
Your progress in the study: 5%
In this study, you can engage in helping to mitigate climate change and earn money.
Climate change is seen by the vast majority of experts as one of the greatest challenges of our time.
Through your actions in this study, you can support the fight against climate change. Depending on your actions throughout the study, we offset CO_2 by donating to the non-profit organisation *Atmosfair*.
Atmosfair actively contributes to CO_2 mitigation by promoting, developing, and financing renewable energies in over 15 countries worldwide. They build and maintain, for example, solar energy, hydropower, biogas, and wind energy. Currently, 90% of *Atmosfair*'s carbon offset projects adhere to the Clean Development Mechanism Gold Standard, the strictest standard available for climate protection projects.
If you want to learn more about *Atmosfair*, you can access their website https://www.atmosfair.de/en/.
All the CO_2 offsets you purchase are real.
You can verify the donation:
A couple of days after your participation, we will send you a link via Prolific private messages. The link will include a donation certificate stating how many CO_2 certificates were purchased in total due to this study.
Next


General Instructions 2/2
Your progress in the study: 10%
This study consists of 3 parts.
First, you will go through Part 1 of the study; after completion, you will continue with Part 2.
Your decisions in Part 1 will not affect Part 2, nor vice versa.
At the end of the study, the computer will **randomly** select either Part 1 or Part 2 for payment, with equal probability. All payments and CO_2 offsets of the selected part will be conducted.
Part 3 consists of a short questionnaire.
Once you have pressed the 'Next' button at the end of each page, you cannot change the selection you have made. You cannot go back one page and should not try to reload the page.
Next

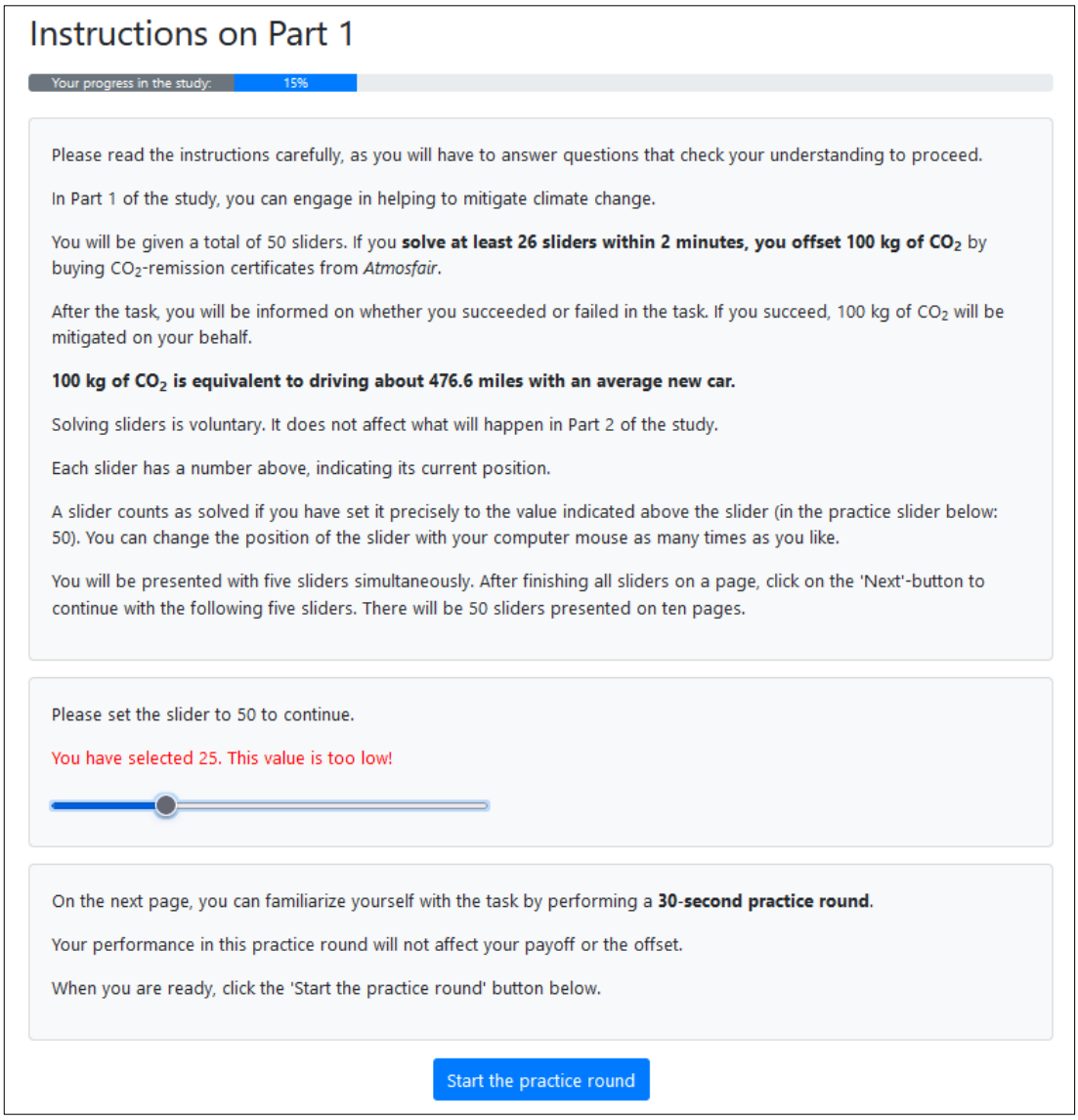
Treatment: SELF

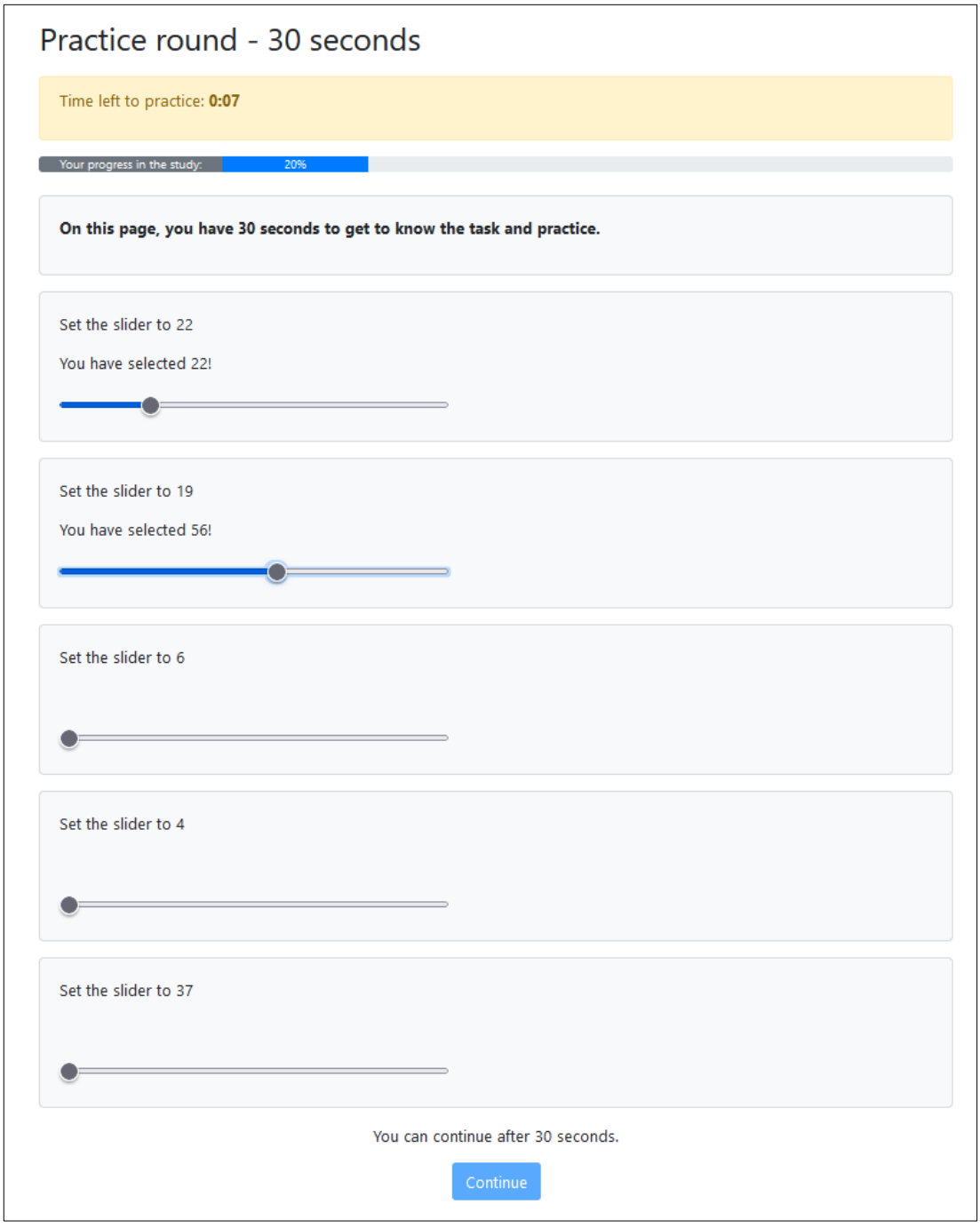


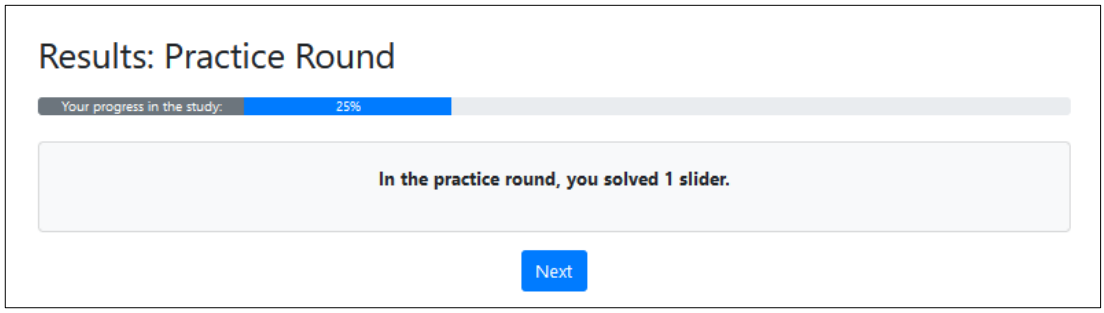
Treatment: LOW



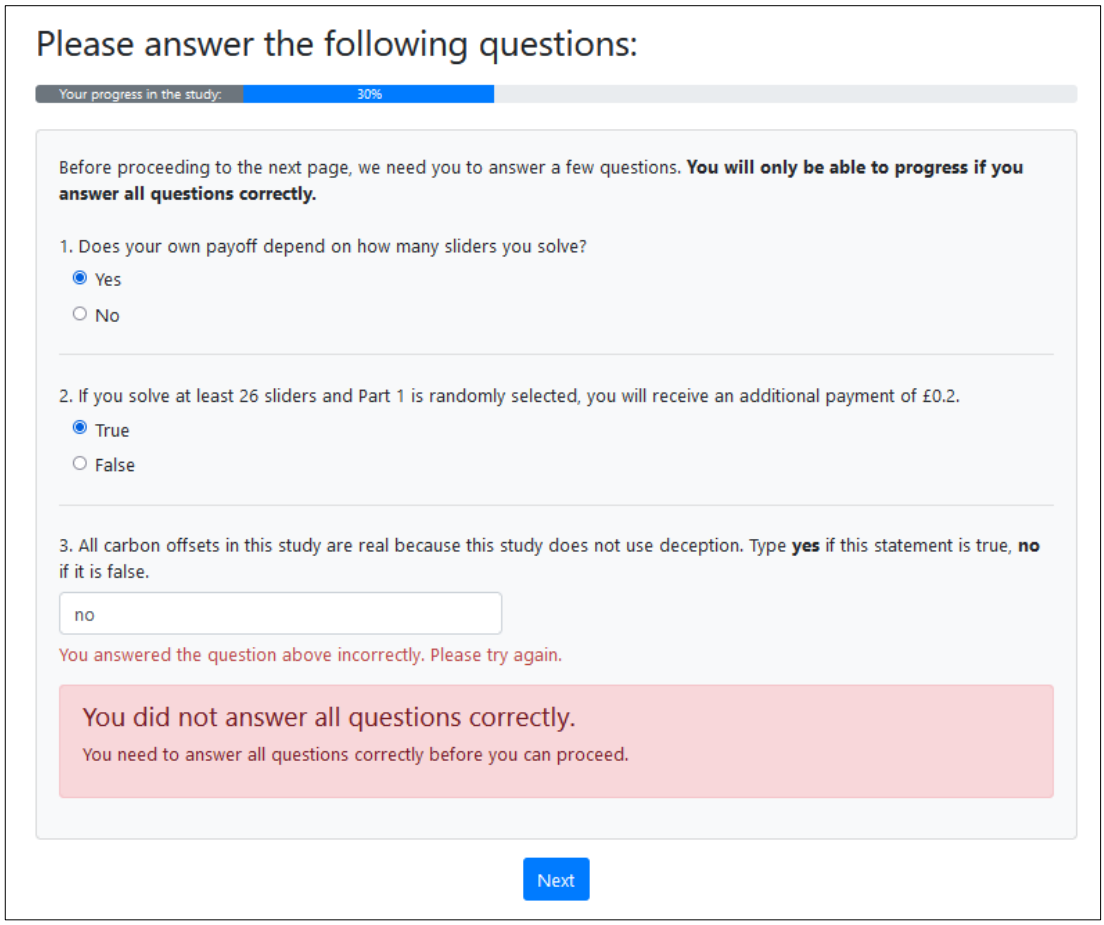
Treatment: HIGH



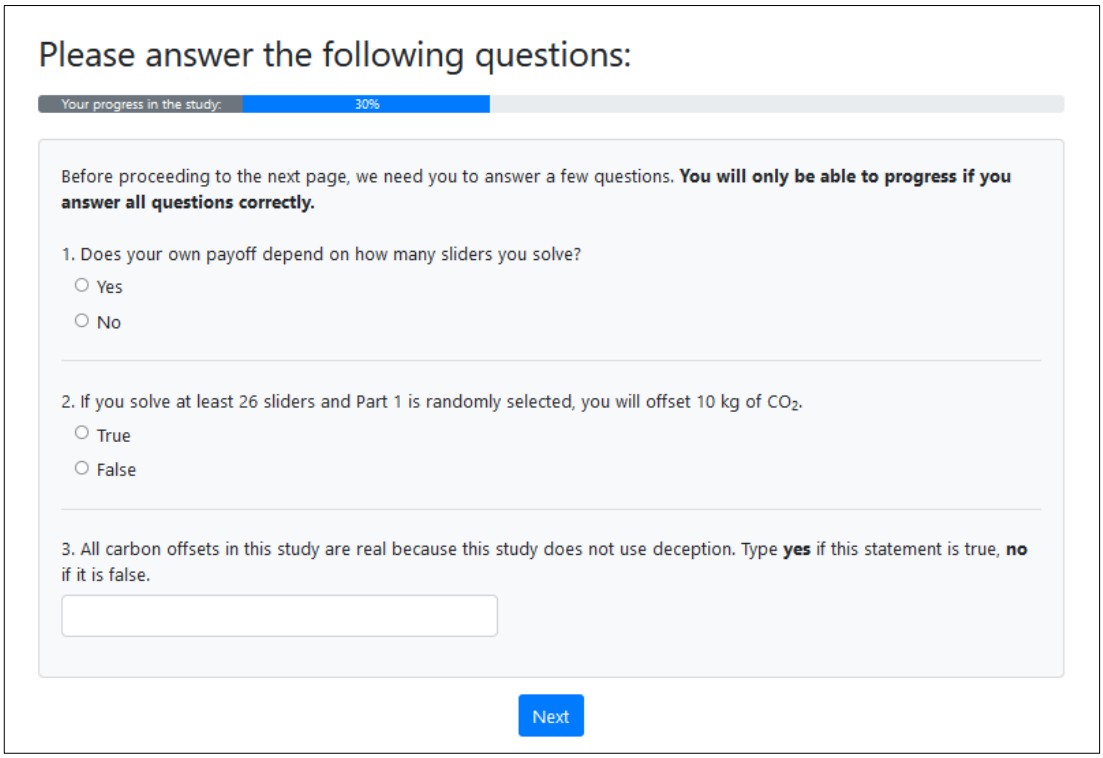

Practice round - 30 seconds
Time left to practice: 0:07
Your progress in the study:
20%
On this page, you have 30 seconds to get to know the task and practice.
Set the slider to 22
You have selected 22!
Set the slider to 19
You have selected 56!
Set the slider to 6
Set the slider to 4
Set the slider to 37
You can continue after 30 seconds.
Continue



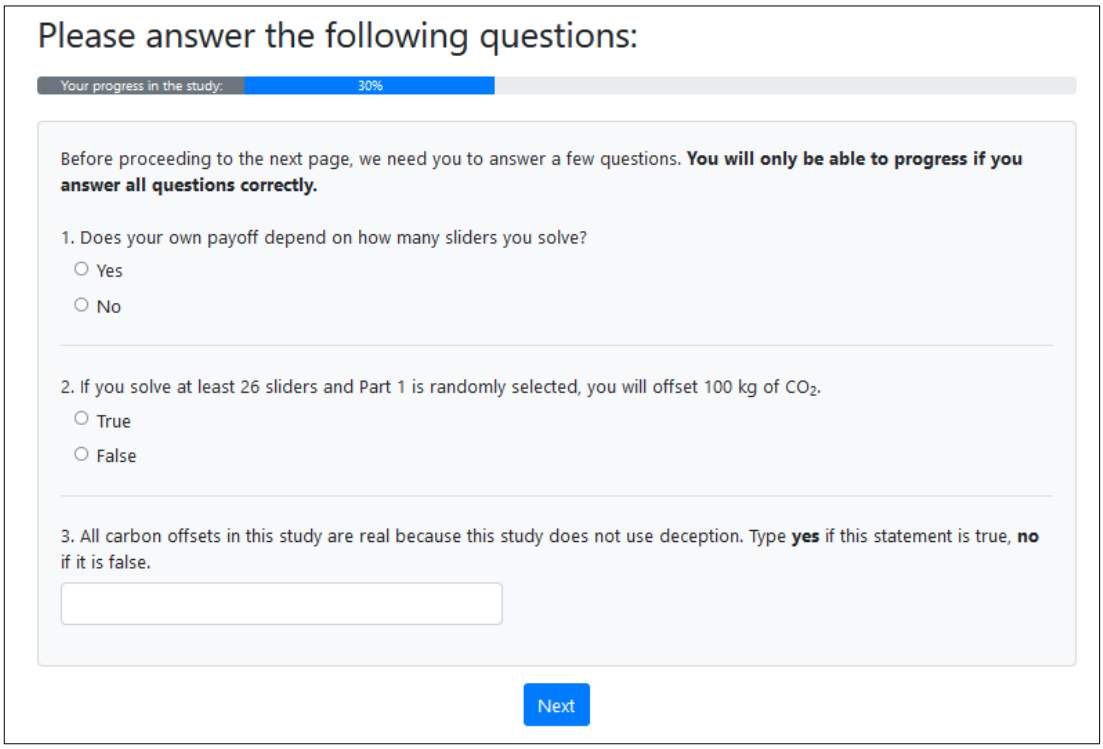
Treatment: SELF



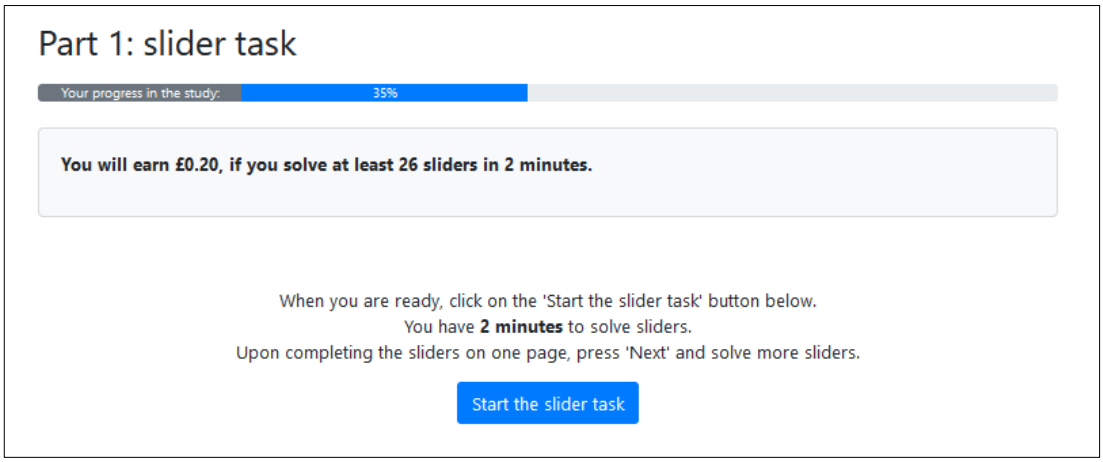
Treatment: LOW



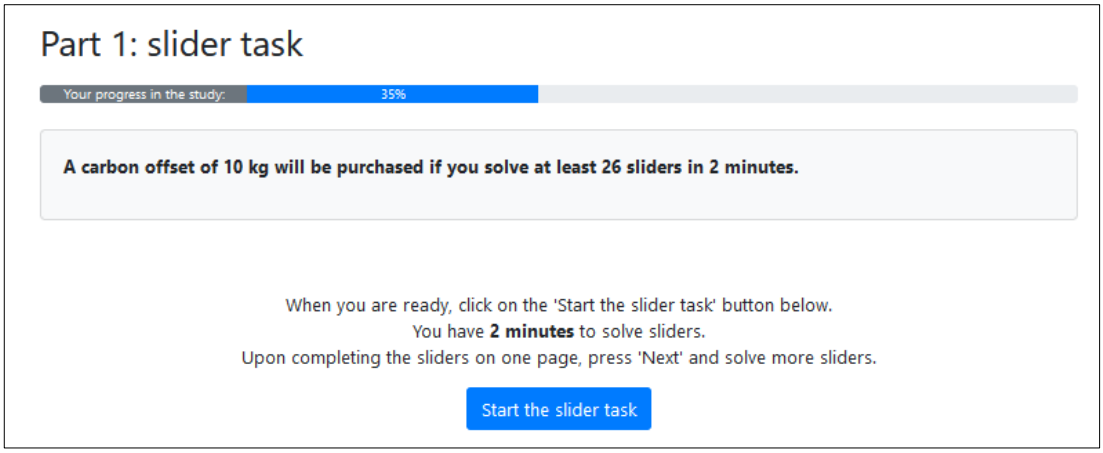
Treatment: HIGH



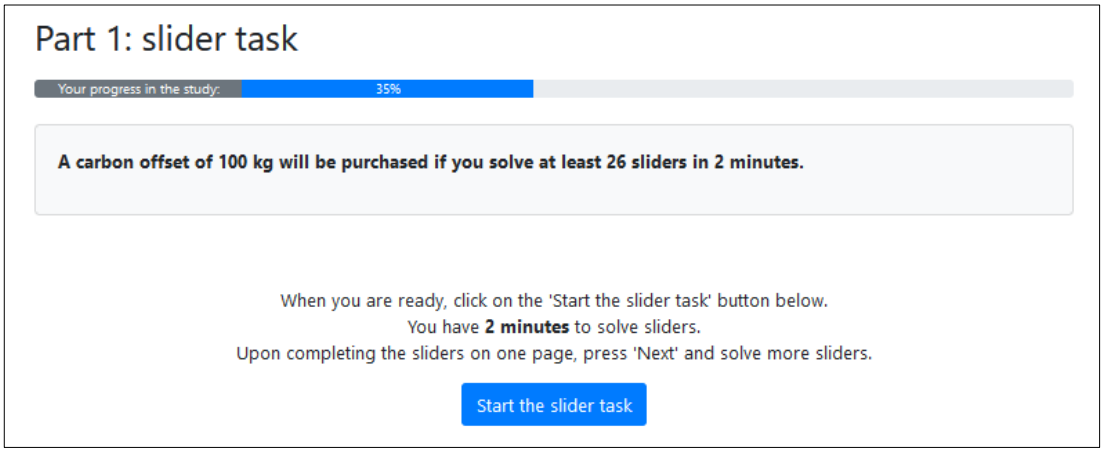
Treatment: SELF



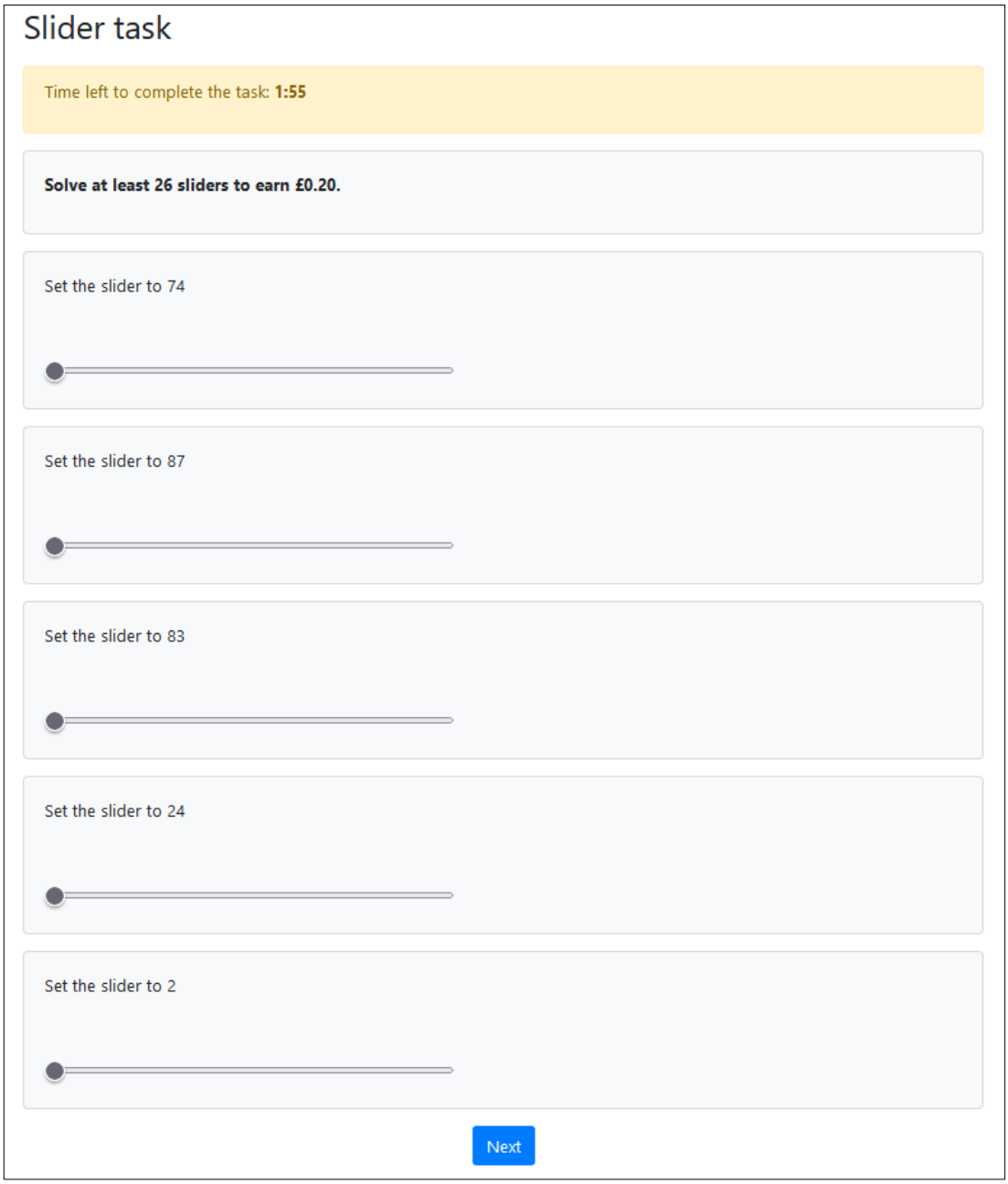
Treatment: LOW



Treatment: HIGH

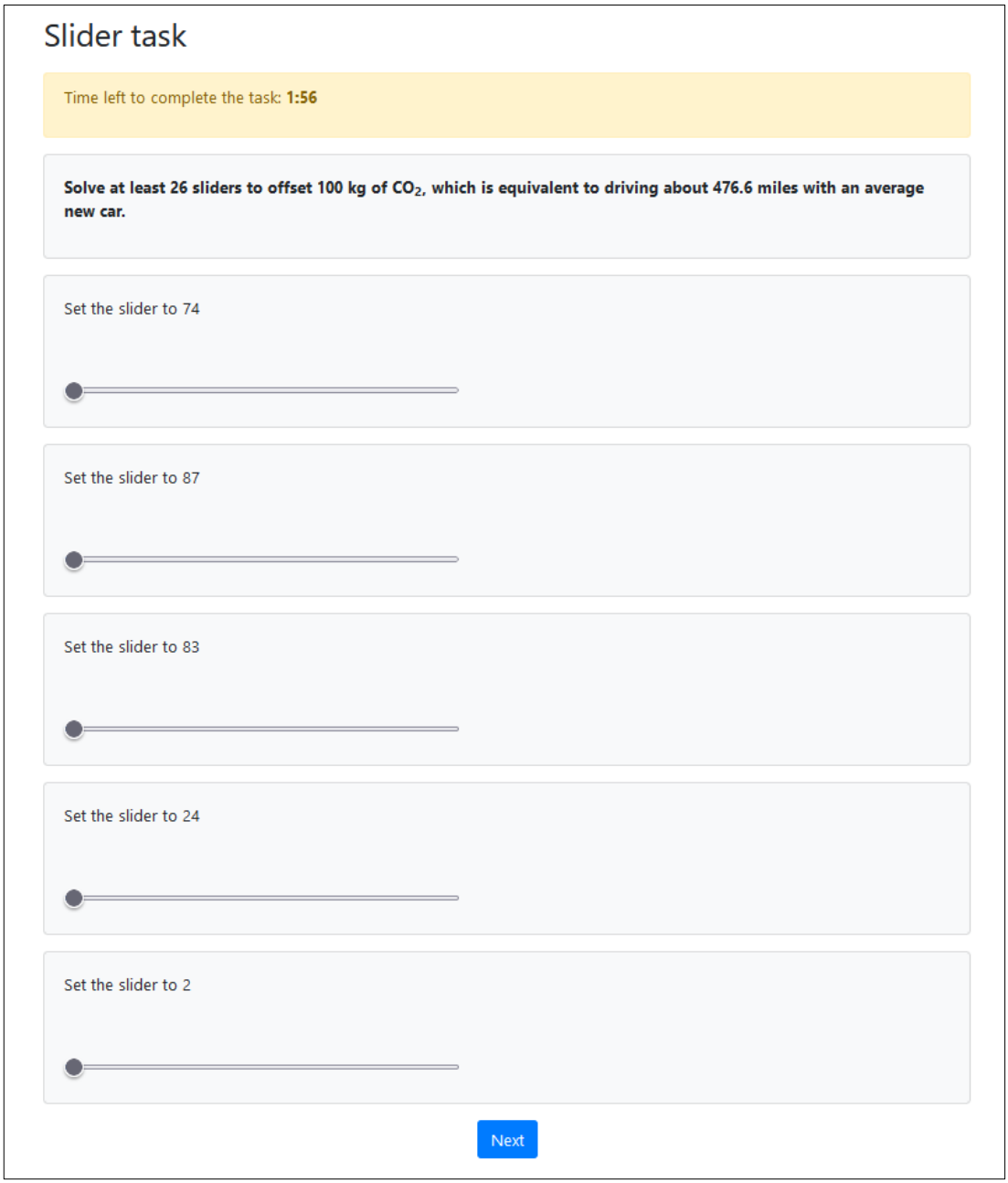


Treatment: SELF

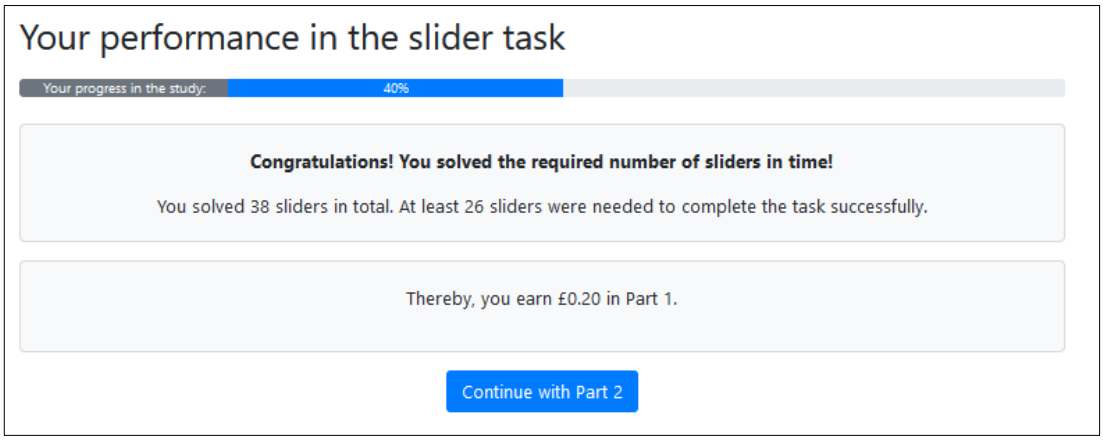


Treatment: LOW

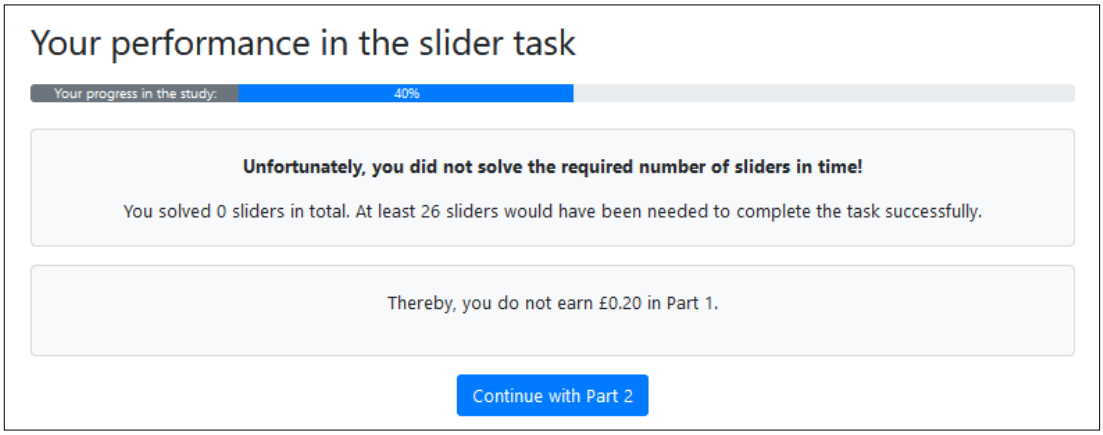
Treatment: HIGH



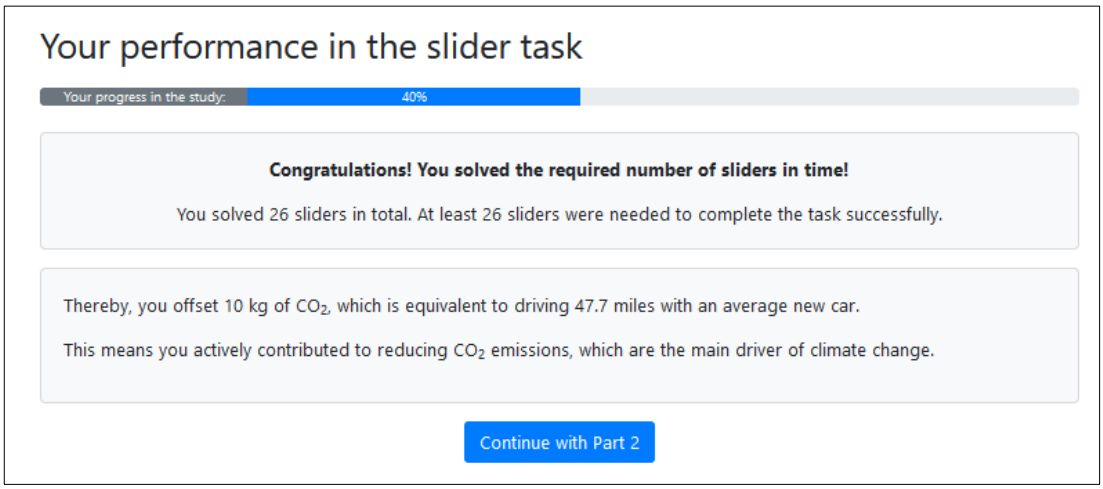
Treatment: SELF, *Success*



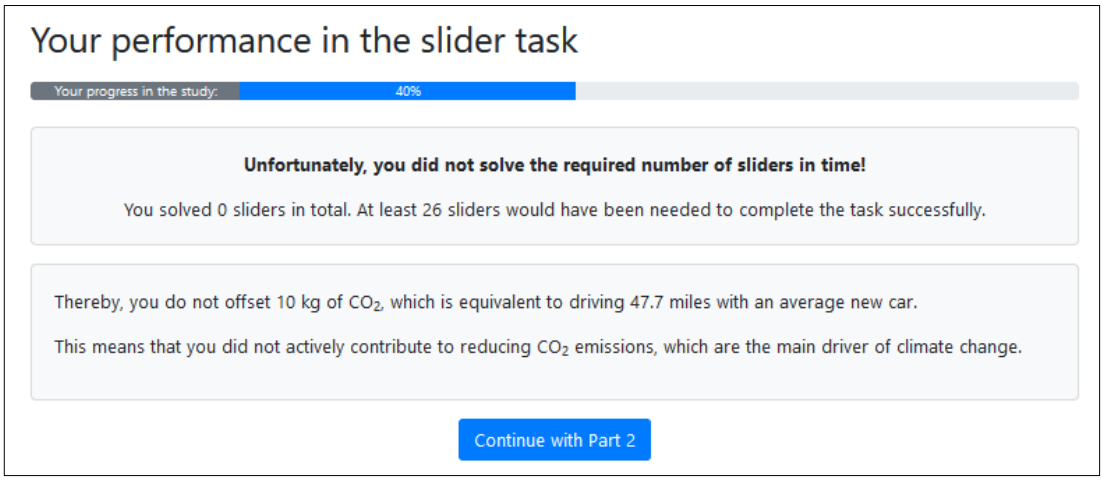
Treatment: SELF, *Failure*



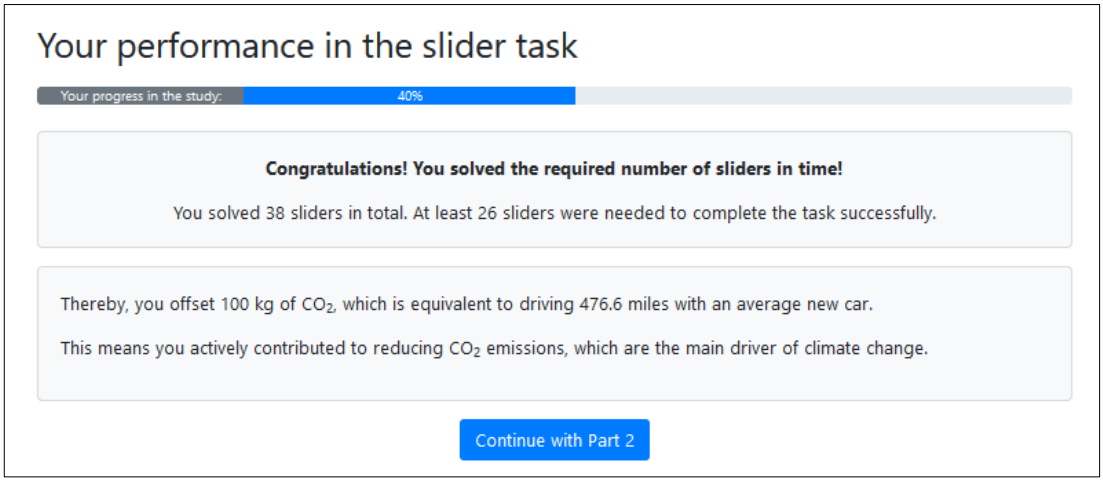
Treatment: LOW, *Success*



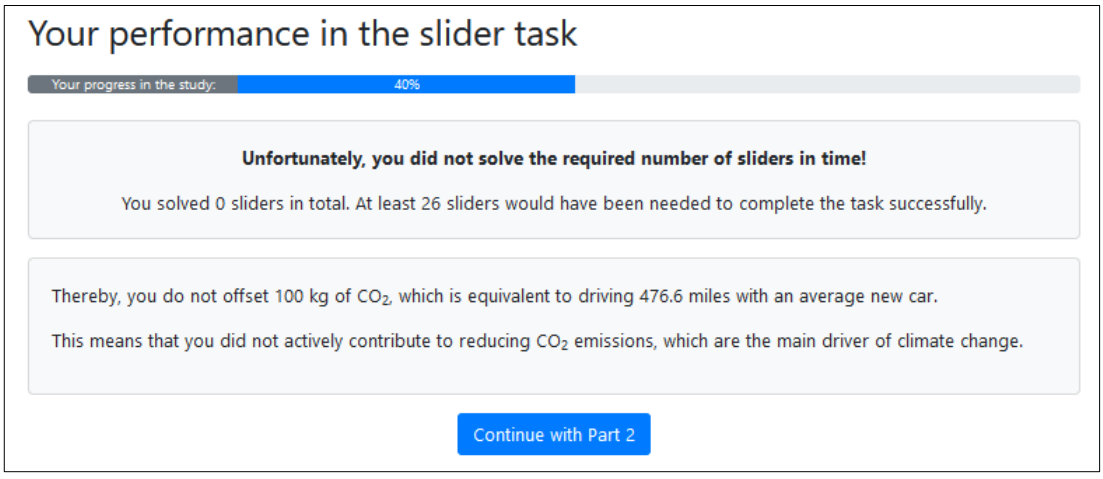
Treatment: LOW, *Failure*

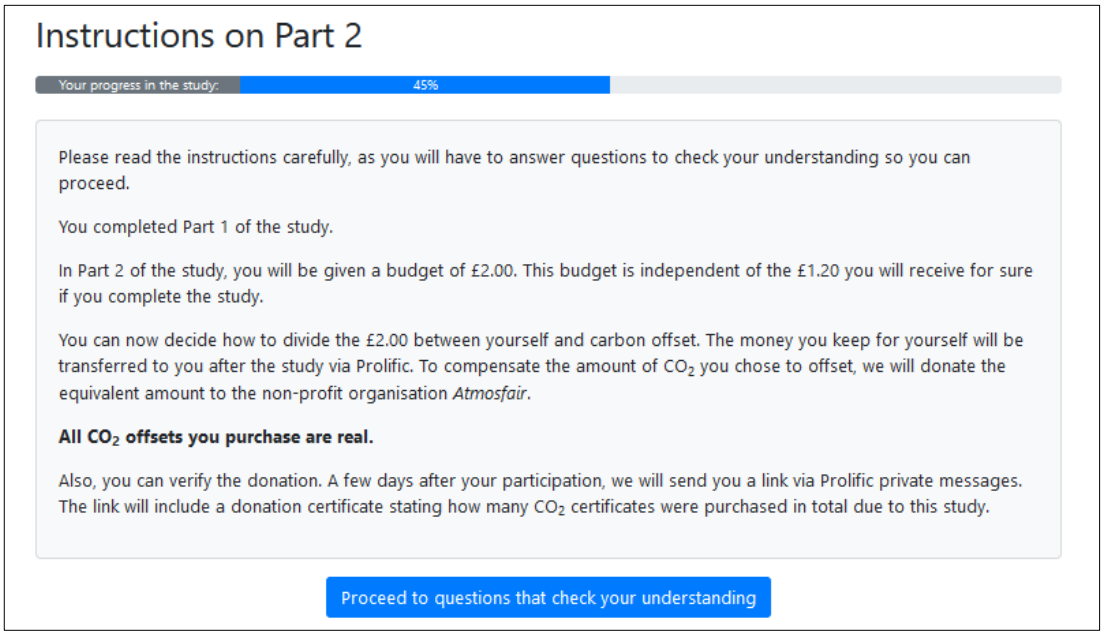


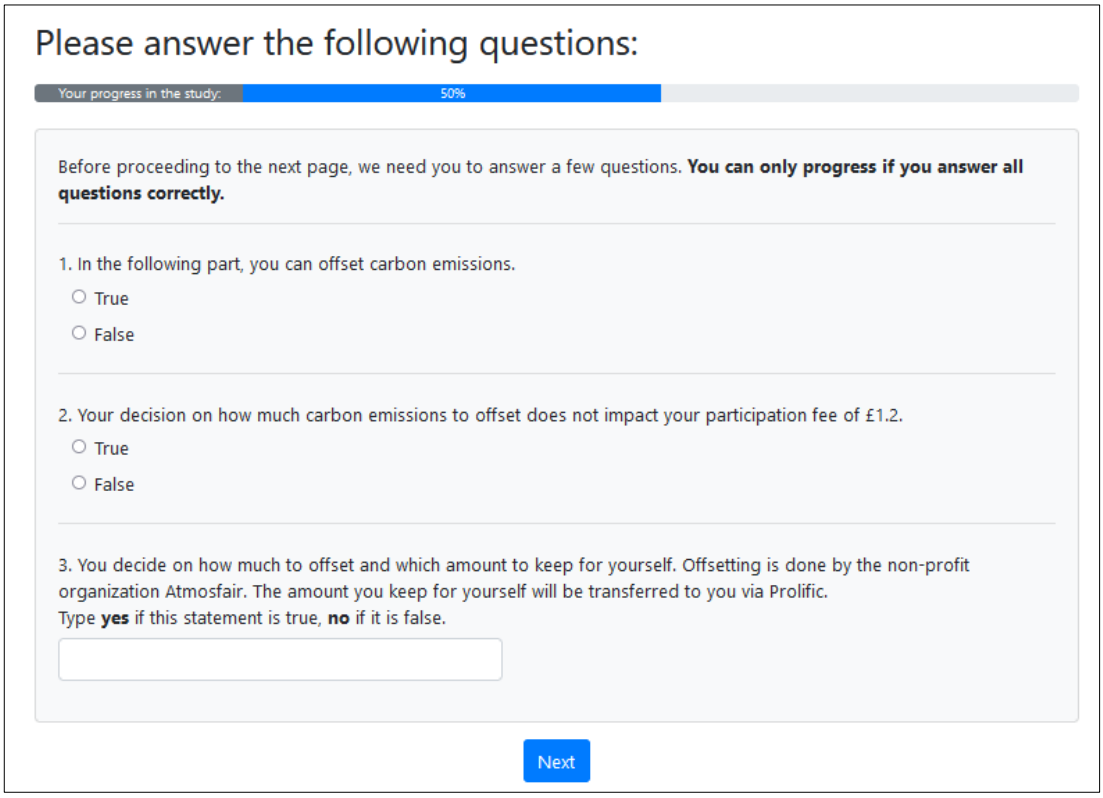
Treatment: HIGH, *Success*

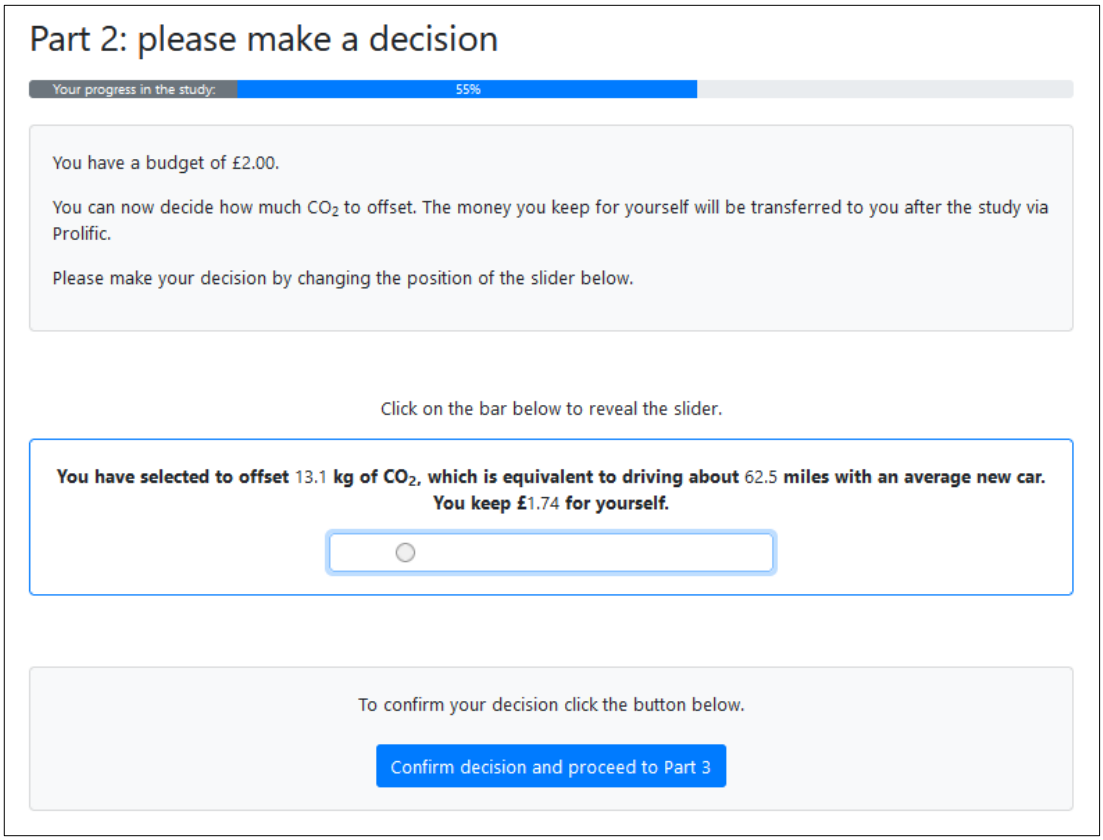


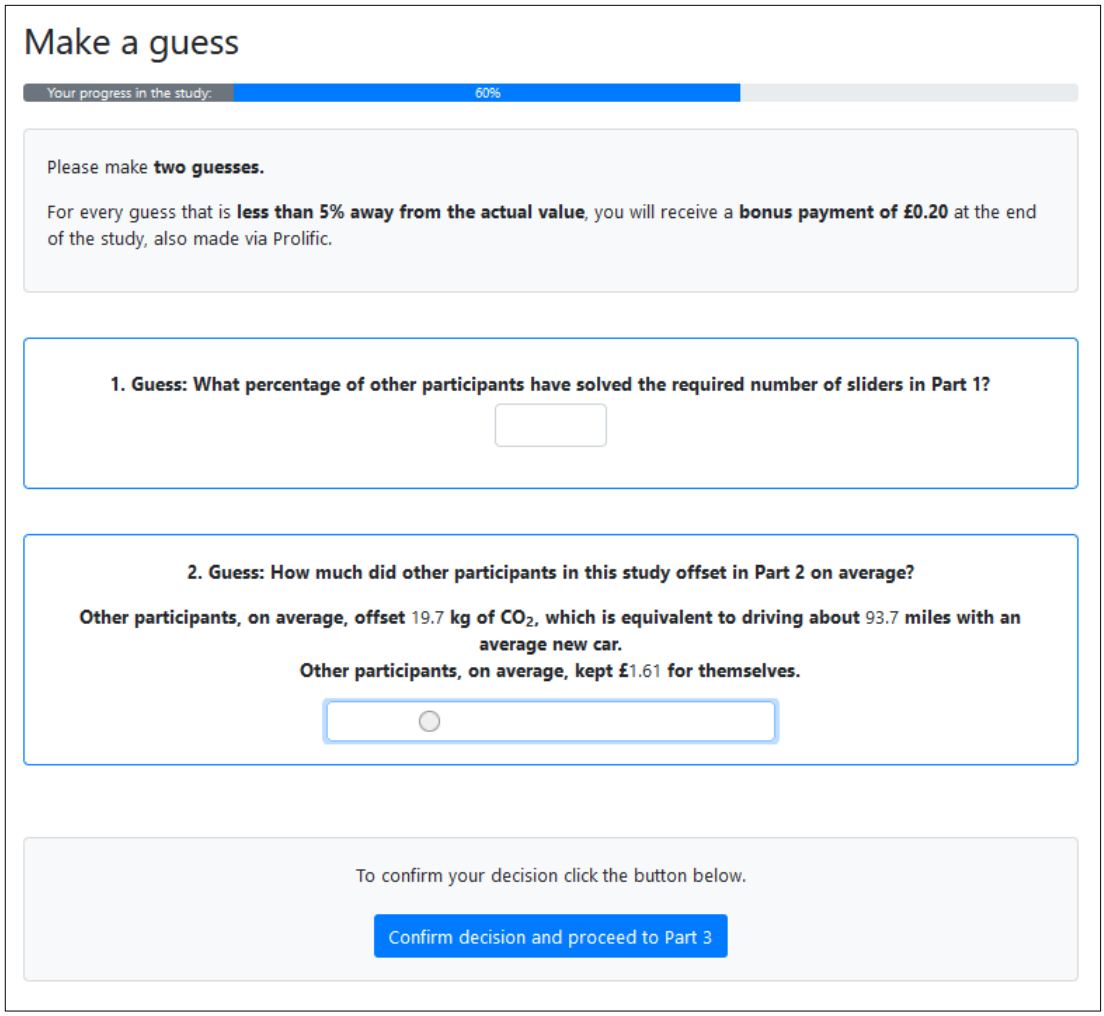
Treatment: HIGH, *Failure*

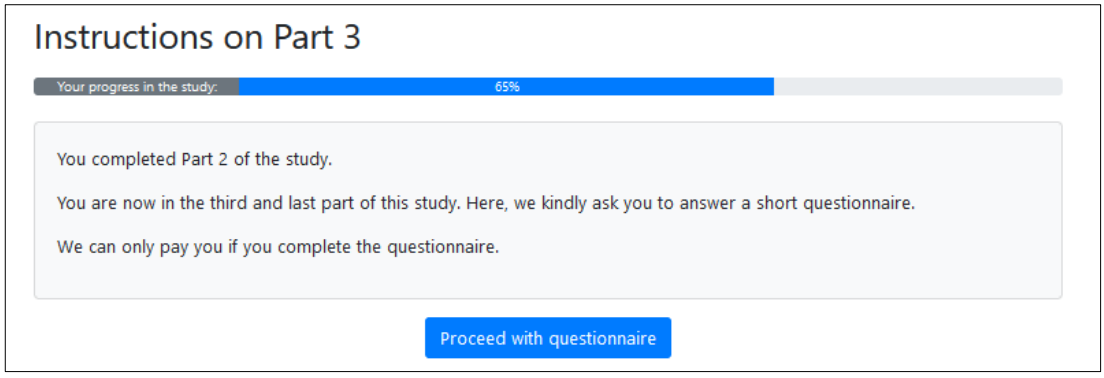


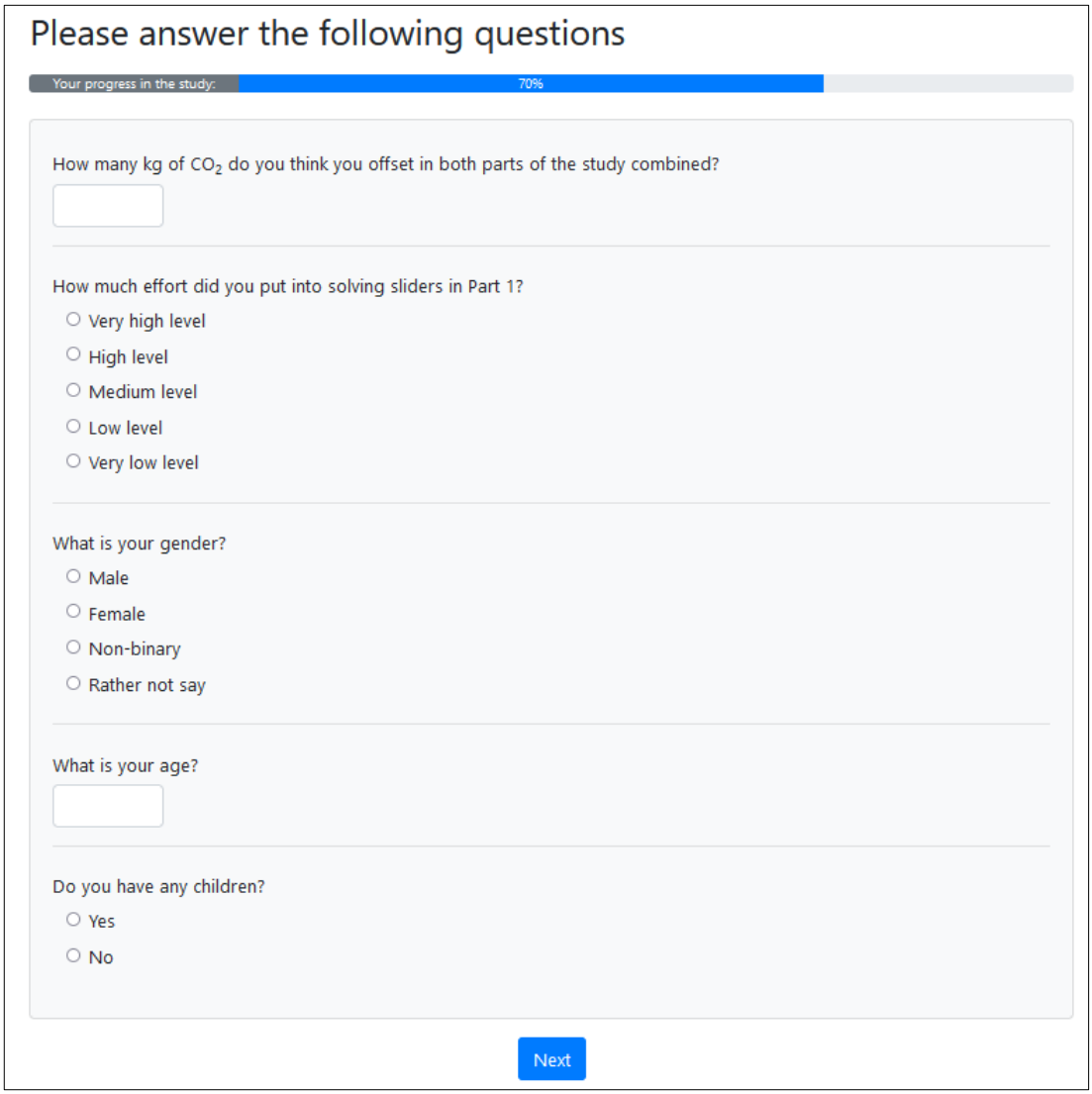

Instructions on Part 2
Your progress in the study:
45%
Please read the instructions carefully, as you will have to answer questions to check your understanding so you can proceed.
You completed Part 1 of the study.
In Part 2 of the study, you will be given a budget of £2.00. This budget is independent of the £1.20 you will receive for sure if you complete the study.
You can now decide how to divide the £2.00 between yourself and carbon offset. The money you keep for yourself will be transferred to you after the study via Prolific. To compensate the amount of CO_2 you chose to offset, we will donate the equivalent amount to the non-profit organisation *Atmosfair*.
All CO_2 offsets you purchase are real.
Also, you can verify the donation. A few days after your participation, we will send you a link via Prolific private messages. The link will include a donation certificate stating how many CO_2 certificates were purchased in total due to this study.
Proceed to questions that check your understanding

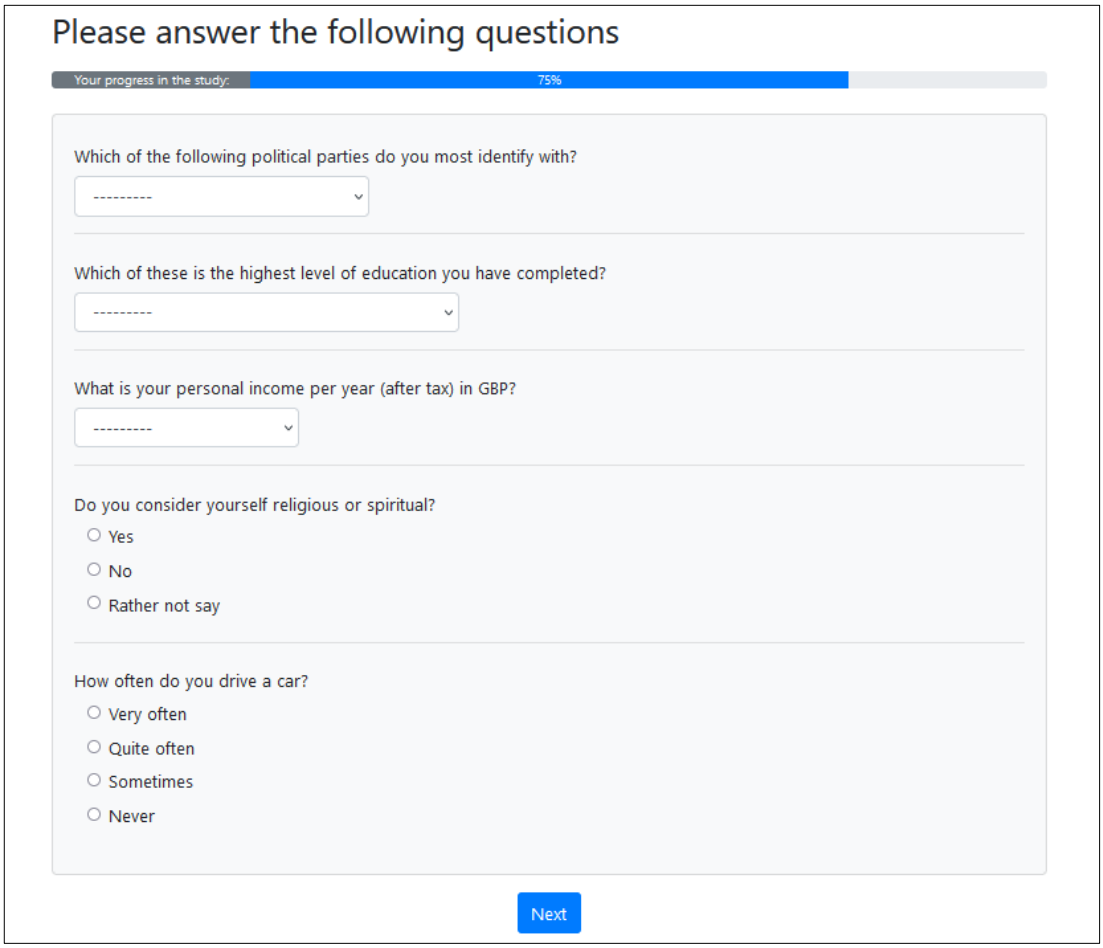

Please answer the following questions:
Your progress in the study:
50%
Before proceeding to the next page, we need you to answer a few questions. **You can only progress if you answer all questions correctly.**
1. In the following part, you can offset carbon emissions.
True
False
2. Your decision on how much carbon emissions to offset does not impact your participation fee of £1.2.
True
False
3. You decide on how much to offset and which amount to keep for yourself. Offsetting is done by the non-profit organization Atmosfair. The amount you keep for yourself will be transferred to you via Prolific.
Type **yes** if this statement is true, **no** if it is false.
Next


Part 2: please make a decision
Your progress in the study:
55%
You have a budget of £2.00.
You can now decide how much CO_2 to offset. The money you keep for yourself will be transferred to you after the study via Prolific.
Please make your decision by changing the position of the slider below.
Click on the bar below to reveal the slider.
You have selected to offset 13.1 kg of CO_2, which is equivalent to driving about 62.5 miles with an average new car.
You keep £1.74 for yourself.
To confirm your decision click the button below.
Confirm decision and proceed to Part 3


Make a guess
Your progress in the study:
60%
Please make two guesses.
For every guess that is less than 5% away from the actual value, you will receive a bonus payment of £0.20 at the end of the study, also made via Prolific.
1. Guess: What percentage of other participants have solved the required number of sliders in Part 1?
2. Guess: How much did other participants in this study offset in Part 2 on average?
Other participants, on average, offset 19.7 kg of CO_2, which is equivalent to driving about 93.7 miles with an average new car.
Other participants, on average, kept £1.61 for themselves.
To confirm your decision click the button below.
Confirm decision and proceed to Part 3


Instructions on Part 3
Your progress in the study:
65%
You completed Part 2 of the study.
You are now in the third and last part of this study. Here, we kindly ask you to answer a short questionnaire.
We can only pay you if you complete the questionnaire.
Proceed with questionnaire

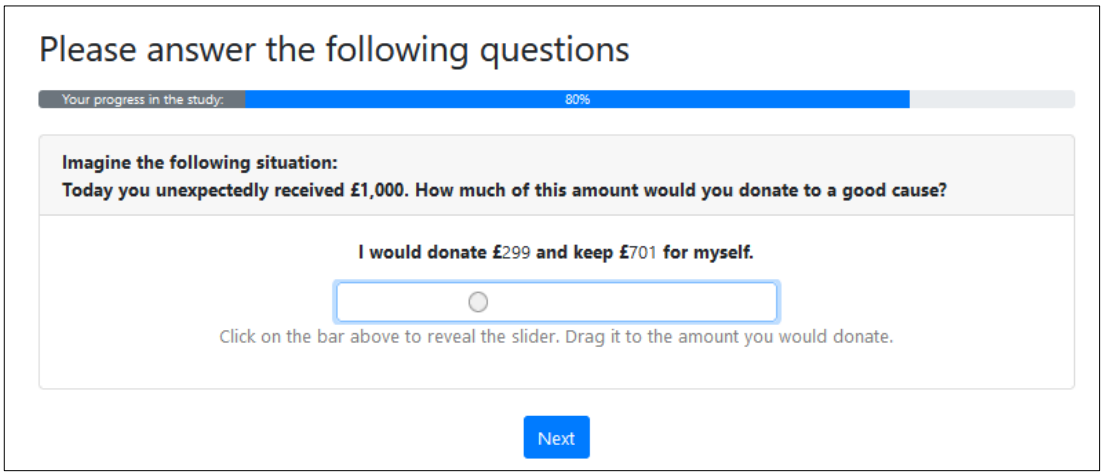

Please answer the following questions
Your progress in the study:
70%
How many kg of CO_2 do you think you offset in both parts of the study combined?
How much effort did you put into solving sliders in Part 1?
Very high level
High level
Medium level
Low level
Very low level
What is your gender?
Male
Female
Non-binary
Rather not say
What is your age?
Do you have any children?
Yes
No
Next

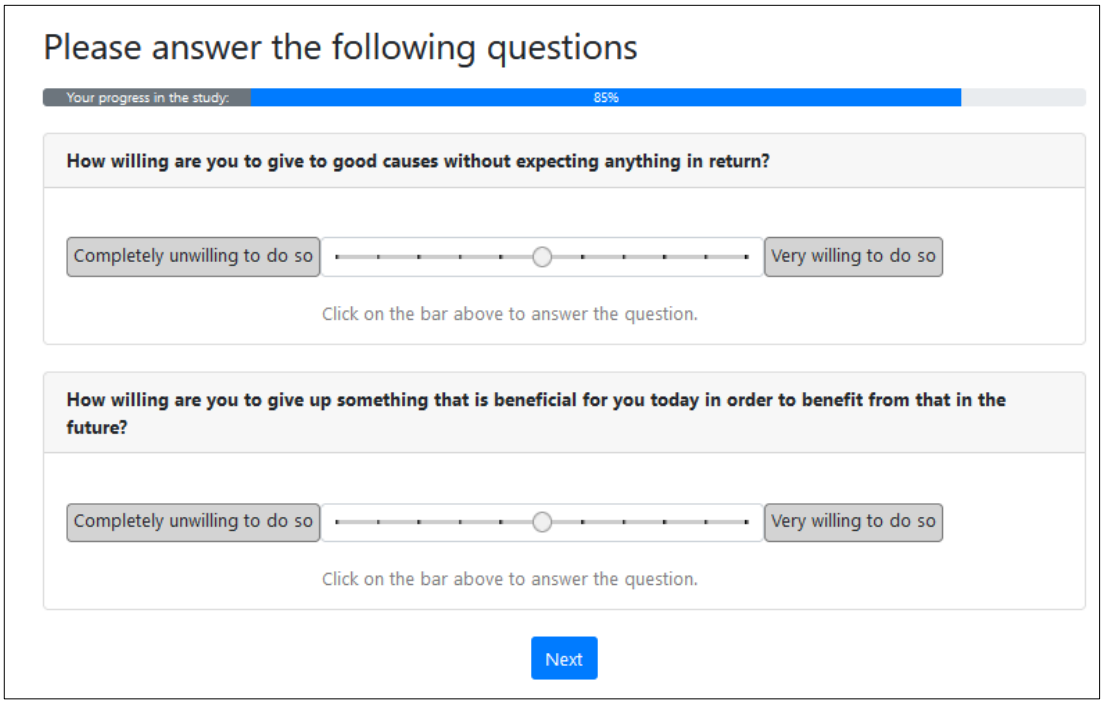

Please answer the following questions
Your progress in the study:
75%
Which of the following political parties do you most identify with?

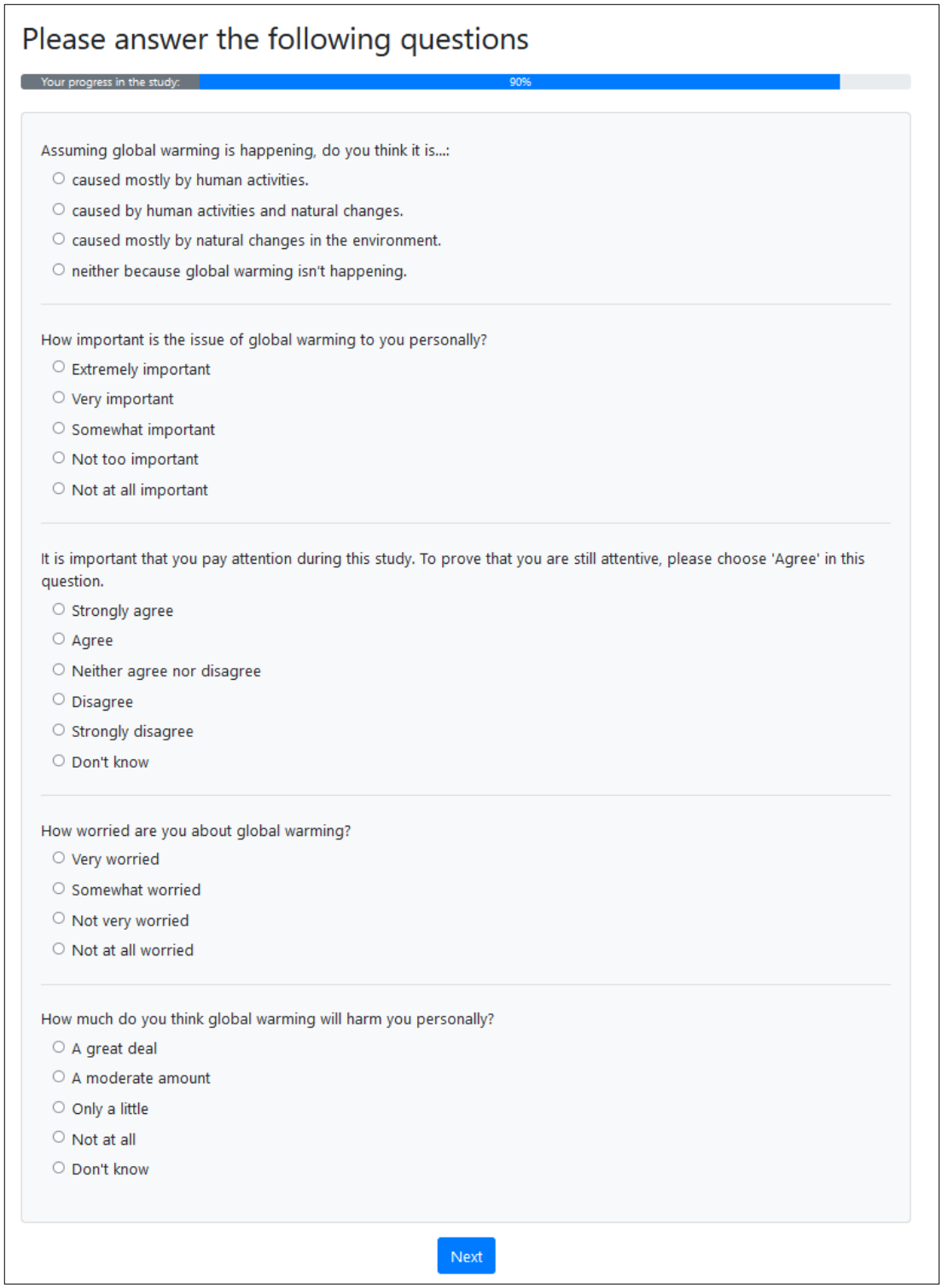
Which of these is the highest level of education you have completed?

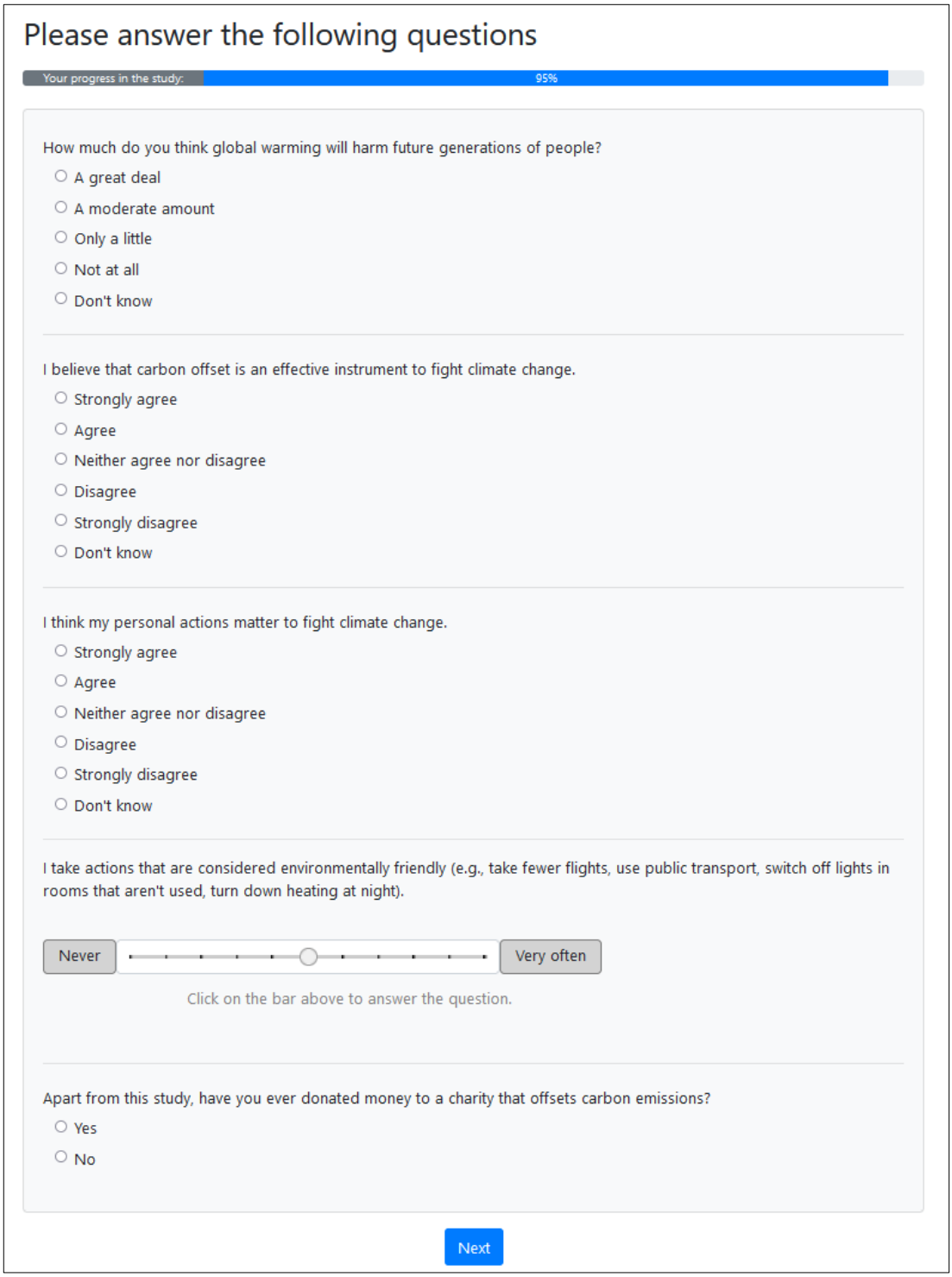
What is your personal income per year (after tax) in GBP?

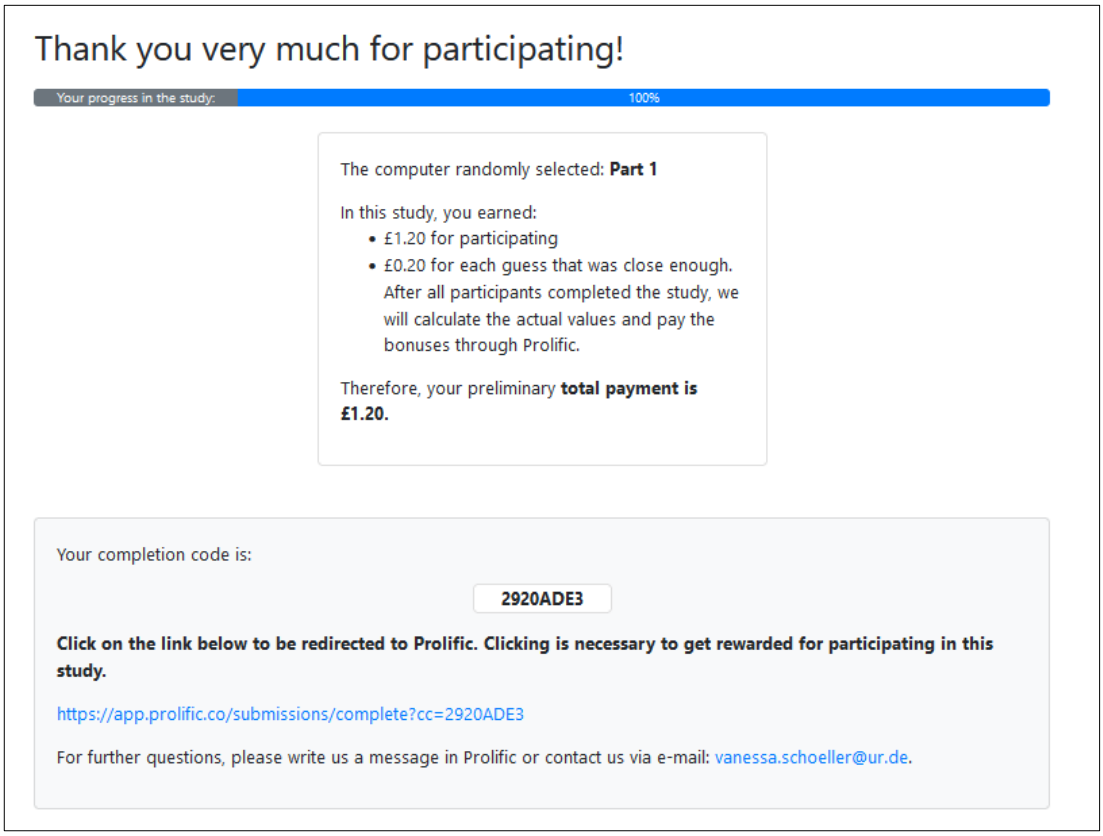
Do you consider yourself religious or spiritual?
Yes
No
Rather not say
How often do you drive a car?
Very often
Quite often
Sometimes
Never
Next


Please answer the following questions
Your progress in the study:
80%
Imagine the following situation:
Today you unexpectedly received £1,000. How much of this amount would you donate to a good cause?
I would donate £299 and keep £701 for myself.
Click on the bar above to reveal the slider. Drag it to the amount you would donate.
Next


Please answer the following questions
Your progress in the study:
85%
How willing are you to give to good causes without expecting anything in return?
Completely unwilling to do so
Very willing to do so
Click on the bar above to answer the question.
How willing are you to give up something that is beneficial for you today in order to benefit from that in the future?
Completely unwilling to do so
Very willing to do so
Click on the bar above to answer the question.
Next


Please answer the following questions
Your progress in the study:
90%
Assuming global warming is happening, do you think it is...:
caused mostly by human activities.
caused by human activities and natural changes.
caused mostly by natural changes in the environment.
neither because global warming isn't happening.
How important is the issue of global warming to you personally?
Extremely important
Very important
Somewhat important
Not too important
Not at all important
It is important that you pay attention during this study. To prove that you are still attentive, please choose 'Agree' in this question.
Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree
Don't know
How worried are you about global warming?
Very worried
Somewhat worried
Not very worried
Not at all worried
How much do you think global warming will harm you personally?
A great deal
A moderate amount
Only a little
Not at all
Don't know
Next


Please answer the following questions
Your progress in the study:
95%
How much do you think global warming will harm future generations of people?
A great deal
A moderate amount
Only a little
Not at all
Don't know
I believe that carbon offset is an effective instrument to fight climate change.
Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree
Don't know
I think my personal actions matter to fight climate change.
Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree
Don't know
I take actions that are considered environmentally friendly (e.g., take fewer flights, use public transport, switch off lights in rooms that aren't used, turn down heating at night).
Never
Very often
Click on the bar above to answer the question.
Apart from this study, have you ever donated money to a charity that offsets carbon emissions?
Yes
No
Next


Thank you very much for participating!
Your progress in the study:
100%
The computer randomly selected: Part 1
In this study, you earned:
£1.20 for participating
£0.20 for each guess that was close enough. After all participants completed the study, we will calculate the actual values and pay the bonuses through Prolific.
Therefore, your preliminary total payment is £1.20.
Your completion code is:
2920ADE3
Click on the link below to be redirected to Prolific. Clicking is necessary to get rewarded for participating in this study.
https://app.prolific.co/submissions/complete?cc=2920ADE3
For further questions, please write us a message in Prolific or contact us via e-mail: vanessa.schoeller@ur.de.

Appendix C

Gender Differences

C.1 Summary statistics

Table C.1: Descriptives for the cisgender and transgender samples.

		Gender		
	Total	Cisgender	Transgender	p-value
	(N=780)	(N=425)	(N=355)	
Treatment				0.933
NEUTRAL	259 (33.2%)	143 (33.6%)	116 (32.7%)	
FEMININE	263 (33.7%)	141 (33.2%)	122 (34.4%)	
MASCULINE	258 (33.1%)	141 (33.2%)	117 (33.0%)	
Age (years)				0.516
Mean (SD)	24.4 (6.60)	24.3 (6.52)	24.6 (6.71)	
Height (cm)				0.002
Mean (SD)	170 (10.8)	171 (11.0)	169 (10.5)	
Student status				0.830
Yes	368 (47.2%)	202 (47.5%)	166 (46.8%)	
No	412 (52.8%)	223 (52.5%)	189 (53.2%)	
Highest education				0.094
University degree	266 (34.1%)	159 (37.4%)	107 (30.1%)	
High school diploma/A-levels	361 (46.3%)	189 (44.5%)	172 (48.5%)	
Other	153 (19.6%)	77 (18.1%)	76 (21.4%)	
Income: Less than 20,000 GBP				0.171
Yes	541 (69.4%)	286 (67.3%)	255 (71.8%)	
No	239 (30.6%)	139 (32.7%)	100 (28.2%)	
Religion				0.891
Non-religious	547 (70.1%)	295 (69.4%)	252 (71.0%)	
Religious	201 (25.8%)	112 (26.4%)	89 (25.1%)	
Not say	32 (4.1%)	18 (4.2%)	14 (3.9%)	
Residence				<0.001
Continental Europe	250 (32.1%)	169 (39.8%)	81 (22.8%)	
United Kingdom	205 (26.3%)	101 (23.8%)	104 (29.3%)	
United States	265 (34.0%)	133 (31.3%)	132 (37.2%)	
Other	60 (7.7%)	22 (5.2%)	38 (10.7%)	
BEM group:				0.002
Androgynous	188 (24.1%)	116 (27.3%)	72 (20.3%)	
Feminine	222 (28.5%)	104 (24.5%)	118 (33.2%)	
Masculine	151 (19.4%)	95 (22.4%)	56 (15.8%)	
Undifferentiated	219 (28.1%)	110 (25.9%)	109 (30.7%)	
BEM score: Feminine				0.730
Mean (SD)	41.8 (8.58)	41.8 (8.19)	41.7 (9.03)	
BEM score: Masculine				<0.001
Mean (SD)	33.9 (7.95)	35.0 (7.64)	32.5 (8.11)	
TCS				<0.001
Mean (SD)	3.67 (1.14)	4.47 (0.570)	2.71 (0.865)	
STT				<0.001
Mean (SD)	4.35 (4.59)	0.998 (1.47)	8.37 (3.76)	

Note: The table summarizes the characteristics of the cisgender and transgender samples. The education category other includes subjects that replied technical/community college, secondary education (e.g. GED/GCSE), no formal qualification, or don't know/not applicable. The religion category religious includes subjects that replied Buddhism, Christianity, Hinduism, Islam, Judaism, Paganism, Sikhism, or Spiritualism. The residence category other includes subjects that replied Australia, Canada, Chile, Israel, Japan, Mexico, New Zealand, or South Africa. The column p-value reports the p-values of χ^2-tests for categorical variables and the p-values of Wilcoxon-Mann Whitney tests for numerical variables between the cisgender and transgender column.

Table C.2: Descriptives by treatment for cismen.

		Treatment			
	Total	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen					
	(N=214)	(N=72)	(N=71)	(N=71)	
Age (years)					0.042
Mean (SD)	24.1 (5.74)	25.8 (7.70)	24.1 (4.79)	22.5 (3.44)	
Height (cm)					0.449
Mean (SD)	178 (9.08)	180 (10.2)	177 (7.95)	177 (8.77)	
Student status					0.754
Yes	102 (47.7%)	32 (44.4%)	34 (47.9%)	36 (50.7%)	
No	112 (52.3%)	40 (55.6%)	37 (52.1%)	35 (49.3%)	
Highest education					0.237
University degree	72 (33.6%)	26 (36.1%)	27 (38.0%)	19 (26.8%)	
High school diploma/A-levels	94 (43.9%)	27 (37.5%)	28 (39.4%)	39 (54.9%)	
Other	48 (22.4%)	19 (26.4%)	16 (22.5%)	13 (18.3%)	
Income: Less than 20,000 GBP					0.841
Yes	135 (63.1%)	47 (65.3%)	43 (60.6%)	45 (63.4%)	
No	79 (36.9%)	25 (34.7%)	28 (39.4%)	26 (36.6%)	
Religion					0.820
Non-religious	144 (67.3%)	48 (66.7%)	47 (66.2%)	49 (69.0%)	
Religious	60 (28.0%)	21 (29.2%)	19 (26.8%)	20 (28.2%)	
Not say	10 (4.7%)	3 (4.2%)	5 (7.0%)	2 (2.8%)	
Residence					0.972
Continental Europe	95 (44.4%)	31 (43.1%)	30 (42.3%)	34 (47.9%)	
United Kingdom	50 (23.4%)	17 (23.6%)	17 (23.9%)	16 (22.5%)	
United States	65 (30.4%)	23 (31.9%)	23 (32.4%)	19 (26.8%)	
Other	4 (1.9%)	1 (1.4%)	1 (1.4%)	2 (2.8%)	
BEM group:					0.490
Androgynous	56 (26.2%)	19 (26.4%)	19 (26.8%)	18 (25.4%)	
Feminine	40 (18.7%)	15 (20.8%)	11 (15.5%)	14 (19.7%)	
Masculine	59 (27.6%)	22 (30.6%)	15 (21.1%)	22 (31.0%)	
Undifferentiated	59 (27.6%)	16 (22.2%)	26 (36.6%)	17 (23.9%)	
BEM score: Feminine					0.644
Mean (SD)	40.0 (8.55)	40.4 (8.52)	39.3 (9.13)	40.4 (8.03)	
BEM score: Masculine					0.522
Mean (SD)	35.7 (7.68)	36.2 (6.93)	34.9 (7.71)	36.0 (8.41)	
TCS					0.620
Mean (SD)	4.47 (0.591)	4.45 (0.541)	4.49 (0.590)	4.46 (0.645)	
STT					0.001
Mean (SD)	0.986 (1.46)	1.18 (1.09)	0.535 (0.939)	1.24 (2.01)	

Note: The table summarizes the characteristics of the cisgender and transgender samples. The education category other includes subjects that replied technical/community college, secondary education (e.g. GED/GCSE), no formal qualification, or don't know/not applicable. The religion category religious includes subjects that replied Buddhism, Christianity, Hinduism, Islam, Judaism, Paganism, Sikhism, or Spiritualism. The residence category other includes subjects that replied Australia, Canada, Chile, Israel, Japan, Mexico, New Zealand, or South Africa. The column p-value reports the p-values of χ^2-tests for categorical variables and the p-values of the Kruskal Wallis test for numerical variables between the treatment columns.

Table C.3: Descriptives by treatment for ciswomen.

		Treatment			
	Total	NEUTRAL	FEMININE	MASCULINE	p-value
Ciswomen					
	(N=211)	(N=71)	(N=70)	(N=70)	
Age (years)					0.644
Mean (SD)	24.6 (7.23)	25.0 (7.83)	25.1 (7.79)	23.6 (5.90)	
Height (cm)					0.541
Mean (SD)	164 (7.96)	164 (9.88)	164 (7.02)	165 (6.60)	
Student status					0.813
Yes	100 (47.4%)	35 (49.3%)	34 (48.6%)	31 (44.3%)	
No	111 (52.6%)	36 (50.7%)	36 (51.4%)	39 (55.7%)	
Highest education					0.667
University degree	87 (41.2%)	32 (45.1%)	31 (44.3%)	24 (34.3%)	
High school diploma/A-levels	95 (45.0%)	29 (40.8%)	31 (44.3%)	35 (50.0%)	
Other	29 (13.7%)	10 (14.1%)	8 (11.4%)	11 (15.7%)	
Income: Less than 20,000 GBP					0.253
Yes	151 (71.6%)	53 (74.6%)	53 (75.7%)	45 (64.3%)	
No	60 (28.4%)	18 (25.4%)	17 (24.3%)	25 (35.7%)	
Religion					0.990
Non-religious	151 (71.6%)	51 (71.8%)	50 (71.4%)	50 (71.4%)	
Religious	52 (24.6%)	17 (23.9%)	17 (24.3%)	18 (25.7%)	
Not say	8 (3.8%)	3 (4.2%)	3 (4.3%)	2 (2.9%)	
Residence					0.589
Continental Europe	74 (35.1%)	28 (39.4%)	25 (35.7%)	21 (30.0%)	
United Kingdom	51 (24.2%)	19 (26.8%)	15 (21.4%)	17 (24.3%)	
United States	68 (32.2%)	21 (29.6%)	24 (34.3%)	23 (32.9%)	
Other	18 (8.5%)	3 (4.2%)	6 (8.6%)	9 (12.9%)	
BEM group:					0.187
Androgynous	60 (28.4%)	17 (23.9%)	28 (40.0%)	15 (21.4%)	
Feminine	64 (30.3%)	23 (32.4%)	19 (27.1%)	22 (31.4%)	
Masculine	36 (17.1%)	14 (19.7%)	11 (15.7%)	11 (15.7%)	
Undifferentiated	51 (24.2%)	17 (23.9%)	12 (17.1%)	22 (31.4%)	
BEM score: Feminine					0.212
Mean (SD)	43.5 (7.42)	43.3 (7.35)	44.7 (7.66)	42.6 (7.22)	
BEM score: Masculine					0.099
Mean (SD)	34.3 (7.54)	33.9 (7.31)	35.9 (8.18)	33.2 (6.93)	
TCS					0.878
Mean (SD)	4.48 (0.550)	4.55 (0.413)	4.47 (0.585)	4.42 (0.630)	
STT					0.906
Mean (SD)	1.01 (1.49)	1.15 (2.07)	0.957 (1.04)	0.914 (1.14)	

Note: The table summarizes the characteristics of the cisgender and transgender samples. The education category other includes subjects that replied technical/community college, secondary education (e.g. GED/GCSE), no formal qualification, or don't know/not applicable. The religion category religious includes subjects that replied Buddhism, Christianity, Hinduism, Islam, Judaism, Paganism, Sikhism, or Spiritualism. The residence category other includes subjects that replied Australia, Canada, Chile, Israel, Japan, Mexico, New Zealand, or South Africa. The column p-value reports the p-values of χ^2-tests for categorical variables and the p-values of the Kruskal Wallis test for numerical variables between the treatment columns.

Table C.4: Descriptives by treatment for transmen.

		Treatment			
	Total	NEUTRAL	FEMININE	MASCULINE	p-value
Transmen					
	(N=215)	(N=72)	(N=72)	(N=71)	
Age (years)					0.775
Mean (SD)	24.3 (6.40)	25.1 (8.07)	24.0 (5.61)	23.7 (5.12)	
Height (cm)					0.301
Mean (SD)	164 (8.52)	165 (10.4)	164 (7.32)	164 (7.64)	
Student status					0.376
Yes	108 (50.2%)	34 (47.2%)	41 (56.9%)	33 (46.5%)	
No	107 (49.8%)	38 (52.8%)	31 (43.1%)	38 (53.5%)	
Highest education					0.891
University degree	63 (29.3%)	22 (30.6%)	19 (26.4%)	22 (31.0%)	
High school diploma/A-levels	109 (50.7%)	37 (51.4%)	39 (54.2%)	33 (46.5%)	
Other	43 (20.0%)	13 (18.1%)	14 (19.4%)	16 (22.5%)	
Income: Less than 20,000 GBP					0.355
Yes	155 (72.1%)	48 (66.7%)	52 (72.2%)	55 (77.5%)	
No	60 (27.9%)	24 (33.3%)	20 (27.8%)	16 (22.5%)	
Religion					0.892
Non-religious	144 (67.0%)	47 (65.3%)	49 (68.1%)	48 (67.6%)	
Religious	62 (28.8%)	23 (31.9%)	19 (26.4%)	20 (28.2%)	
Not say	9 (4.2%)	2 (2.8%)	4 (5.6%)	3 (4.2%)	
Residence					0.939
Continental Europe	47 (21.9%)	14 (19.4%)	17 (23.6%)	16 (22.5%)	
United Kingdom	64 (29.8%)	22 (30.6%)	23 (31.9%)	19 (26.8%)	
United States	85 (39.5%)	31 (43.1%)	25 (34.7%)	29 (40.8%)	
Other	19 (8.8%)	5 (6.9%)	7 (9.7%)	7 (9.9%)	
BEM group:					0.927
Androgynous	44 (20.5%)	17 (23.6%)	13 (18.1%)	14 (19.7%)	
Feminine	69 (32.1%)	20 (27.8%)	24 (33.3%)	25 (35.2%)	
Masculine	42 (19.5%)	16 (22.2%)	13 (18.1%)	13 (18.3%)	
Undifferentiated	60 (27.9%)	19 (26.4%)	22 (30.6%)	19 (26.8%)	
BEM score: Feminine					0.809
Mean (SD)	41.4 (8.90)	41.1 (8.18)	41.5 (9.47)	41.6 (9.13)	
BEM score: Masculine					0.597
Mean (SD)	33.6 (7.43)	34.0 (7.42)	33.1 (6.98)	33.6 (7.95)	
TCS					0.692
Mean (SD)	2.82 (0.868)	2.88 (0.946)	2.75 (0.857)	2.84 (0.800)	
STT					0.910
Mean (SD)	9.26 (3.15)	9.29 (3.50)	9.21 (3.01)	9.27 (2.96)	

Note: The table summarizes the characteristics of the cisgender and transgender samples. The education category other includes subjects that replied technical/community college, secondary education (e.g. GED/GCSE), no formal qualification, or don't know/not applicable. The religion category religious includes subjects that replied Buddhism, Christianity, Hinduism, Islam, Judaism, Paganism, Sikhism, or Spiritualism. The residence category other includes subjects that replied Australia, Canada, Chile, Israel, Japan, Mexico, New Zealand, or South Africa. The column p-value reports the p-values of χ^2-tests for categorical variables and the p-values of the Kruskal Wallis test for numerical variables between the treatment columns.

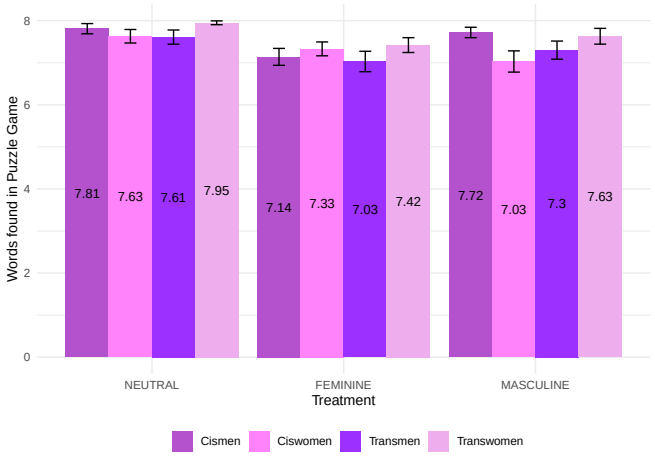
Table C.5: Descriptives by treatment for transwomen.

		Treatment			
	Total	NEUTRAL	FEMININE	MASCULINE	p-value
Transwomen					
	(N=140)	(N=44)	(N=50)	(N=46)	
Age (years)					0.345
Mean (SD)	25.1 (7.15)	25.3 (5.91)	25.6 (9.01)	24.2 (5.89)	
Height (cm)					0.864
Mean (SD)	175 (10.1)	176 (8.41)	174 (13.4)	175 (7.24)	
Student status					0.939
Yes	58 (41.4%)	18 (40.9%)	20 (40.0%)	20 (43.5%)	
No	82 (58.6%)	26 (59.1%)	30 (60.0%)	26 (56.5%)	
Highest education					0.090
University degree	44 (31.4%)	13 (29.5%)	20 (40.0%)	11 (23.9%)	
High school diploma/A-levels	63 (45.0%)	16 (36.4%)	20 (40.0%)	27 (58.7%)	
Other	33 (23.6%)	15 (34.1%)	10 (20.0%)	8 (17.4%)	
Income: Less than 20,000 GBP					0.070
Yes	100 (71.4%)	37 (84.1%)	34 (68.0%)	29 (63.0%)	
No	40 (28.6%)	7 (15.9%)	16 (32.0%)	17 (37.0%)	
Religion					0.664
Non-religious	108 (77.1%)	33 (75.0%)	39 (78.0%)	36 (78.3%)	
Religious	27 (19.3%)	10 (22.7%)	10 (20.0%)	7 (15.2%)	
Not say	5 (3.6%)	1 (2.3%)	1 (2.0%)	3 (6.5%)	
Residence					0.257
Continental Europe	34 (24.3%)	11 (25.0%)	9 (18.0%)	14 (30.4%)	
United Kingdom	40 (28.6%)	11 (25.0%)	17 (34.0%)	12 (26.1%)	
United States	47 (33.6%)	13 (29.5%)	21 (42.0%)	13 (28.3%)	
Other	19 (13.6%)	9 (20.5%)	3 (6.0%)	7 (15.2%)	
BEM group:					0.333
Androgynous	28 (20.0%)	5 (11.4%)	12 (24.0%)	11 (23.9%)	
Feminine	49 (35.0%)	19 (43.2%)	17 (34.0%)	13 (28.3%)	
Masculine	14 (10.0%)	7 (15.9%)	3 (6.0%)	4 (8.7%)	
Undifferentiated	49 (35.0%)	13 (29.5%)	18 (36.0%)	18 (39.1%)	
BEM score: Feminine					0.973
Mean (SD)	42.3 (9.22)	42.8 (7.72)	41.3 (10.4)	42.8 (9.28)	
BEM score: Masculine					0.996
Mean (SD)	30.9 (8.83)	31.0 (9.17)	30.8 (8.75)	30.9 (8.80)	
TCS					0.745
Mean (SD)	2.54 (0.835)	2.54 (0.755)	2.55 (0.872)	2.52 (0.883)	
STT					0.027
Mean (SD)	7.01 (4.19)	7.93 (4.05)	5.80 (4.35)	7.46 (3.91)	

Note: The table summarizes the characteristics of the cisgender and transgender samples. The education category other includes subjects that replied technical/community college, secondary education (e.g. GED/GCSE), no formal qualification, or don't know/not applicable. The religion category religious includes subjects that replied Buddhism, Christianity, Hinduism, Islam, Judaism, Paganism, Sikhism, or Spiritualism. The residence category other includes subjects that replied Australia, Canada, Chile, Israel, Japan, Mexico, New Zealand, or South Africa. The column p-value reports the p-values of χ^2-tests for categorical variables and the p-values of the Kruskal Wallis test for numerical variables between the treatment columns.

C.2 Priming (Part 1)

Figure C.1 presents the number of marked words divided by treatment and subject group. We do not find any differences in marked words within one priming condition across subject groups (KW, NEUTRAL: $p = 0.349$, FEMININE: $p = 0.874$, MASCULINE: $p = 0.112$). For the different subject groups separately across priming conditions, only the number of words marked by transmen did not differ across priming conditions (KW, cismen: $p < 0.001$, ciswomen: $p = 0.038$, transmen: $p = 0.123$, transwomen: $p = 0.014$). Concerning gender differences, we do not see significant variations (MWU, $p < 0.675$). The same is true for sex difference (MWU, $p < 0.060$). As we did not pre-register to control for the number of words marked in our regressions, we do not add this variable in the reported analysis. However, please note that all main results remain qualitatively the same when we account for the heterogeneity in the number of marked words. The additional analyses are available on request.



Note: The bars show the average amount of marked words, and the error bars represent the standard error of the mean.

Figure C.1: Marked words in Part 1 by treatment and subject groups in alphabetical order ($n = 780$).

Table C.6: Words found in the priming task across treatments and subject groups.

Panel A: Priming across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	7.806	7.141	7.718	<0.001
Ciswomen	7.634	7.329	7.029	0.038
Transmen	7.611	7.028	7.296	0.123
Transwomen	7.955	7.420	7.630	0.014

Panel B: Priming across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	7.806	7.634	7.611	7.955	0.349
FEMININE	7.141	7.329	7.028	7.420	0.874
MASCULINE	7.718	7.029	7.296	7.630	0.112

Panel C: Priming across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	7.720	Transgender	7.741	0.816
Case 2	Cismen	7.806	Ciswomen	7.634	0.339
Case 3	Transmen	7.611	Transwomen	7.955	0.122
Case 4	Female	7.622	Male	7.862	0.091
Case 5	Feminine	7.757	Masculine	7.708	0.820

Panel D: Priming in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	7.806	7.141	<0.001	7.718	0.345
Ciswomen	7.634	7.329	0.012	7.029	0.035
Transmen	7.611	7.028	0.040	7.296	0.207
Transwomen	7.955	7.420	0.004	7.630	0.100

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.3 Performance in the real effort math task (Part 2, 3, and 4)

The following tables summarize the performance in the math task by treatment and subject groups for Part 2 (Table C.7) and Part 3 (Table C.8). By treatment, ciswomen and cismen have differences in performance in MASCULINE in Part 2 (MWU, NEUTRAL: $p = 0.080$, FEMININE: $p = 0.205$, MASCULINE: $p = 0.037$) and across all treatments in Part 3 (MWU, NEUTRAL: $p = 0.004$, Part 3 FEMININE: $p = 0.010$, Part 3 MASCULINE: $p = 0.028$). Also, Buser et al. (2021) (which set up the online version of this math task) does find gender differences in performance. Thomas Buser states in a personal communication that women "perform significantly worse". In their paper, this is true for their first round (our Part 2), where "women score 1.3 fewer correct answers" than men (male average: 10.0). Moreover, in their second round (our Part 4), males score on average 10.0, but women score 0.7 fewer correct answers.

Transgender participants show performance differences in NEUTRAL in Part 2 and 3 (MWU, Part 2: NEUTRAL: $p = 0.007$, FEMININE: $p = 0.555$, MASCULINE: $p = 0.181$, Part 3: NEUTRAL: $p = 0.015$, FEMININE: $p = 0.600$, MASCULINE: $p = 0.053$). Concerning sex differences, male participants always have a higher performance than female ones in NEUTRAL and MASCULINE when facing piece-rate incentives (MWU, Part 2: NEUTRAL: $p = 0.003$, FEMININE: $p = 0.164$, MASCULINE: $p = 0.010$). Interestingly, this is true when they compete in Part 3 for all three treatments (MWU, NEUTRAL: $p < 0.001$, FEMININE: $p = 0.014$, MASCULINE: $p = 0.003$).

However, performances do not differ by the individual's gender (MWU, Part 2: NEUTRAL: $p = 0.755$, FEMININE: $p = 0.621$, MASCULINE: $p = 0.553$, Part 3: NEUTRAL: $p = 0.575$, FEMININE: $p = 0.161$, MASCULINE: $p = 0.675$). All differences vanish in Part 4 when we split up the data by those in the tournament (see the respective p-values in Table C.9 and Table C.9). For the priming intervention, we have no evidence of priming influencing the performance, independent of the part or subject group (KW, cismen: $p > 0.478$, ciswomen: $p > 0.562$, transmen: $p > 0.956$, or transwomen: $p > 0.170$).

Please note that we can not exclude that the math task is not influenced by a participant's gender and sex, combinations of it, in addition to interactions with priming. However, we can control how performance heterogeneity affects competitiveness by adding individual performances to our regressions measuring competitiveness. See Table C.15 to Table C.16.

C.3.1 Performance Part 2

Table C.7: Performance in Part 2 across treatments and subject groups.

Panel A: Performance in Part 2 across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	8.458	8.704	9.423	0.478
Ciswomen	7.535	7.929	8.014	0.719
Transmen	7.750	7.778	7.718	0.956
Transwomen	9.409	8.740	8.739	0.285

Panel B: Performance in Part 2 across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	8.458	7.535	7.750	9.409	0.011
FEMININE	8.704	7.929	7.778	8.740	0.529
MASCULINE	9.423	8.014	7.718	8.739	0.062

Panel C: Performance in Part 2 across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	8.000	Transgender	8.379	0.405
Case 2	Cismen	8.458	Ciswomen	7.535	0.080
Case 3	Transmen	7.750	Transwomen	9.409	0.007
Case 4	Female	7.643	Male	8.819	0.003
Case 5	Feminine	8.252	Masculine	8.104	0.755

Panel D: Performance in Part 2 in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	8.458	8.704	0.797	9.423	0.231
Ciswomen	7.535	7.929	0.488	8.014	0.476
Transmen	7.750	7.778	0.832	7.718	0.932
Transwomen	9.409	8.740	0.125	8.739	0.255

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.3.2 Performance Part 3

Table C.8: Performance in Part 3 across treatments and subject groups.

Panel A: Performance in Part 3 across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	9.333	10.070	9.930	0.593
Ciswomen	7.423	7.957	8.271	0.562
Transmen	7.833	7.736	7.930	0.979
Transwomen	9.659	8.500	9.326	0.170

Panel B: Performance in Part 3 across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	9.333	7.423	7.833	9.659	0.002
FEMININE	10.070	7.957	7.736	8.500	0.021
MASCULINE	9.930	8.271	7.930	9.326	0.024

Panel C: Performance in Part 3 across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	8.385	Transgender	8.526	0.612
Case 2	Cismen	9.333	Ciswomen	7.423	0.004
Case 3	Transmen	7.833	Transwomen	9.659	0.015
Case 4	Female	7.629	Male	9.457	$<$0.001
Case 5	Feminine	8.278	Masculine	8.583	0.575

Panel D: Performance in Part 3 in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	9.333	10.070	0.353	9.930	0.406
Ciswomen	7.423	7.957	0.396	8.271	0.325
Transmen	7.833	7.736	0.920	7.930	0.832
Transwomen	9.659	8.500	0.081	9.326	0.686

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.3.3 Performance Part 4

Table C.9: Performance in Part 4 of *competing* subjects across treatments and subject groups.

Panel A: Performance in Part 4 of competing subjects across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	9.500	10.321	9.312	0.765
Ciswomen	8.391	8.947	9.053	0.763
Transmen	9.000	8.211	8.474	0.839
Transwomen	9.429	8.333	9.167	0.670

Panel B: Performance in Part 4 of competing subjects across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	9.500	8.391	9.000	9.429	0.705
FEMININE	10.321	8.947	8.211	8.333	0.389
MASCULINE	9.312	9.053	8.474	9.167	0.923

Panel C: Performance in Part 4 of competing subjects across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	8.933	Transgender	9.176	0.651
Case 2	Cismen	9.500	Ciswomen	8.391	0.278
Case 3	Transmen	9.000	Transwomen	9.429	0.860
Case 4	Female	8.674	Male	9.472	0.358
Case 5	Feminine	8.784	Masculine	9.262	0.475

Panel D: Performance in Part 4 of competing subjects in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	9.500	10.321	0.556	9.312	0.965
Ciswomen	8.391	8.947	0.638	9.053	0.494
Transmen	9.000	8.211	0.563	8.474	0.671
Transwomen	9.429	8.333	0.360	9.167	0.797

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

Table C.10: Performance in Part 4 of *non-competing* subjects across treatments and subject groups.

Panel A: Performance in Part 4 of non-competing subjects across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	8.960	9.093	9.026	0.994
Ciswomen	7.979	9.098	8.725	0.341
Transmen	8.712	8.453	8.288	0.825
Transwomen	8.800	8.906	9.500	0.703

Panel B: Performance in Part 4 of non-competing subjects across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	8.960	7.979	8.712	8.800	0.577
FEMININE	9.093	9.098	8.453	8.906	0.929
MASCULINE	9.026	8.725	8.288	9.500	0.376

Panel C: Performance in Part 4 of non-competing subjects across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	8.480	Transgender	8.744	0.874
Case 2	Cismen	8.960	Ciswomen	7.979	0.159
Case 3	Transmen	8.712	Transwomen	8.800	0.977
Case 4	Female	8.360	Male	8.900	0.292
Case 5	Feminine	8.295	Masculine	8.833	0.309

Panel D: Performance in Part 4 of non-competing subjects in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	8.960	9.093	0.914	9.026	0.960
Ciswomen	7.979	9.098	0.135	8.725	0.529
Transmen	8.712	8.453	0.921	8.288	0.606
Transwomen	8.800	8.906	0.843	9.500	0.445

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.4 Beliefs (Part 3)

C.4.1 Non-parametric tests

Table C.11: Beliefs in Part 3 across treatments and subject groups.

Panel A: Beliefs in Part 3 across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	2.139	1.944	2.070	0.391
Ciswomen	2.606	2.500	2.571	0.793
Transmen	2.542	2.653	2.704	0.633
Transwomen	2.205	2.440	2.304	0.396

Panel B: Beliefs in Part 3 across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	2.139	2.606	2.542	2.205	0.006
FEMININE	1.944	2.500	2.653	2.440	<0.001
MASCULINE	2.070	2.571	2.704	2.304	<0.001

Panel C: Beliefs in Part 3 across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	2.371	Transgender	2.414	0.746
Case 2	Cismen	2.139	Ciswomen	2.606	0.003
Case 3	Transmen	2.542	Transwomen	2.205	0.061
Case 4	Female	2.573	Male	2.164	0.001
Case 5	Feminine	2.452	Masculine	2.340	0.362

Panel D: Beliefs in Part 3 in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	2.139	1.944	0.177	2.070	0.567
Ciswomen	2.606	2.500	0.496	2.571	0.772
Transmen	2.542	2.653	0.529	2.704	0.354
Transwomen	2.205	2.440	0.178	2.304	0.537

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.4.2 Regressions

Table C.12: OLS regression for NEUTRAL. Dependent variable: Beliefs in Part 3.

	(1)	(2)	(3)
Ciswomen	0.467 **	0.586 **	0.494 **
	(0.155)	(0.176)	(0.156)
Transmen	0.403 *	0.511 **	-0.155
	(0.159)	(0.180)	(0.322)
Transwomen	0.066	0.194	-0.551
	(0.172)	(0.173)	(0.324)
Age		0.000	
		(0.009)	
Height		0.005	
		(0.006)	
Student status		-0.169	
		(0.136)	
Income: <20,000 GBP		-0.077	
		(0.135)	
Religion: Religious		0.243	
		(0.135)	
Religion: Not say		0.460	
		(0.313)	
Residence: US		-0.080	
		(0.152)	
Residence: UK		0.097	
		(0.170)	
Residence: Other		-0.408 *	
		(0.202)	
TCS			-0.253 **
			(0.093)
STT			0.020
			(0.022)
Const.	2.139 ***	1.226	3.242 ***
	(0.105)	(1.195)	(0.436)
N	259	259	259
Adj. R2	0.035	0.057	0.057
H_0: Sex	0.001	0.001	0.000
H_0: Gender	0.587	0.259	0.682

Note: The beliefs in Part 3 are the participants belief about how their performance ranks within the group (1 = best to 4 = worst). Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** p <0.001; ** p <0.01; * p <0.05. H_0: Sex reports the p-values of a joint coefficient test comparing Male (Cismen and Transwomen) with Female (Ciswomen and Transmen). H_0: Gender reports the p-values of a joint coefficient test comparing Masculine (Cismen and Transmen) with Feminine (Ciswomen and Transwomen).

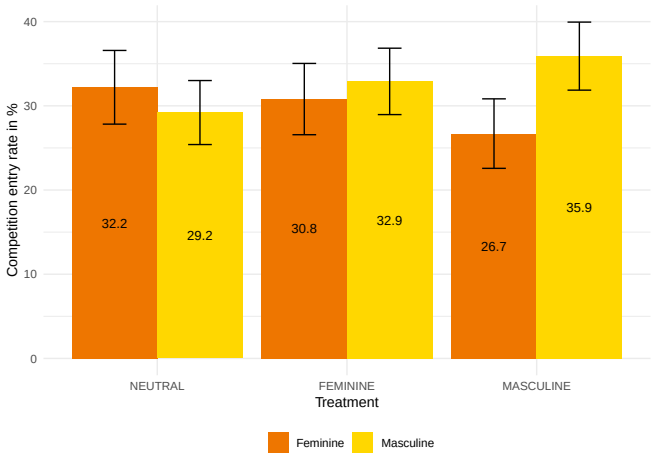
Table C.13: OLS regression for all treatments. Dependent variable: Beliefs in Part 3.

	(1)	(2)	(3)
Ciswomen	0.467 **	0.444 **	0.484 **
	(0.155)	(0.162)	(0.155)
Transmen	0.403 *	0.405 *	0.093
	(0.159)	(0.164)	(0.229)
Transwomen	0.066	0.080	-0.291
	(0.172)	(0.174)	(0.234)
Treatment: FEMININE	-0.195	-0.187	-0.184
	(0.149)	(0.146)	(0.148)
Treatment: MASCULINE	-0.068	-0.045	-0.067
	(0.151)	(0.148)	(0.150)
FEMININE x Ciswomen	0.090	0.094	0.065
	(0.218)	(0.216)	(0.217)
MASCULINE x Ciswomen	0.034	0.034	0.013
	(0.220)	(0.215)	(0.219)
FEMININE x Transmen	0.306	0.309	0.273
	(0.218)	(0.216)	(0.215)
MASCULINE x Transmen	0.231	0.210	0.222
	(0.222)	(0.220)	(0.219)
FEMININE x Transwomen	0.431	0.421	0.434
	(0.239)	(0.238)	(0.236)
MASCULINE x Transwomen	0.168	0.145	0.166
	(0.242)	(0.242)	(0.241)
Const.	2.139 ***	2.547 ***	2.860 ***
	(0.105)	(0.693)	(0.250)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.055	0.074	0.067
H_0: FEMININE on Cismen	0.192	0.201	0.213
H_0: MASCULINE on Cismen	0.651	0.762	0.654
H_0: FEMININE on Ciswomen	0.504	0.553	0.454
H_0: MASCULINE on Ciswomen	0.830	0.943	0.736
H_0: FEMININE on Transmen	0.483	0.443	0.568
H_0: MASCULINE on Transmen	0.317	0.309	0.331
H_0: FEMININE on Transwomen	0.207	0.211	0.178
H_0: MASCULINE on Transwomen	0.597	0.601	0.599

Note: The beliefs in Part 3 are the participants belief about how their performance ranks within the group (1 = best to 4 = worst). Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** p <0.001; ** p <0.01; * p <0.05. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

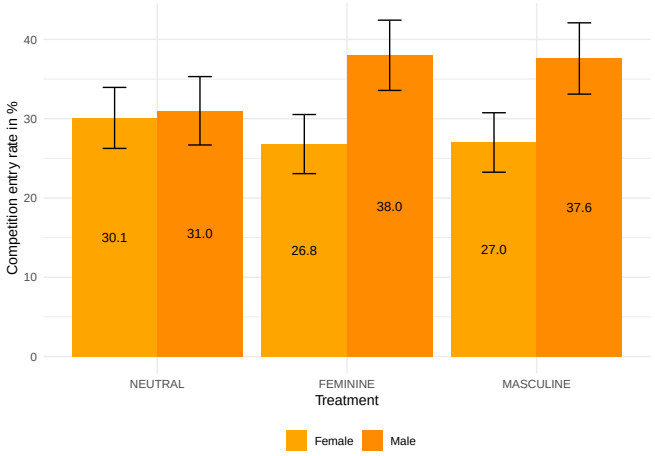
C.5 Competitiveness (Part 4)

C.5.1 Bar graphs



Note: The bars show the percentage of participants (between 0 and 100) who chose to compete rather than to perform under piece-rate incentives. The error bars represent the standard error of the mean.

Figure C.2: Tournament entry rates in Part 4 by treatment and gender ($n = 780$).



Note: The bars show the percentage of participants (between 0 and 100) who chose to compete rather than to perform under piece-rate incentives. The error bars represent the standard error of the mean.

Figure C.3: Tournament entry rates in Part 4 by treatment and sex ($n = 780$).

C.5.2 Non-parametric tests

Table C.14: Tournament entry rates across treatments and subject groups.

Panel A: Competitiveness across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	30.6%	39.4%	45.1%	0.198
Ciswomen	32.4%	27.1%	27.1%	0.729
Transmen	27.8%	26.4%	26.8%	0.981
Transwomen	31.8%	36.0%	26.1%	0.578

Panel B: Competitiveness across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	30.6%	32.4%	27.8%	31.8%	0.939
FEMININE	39.4%	27.1%	26.4%	36.0%	0.264
MASCULINE	45.1%	27.1%	26.8%	26.1%	0.046

Panel C: Competitiveness across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	31.5%	Transgender	29.3%	0.708
Case 2	Cismen	30.6%	Ciswomen	32.4%	0.813
Case 3	Transmen	27.8%	Transwomen	31.8%	0.643
Case 4	Female	30.1%	Male	31.0%	0.867
Case 5	Feminine	32.2%	Masculine	29.2%	0.601

Panel D: Competitiveness in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	30.6%	39.4%	0.265	45.1%	0.073
Ciswomen	32.4%	27.1%	0.495	27.1%	0.495
Transmen	27.8%	26.4%	0.851	26.8%	0.891
Transwomen	31.8%	36.0%	0.669	26.1%	0.549

Note: The column p-value reports the results of χ^2 tests performed column-wise.

C.5.3 Regressions

Table C.15: Probit regression for NEUTRAL. Dependent variable: Competition.

	(1)	(2)	(3)	(4)	(5)
Ciswomen	0.052	0.173	0.281	0.360	0.268
	(0.220)	(0.231)	(0.254)	(0.297)	(0.254)
Transmen	-0.081	0.013	0.106	0.114	0.516
	(0.223)	(0.231)	(0.239)	(0.292)	(0.462)
Transwomen	0.036	0.108	0.084	-0.076	0.478
	(0.252)	(0.261)	(0.265)	(0.306)	(0.472)
Perf. tournament		0.003	-0.064 *	-0.054	-0.064 *
		(0.025)	(0.029)	(0.030)	(0.029)
Delta perf.		0.122 ***	0.096 *	0.103 **	0.095 *
		(0.035)	(0.039)	(0.038)	(0.040)
Belief tournament			-0.522 ***	-0.516 ***	-0.518 ***
			(0.123)	(0.127)	(0.124)
Risk			0.124	0.137	0.130
			(0.074)	(0.077)	(0.074)
Age				0.026 *	
				(0.012)	
Height				0.007	
				(0.010)	
Student status				-0.073	
				(0.209)	
Income: < 20,000 GBP				0.442	
				(0.232)	
Religion: Religious				0.081	
				(0.207)	
Religion: Not say				0.268	
				(0.485)	
Residence: US				0.159	
				(0.254)	
Residence: UK				0.421	
				(0.270)	
Residence: Other				0.564	
				(0.391)	
TCS					0.094
					(0.139)
STT					-0.033
					(0.033)
Const.	-0.508 **	-0.670 *	0.823	-1.618	0.419
	(0.156)	(0.272)	(0.468)	(2.000)	(0.789)
N	259	259	259	259	259
Pseudo R2 (McFadden)	0.001	0.059	0.129	0.168	0.132
H_0: Sex	0.846	0.823	0.415	0.214	0.420
H_0: Gender	0.614	0.437	0.470	0.656	0.527

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** p <0.001; ** p <0.01; * p <0.05. H_0: Sex reports the p-values of a joint coefficient test comparing Male (Cismen and Transwomen) with Female (Ciswomen and Transmen). H_0: Gender reports the p-values of a joint coefficient test comparing Masculine (Cismen and Transmen) with Feminine (Ciswomen and Transwomen).

Table C.16: Probit regression for all treatments. Dependent variable: Competition.

	(1)	(2)	(3)	(4)	(5)
Ciswomen	0.052	0.167	0.307	0.461	0.280
	(0.220)	(0.229)	(0.255)	(0.270)	(0.255)
Transmen	-0.081	0.008	0.127	0.189	0.634
	(0.223)	(0.230)	(0.241)	(0.270)	(0.343)
Transwomen	0.036	0.122	0.078	0.079	0.635
	(0.252)	(0.262)	(0.272)	(0.287)	(0.359)
Treatment: FEMININE	0.241	0.187	0.129	0.167	0.104
	(0.218)	(0.223)	(0.227)	(0.241)	(0.231)
Treatment: MASCULINE	0.385	0.473 *	0.428	0.544 *	0.423
	(0.217)	(0.225)	(0.232)	(0.246)	(0.234)
FEMININE x Ciswomen	-0.392	-0.339	-0.318	-0.383	-0.285
	(0.313)	(0.329)	(0.360)	(0.365)	(0.360)
MASCULINE x Ciswomen	-0.536	-0.678 *	-0.645	-0.755 *	-0.614
	(0.312)	(0.323)	(0.342)	(0.348)	(0.343)
FEMININE x Transmen	-0.282	-0.201	-0.107	-0.112	-0.052
	(0.313)	(0.325)	(0.329)	(0.340)	(0.333)
MASCULINE x Transmen	-0.415	-0.543	-0.405	-0.475	-0.369
	(0.313)	(0.319)	(0.327)	(0.341)	(0.333)
FEMININE x Transwomen	-0.126	-0.018	0.187	0.131	0.150
	(0.347)	(0.352)	(0.357)	(0.366)	(0.360)
MASCULINE x Transwomen	-0.552	-0.690	-0.573	-0.651	-0.583
	(0.356)	(0.366)	(0.377)	(0.393)	(0.379)
Perf. tournament		-0.008	-0.067 ***	-0.066 ***	-0.066 ***
		(0.014)	(0.016)	(0.016)	(0.016)
Delta perf.		0.136 ***	0.108 ***	0.114 ***	0.110 ***
		(0.020)	(0.021)	(0.022)	(0.021)
Belief tournament			-0.561 ***	-0.557 ***	-0.543 ***
			(0.071)	(0.072)	(0.071)
Risk			0.150 ***	0.153 ***	0.153 ***
			(0.045)	(0.046)	(0.045)
Const.	-0.508 **	-0.587 **	0.865 **	-1.757	-0.072
	(0.156)	(0.200)	(0.310)	(1.255)	(0.480)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	780	780	780	780	780
Pseudo R2 (McFadden)	0.013	0.082	0.161	0.178	0.167
H_0: FEMININE on Cismen	0.266	0.402	0.579	0.478	0.656
H_0: MASCULINE on Cismen	0.074	0.034	0.063	0.021	0.068
H_0: FEMININE on Ciswomen	0.495	0.508	0.432	0.374	0.452
H_0: MASCULINE on Ciswomen	0.495	0.374	0.368	0.387	0.426
H_0: FEMININE on Transmen	0.851	0.953	0.928	0.821	0.832
H_0: MASCULINE on Transmen	0.891	0.761	0.922	0.779	0.828
H_0: FEMININE on Transwomen	0.669	0.541	0.274	0.310	0.386
H_0: MASCULINE on Transwomen	0.549	0.449	0.627	0.724	0.591

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.5.4 Cohen's *d*

Table C.17: Cohen's d analysis: Competitiveness

	—Cohen's d—	Cohen's d	CI.95-Lower	CI.95-Upper	t-statistic	p-value
Within NEUTRAL						
Female vs. Male	0.021	-0.021	-0.266	0.224	-0.169	0.866
Feminine vs. Masculine	0.065	0.065	-0.180	0.310	0.526	0.599
Cismen vs. Ciswomen	0.040	-0.040	-0.367	0.288	-0.237	0.813
Cismen vs. Transmen	0.061	0.061	-0.266	0.388	0.367	0.714
Cismen vs. Transwomen	0.027	-0.027	-0.402	0.348	-0.147	0.883
Ciswomen vs. Transmen	0.101	0.101	-0.228	0.429	0.603	0.548
Ciswomen vs. Transwomen	0.012	0.012	-0.364	0.388	0.066	0.947
Transmen vs. Transwomen	0.089	-0.089	-0.464	0.287	-0.478	0.633
Between treatments						
NEUTRAL vs. FEMININE	0.031	-0.031	-0.203	0.141	-0.354	0.723
Cismen	0.187	-0.187	-0.516	0.143	-1.118	0.265
Ciswomen	0.115	0.115	-0.216	0.445	0.683	0.496
Transmen	0.031	0.031	-0.296	0.358	0.188	0.852
Transwomen	0.088	-0.088	-0.493	0.318	-0.428	0.669
NEUTRAL vs. MASCULINE	0.028	-0.028	-0.200	0.145	-0.315	0.753
Cismen	0.303	-0.303	-0.633	0.030	-1.811	0.072
Ciswomen	0.115	0.115	-0.216	0.445	0.683	0.496
Transmen	0.023	0.023	-0.305	0.351	0.137	0.892
Transwomen	0.127	0.127	-0.288	0.540	0.601	0.549

C.6 Payoffs (Part 4)

C.6.1 Regressions

Table C.18: OLS regression for NEUTRAL. Dependent variable: Payoff in Part 4.

	(1)	(2)	(3)
Ciswomen	-0.156	1.315	-0.267
	(1.183)	(1.326)	(1.177)
Transmen	0.951	2.316	3.377
	(1.256)	(1.358)	(2.170)
Transwomen	-1.071	-0.560	1.571
	(1.244)	(1.278)	(2.215)
Age		-0.040	
		(0.047)	
Height		0.113 **	
		(0.041)	
Student status		3.479 ***	
		(1.017)	
Income: < 20,000 GBP		0.402	
		(0.922)	
Religion: Religious		-2.336 *	
		(0.953)	
Religion: Not say		-3.673 **	
		(1.247)	
Residence: US		1.421	
		(1.212)	
Residence: UK		1.468	
		(1.176)	
Residence: Other		-1.269	
		(1.576)	
TCS			1.036
			(0.634)
STT			-0.099
			(0.153)
Const.	5.389 ***	-15.595 *	0.895
	(0.851)	(7.668)	(2.853)
N	259	259	259
Adj. R2	-0.003	0.073	-0.002
H_0: Sex	0.289	0.056	0.366
H_0: Gender	0.215	0.381	0.234

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. H_0: Sex reports the p-values of a joint coefficient test comparing Male (Cismen and Transwomen) with Female (Ciswomen and Transmen). H_0: Gender reports the p-values of a joint coefficient test comparing Masculine (Cismen and Transmen) with Feminine (Ciswomen and Transwomen).

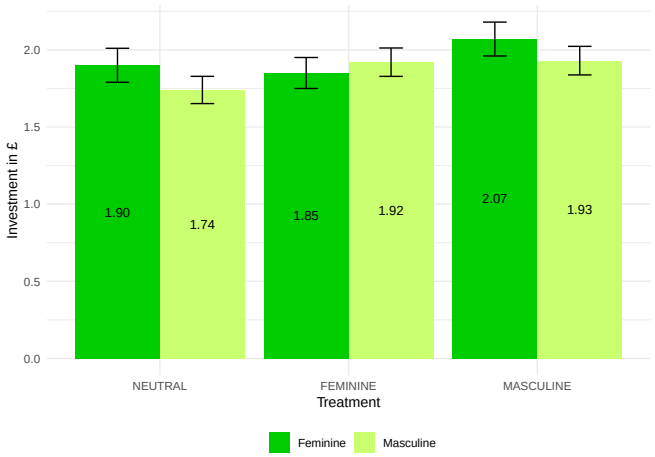
Table C.19: OLS regression for all treatments. Dependent variable: Payoff in Part 4.

	(1)	(2)	(3)
Ciswomen	-0.156	0.922	-0.225
	(1.183)	(1.245)	(1.178)
Transmen	0.951	1.785	1.993
	(1.256)	(1.279)	(1.614)
Transwomen	-1.071	-0.976	0.193
	(1.244)	(1.250)	(1.643)
Treatment: FEMININE	2.660	2.622	2.632
	(1.516)	(1.483)	(1.525)
Treatment: MASCULINE	0.442	0.354	0.435
	(1.331)	(1.343)	(1.328)
FEMININE x Ciswomen	-2.293	-2.286	-2.207
	(1.925)	(1.892)	(1.931)
MASCULINE x Ciswomen	-1.096	-1.243	-1.003
	(1.700)	(1.717)	(1.692)
FEMININE x Transmen	-4.501 *	-4.672 *	-4.383 *
	(1.905)	(1.868)	(1.909)
MASCULINE x Transmen	-0.818	-0.744	-0.781
	(1.859)	(1.850)	(1.849)
FEMININE x Transwomen	-1.689	-1.536	-1.664
	(2.111)	(2.116)	(2.108)
MASCULINE x Transwomen	-0.075	-0.088	-0.054
	(1.813)	(1.770)	(1.827)
Const.	5.389 ***	-8.821	2.454
	(0.851)	(5.310)	(1.658)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.003	0.034	0.005
H_0: FEMININE on Cismen	0.080	0.077	0.085
H_0: MASCULINE on Cismen	0.740	0.792	0.743
H_0: FEMININE on Ciswomen	0.757	0.773	0.719
H_0: MASCULINE on Ciswomen	0.536	0.411	0.589
H_0: FEMININE on Transmen	0.111	0.070	0.129
H_0: MASCULINE on Transmen	0.772	0.760	0.788
H_0: FEMININE on Transwomen	0.509	0.472	0.514
H_0: MASCULINE on Transwomen	0.766	0.822	0.762

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

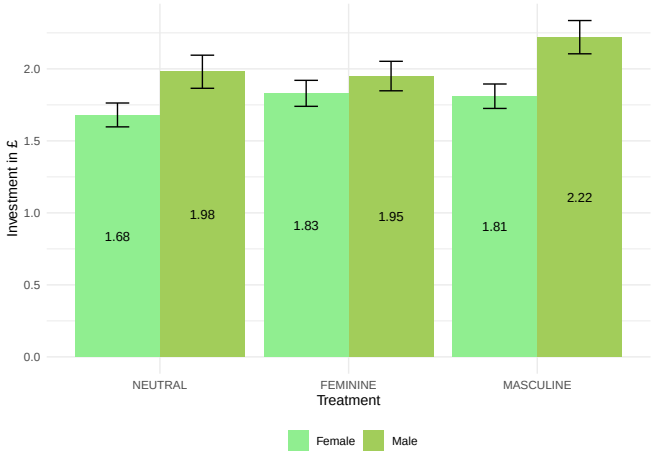
C.7 Risk (Part 5)

C.7.1 Bar graphs



Note: The bars show the average investment rate, and the error bars represent the standard error of the mean.

Figure C.4: Investment into the risky lottery in Part 5 by treatment and gender ($n = 780$).



Note: The bars show the average investment rate, and the error bars represent the standard error of the mean.

Figure C.5: Investment into the risky lottery in Part 5 by treatment and sex ($n = 780$).

C.7.2 Non-parametric tests

Table C.20: Investment into the risky lottery across treatments and subject groups.

Panel A: Risk across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	1.814	2.021	2.208	0.119
Ciswomen	1.690	1.852	1.972	0.446
Transmen	1.673	1.816	1.655	0.660
Transwomen	2.244	1.840	2.227	0.357

Panel B: Risk across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	1.814	1.690	1.673	2.244	0.194
FEMININE	2.021	1.852	1.816	1.840	0.715
MASCULINE	2.208	1.972	1.655	2.227	0.030

Panel C: Risk across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	1.753	Transgender	1.890	0.461
Case 2	Cismen	1.814	Ciswomen	1.690	0.704
Case 3	Transmen	1.673	Transwomen	2.244	0.048
Case 4	Female	1.681	Male	1.977	0.130
Case 5	Feminine	1.902	Masculine	1.743	0.355

Panel D: Risk in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	1.814	2.021	0.262	2.208	0.038
Ciswomen	1.690	1.852	0.550	1.972	0.208
Transmen	1.673	1.816	0.479	1.655	0.881
Transwomen	2.244	1.840	0.206	2.227	0.927

Note: The column p-value reports the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.7.3 Regressions

Table C.21: OLS regression for NEUTRAL. Dependent variable: Risk.

	(1)	(2)	(3)
Ciswomen	-0.124	-0.164	-0.109
	(0.179)	(0.218)	(0.177)
Transmen	-0.141	-0.221	-0.295
	(0.177)	(0.212)	(0.396)
Transwomen	0.430	0.316	0.216
	(0.245)	(0.258)	(0.412)
Age		-0.008	
		(0.010)	
Height		-0.001	
		(0.008)	
Student status		0.064	
		(0.161)	
Income: < 20,000 GBP		-0.125	
		(0.162)	
Religion: Religious		-0.045	
		(0.160)	
Religion: Not say		-0.502	
		(0.414)	
Residence: US		0.052	
		(0.202)	
Residence: UK		0.145	
		(0.194)	
Residence: Other		0.618 *	
		(0.309)	
TCS			-0.143
			(0.105)
STT			-0.009
			(0.029)
Const.	1.814 ***	2.273	2.463 ***
	(0.133)	(1.565)	(0.499)
N	259	259	259
Adj. R2	0.022	0.018	0.024
H_0: Sex	0.020	0.042	0.037
H_0: Gender	0.132	0.228	0.183

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. H_0: Sex reports the p-values of a joint coefficient test comparing Male (Cismen and Transwomen) with Female (Ciswomen and Transmen). H_0: Gender reports the p-values of a joint coefficient test comparing Masculine (Cismen and Transmen) with Feminine (Ciswomen and Transwomen).

Table C.22: OLS regression for all treatments. Dependent variable: Risk.

	(1)	(2)	(3)
Ciswomen	-0.124	-0.123	-0.124
	(0.179)	(0.195)	(0.179)
Transmen	-0.141	-0.154	-0.273
	(0.177)	(0.189)	(0.286)
Transwomen	0.430	0.387	0.325
	(0.245)	(0.248)	(0.324)
Treatment: FEMININE	0.207	0.210	0.218
	(0.192)	(0.193)	(0.192)
Treatment: MASCULINE	0.394 *	0.386 *	0.393 *
	(0.191)	(0.193)	(0.191)
FEMININE x Ciswomen	-0.045	-0.048	-0.052
	(0.263)	(0.263)	(0.263)
MASCULINE x Ciswomen	-0.112	-0.118	-0.106
	(0.256)	(0.259)	(0.255)
FEMININE x Transmen	-0.064	-0.074	-0.073
	(0.255)	(0.257)	(0.255)
MASCULINE x Transmen	-0.412	-0.409	-0.410
	(0.252)	(0.253)	(0.252)
FEMININE x Transwomen	-0.611	-0.593	-0.585
	(0.319)	(0.320)	(0.322)
MASCULINE x Transwomen	-0.412	-0.380	-0.403
	(0.348)	(0.350)	(0.349)
Const.	1.814 ***	1.605	1.762 ***
	(0.133)	(0.913)	(0.308)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.018	0.013	0.017
H_0: FEMININE on Cismen	0.282	0.278	0.257
H_0: MASCULINE on Cismen	0.039	0.046	0.040
H_0: FEMININE on Ciswomen	0.368	0.370	0.356
H_0: MASCULINE on Ciswomen	0.097	0.125	0.092
H_0: FEMININE on Transmen	0.395	0.426	0.388
H_0: MASCULINE on Transmen	0.915	0.889	0.918
H_0: FEMININE on Transwomen	0.113	0.132	0.156
H_0: MASCULINE on Transwomen	0.951	0.984	0.975

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

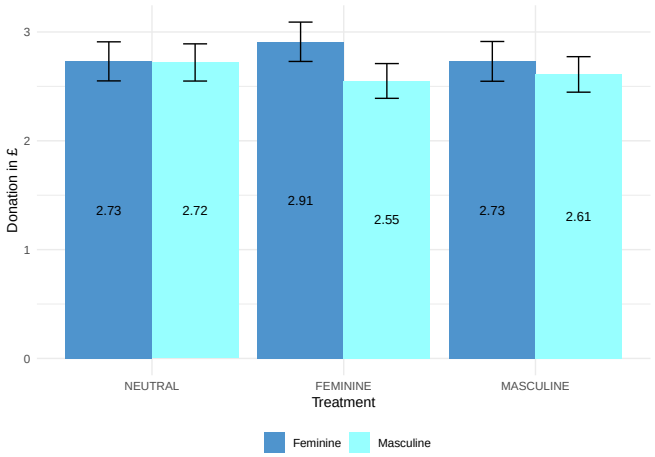
C.7.4 Cohen's *d*

Table C.23: Cohen's d analysis: Risk

	—Cohen's d—	Cohen's d	CI.95-Lower	CI.95-Upper	t-statistic	p-value
Within NEUTRAL						
Female vs. Male	0.268	-0.268	-0.515	-0.020	-2.156	0.032
Feminine vs. Masculine	0.143	0.143	-0.103	0.388	1.150	0.251
Cismen vs. Ciswomen	0.117	0.117	-0.212	0.445	0.698	0.486
Cismen vs. Transmen	0.134	0.134	-0.194	0.461	0.804	0.423
Cismen vs. Transwomen	0.354	-0.354	-0.735	0.030	-1.908	0.059
Ciswomen vs. Transmen	0.017	0.017	-0.310	0.345	0.104	0.917
Ciswomen vs. Transwomen	0.484	-0.484	-0.871	-0.092	-2.596	0.011
Transmen vs. Transwomen	0.504	-0.504	-0.891	-0.111	-2.712	0.008
Between treatments						
NEUTRAL vs. FEMININE	0.065	-0.065	-0.236	0.107	-0.740	0.459
Cismen	0.182	-0.182	-0.510	0.148	-1.086	0.279
Ciswomen	0.153	-0.153	-0.484	0.179	-0.909	0.365
Transmen	0.143	-0.143	-0.470	0.185	-0.857	0.393
Transwomen	0.335	0.335	-0.077	0.744	1.626	0.107
NEUTRAL vs. MASCULINE	0.161	-0.161	-0.334	0.012	-1.833	0.067
Cismen	0.348	-0.348	-0.680	-0.014	-2.080	0.039
Ciswomen	0.282	-0.282	-0.614	0.053	-1.673	0.097
Transmen	0.018	0.018	-0.310	0.346	0.107	0.915
Transwomen	0.013	0.013	-0.400	0.426	0.062	0.951

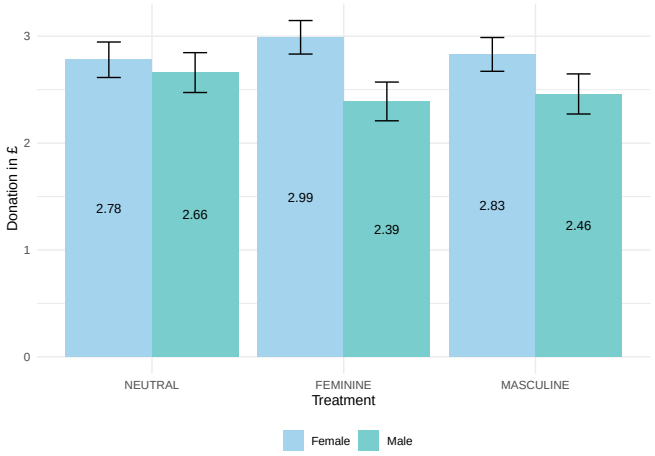
C.8 Altruism (Part 6)

C.8.1 Bar graphs



Note: The average donations are indicated by the bars and the error bars represent the standard error of the mean.

Figure C.6: Donation in Part 6 by treatment and gender ($n = 780$).



Note: The average donations are indicated by the bars and the error bars represent the standard error of the mean.

Figure C.7: Donation in Part 6 by treatment and sex ($n = 780$).

C.8.2 Non-parametric tests

Table C.24: Donations across treatments and subject groups.

Panel A: Donations across treatments

	Treatment			
Subject groups	NEUTRAL	FEMININE	MASCULINE	p-value
Cismen	2.685	2.265	2.423	0.582
Ciswomen	2.803	3.161	2.864	0.478
Transmen	2.762	2.822	2.792	0.999
Transwomen	2.615	2.556	2.525	0.999

Panel B: Donations across subject groups

	Subject groups				
Treatment	Cismen	Ciswomen	Transmen	Transwomen	p-value
NEUTRAL	2.685	2.803	2.762	2.615	0.933
FEMININE	2.265	3.161	2.822	2.556	0.073
MASCULINE	2.423	2.864	2.792	2.525	0.540

Panel C: Donations across groups within NEUTRAL

	Group 1		Group 2		
	Subjects		Subjects		p-value
Case 1	Cisgender	2.743	Transgender	2.706	0.871
Case 2	Cismen	2.685	Ciswomen	2.803	0.759
Case 3	Transmen	2.762	Transwomen	2.615	0.583
Case 4	Female	2.782	Male	2.658	0.564
Case 5	Feminine	2.731	Masculine	2.723	0.914

Panel D: Donations in NEUTRAL *compared to the other treatments*

	NEUTRAL	FEMININE		MASCULINE	
Subject groups			p-value		p-value
Cismen	2.685	2.265	0.322	2.423	0.454
Ciswomen	2.803	3.161	0.260	2.864	0.863
Transmen	2.762	2.822	0.982	2.792	0.977
Transwomen	2.615	2.556	0.948	2.525	0.987

Note: The columns p-value report the results of the tests performed column-wise. For continuous variables, we conducted Mann-Whitney U tests for two groups and Kruskal-Wallis tests for more than two groups.

C.8.3 Regressions

Table C.25: OLS regression for NEUTRAL. Dependent variable: Donations.

	(1)	(2)	(3)
Ciswomen	0.118	0.188	0.128
	(0.335)	(0.403)	(0.337)
Transmen	0.077	0.239	-0.162
	(0.343)	(0.387)	(0.646)
Transwomen	-0.070	0.103	-0.324
	(0.378)	(0.409)	(0.647)
Age		0.040 *	
		(0.016)	
Height		0.002	
		(0.013)	
Student status		0.713 **	
		(0.274)	
Income: < 20,000 GBP		0.025	
		(0.279)	
Religion: Religious		0.800 **	
		(0.274)	
Religion: Not say		0.701	
		(0.736)	
Residence: US		-0.789 *	
		(0.324)	
Residence: UK		-0.470	
		(0.335)	
Residence: Other		-0.391	
		(0.538)	
TCS			-0.091
			(0.199)
STT			0.012
			(0.046)
Const.	2.685 ***	1.123	3.077 ***
	(0.244)	(2.413)	(0.909)
N	259	259	259
Adj. R2	-0.011	0.044	-0.018
H_0: Sex	0.600	0.600	0.580
H_0: Gender	0.955	0.920	0.948

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. H_0: Sex reports the p-values of a joint coefficient test comparing Male (Cismen and Transwomen) with Female (Ciswomen and Transmen). H_0: Gender reports the p-values of a joint coefficient test comparing Masculine (Cismen and Transmen) with Feminine (Ciswomen and Transwomen).

Table C.26: OLS regression for all treatments. Dependent variable: Donations.

	(1)	(2)	(3)
Ciswomen	0.118	0.175	0.128
	(0.335)	(0.358)	(0.336)
Transmen	0.077	0.208	-0.036
	(0.343)	(0.351)	(0.465)
Transwomen	-0.070	-0.004	-0.222
	(0.378)	(0.386)	(0.477)
Treatment: FEMININE	-0.420	-0.377	-0.419
	(0.338)	(0.343)	(0.338)
Treatment: MASCULINE	-0.262	-0.229	-0.261
	(0.340)	(0.346)	(0.340)
FEMININE x Ciswomen	0.778	0.745	0.767
	(0.468)	(0.471)	(0.469)
MASCULINE x Ciswomen	0.323	0.355	0.308
	(0.468)	(0.470)	(0.468)
FEMININE x Transmen	0.480	0.410	0.465
	(0.468)	(0.458)	(0.468)
MASCULINE x Transmen	0.292	0.270	0.287
	(0.474)	(0.466)	(0.474)
FEMININE x Transwomen	0.361	0.444	0.350
	(0.530)	(0.533)	(0.529)
MASCULINE x Transwomen	0.172	0.191	0.166
	(0.543)	(0.547)	(0.545)
Const.	2.685 ***	1.806	3.117 ***
	(0.244)	(1.435)	(0.516)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	-0.000	0.023	-0.002
H_0: FEMININE on Cismen	0.214	0.272	0.216
H_0: MASCULINE on Cismen	0.441	0.509	0.444
H_0: FEMININE on Ciswomen	0.270	0.256	0.282
H_0: MASCULINE on Ciswomen	0.848	0.693	0.882
H_0: FEMININE on Transmen	0.854	0.914	0.886
H_0: MASCULINE on Transmen	0.926	0.895	0.937
H_0: FEMININE on Transwomen	0.885	0.870	0.867
H_0: MASCULINE on Transwomen	0.832	0.930	0.825

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.8.4 Cohen's *d*

Table C.27: Cohen's d analysis: Altruism

	—Cohen's d—	Cohen's d	CI.95-Lower	CI.95-Upper	t-statistic	p-value
Within NEUTRAL						
Female vs. Male	0.062	0.062	-0.183	0.307	0.502	0.616
Feminine vs. Masculine	0.004	0.004	-0.241	0.249	0.031	0.975
Cismen vs. Ciswomen	0.059	-0.059	-0.387	0.269	-0.355	0.723
Cismen vs. Transmen	0.038	-0.038	-0.364	0.289	-0.226	0.821
Cismen vs. Transwomen	0.035	0.035	-0.340	0.410	0.188	0.851
Ciswomen vs. Transmen	0.021	0.021	-0.307	0.348	0.124	0.902
Ciswomen vs. Transwomen	0.098	0.098	-0.279	0.474	0.526	0.600
Transmen vs. Transwomen	0.074	0.074	-0.302	0.449	0.399	0.691
Between treatments						
NEUTRAL vs. FEMININE	0.008	0.008	-0.164	0.180	0.091	0.928
Cismen	0.210	0.210	-0.121	0.538	1.253	0.212
Ciswomen	0.188	-0.188	-0.518	0.145	-1.113	0.267
Transmen	0.031	-0.031	-0.358	0.296	-0.186	0.853
Transwomen	0.030	0.030	-0.375	0.435	0.145	0.885
NEUTRAL vs. MASCULINE	0.033	0.033	-0.140	0.205	0.371	0.711
Cismen	0.130	0.130	-0.199	0.458	0.776	0.439
Ciswomen	0.032	-0.032	-0.363	0.298	-0.193	0.847
Transmen	0.016	-0.016	-0.343	0.312	-0.093	0.926
Transwomen	0.045	0.045	-0.369	0.458	0.214	0.831

C.9 Continuous gender measure (BEM)

C.9.1 Competitiveness

Table C.28: Probit regression for NEUTRAL. Dependent variable: Competition. Gender is measured on a continuous scale.

	(1)	(2)	(3)	(4)	(5)
BEM score: Feminine	0.010	0.014	0.016	0.017	0.016
	(0.011)	(0.011)	(0.012)	(0.013)	(0.012)
BEM score: MASCULINE	0.022	0.019	0.015	0.016	0.015
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)
Perf. tournament		0.004	-0.064 *	-0.057	-0.063 *
		(0.024)	(0.028)	(0.029)	(0.029)
Delta perf.		0.121 ***	0.097 *	0.108 **	0.097 *
		(0.034)	(0.038)	(0.038)	(0.039)
Belief tournament			-0.500 ***	-0.484 ***	-0.501 ***
			(0.123)	(0.126)	(0.122)
Risk			0.129	0.131	0.129
			(0.077)	(0.079)	(0.077)
Age				0.023	
				(0.012)	
Height				-0.000	
				(0.009)	
Student status				-0.114	
				(0.204)	
Income: < 20,000 GBP				0.464 *	
				(0.231)	
Religion: Religious				-0.001	
				(0.212)	
Religion: Not say				0.227	
				(0.471)	
Residence: US				0.094	
				(0.255)	
Residence: UK				0.391	
				(0.260)	
Residence: Other				0.495	
				(0.380)	
TCS					-0.012
					(0.096)
STT					-0.011
					(0.022)
Const.	-1.677 **	-1.847 **	-0.315	-1.493	-0.224
	(0.585)	(0.651)	(0.803)	(1.759)	(0.889)
N	259	259	259	259	259
Pseudo R2 (McFadden)	0.018	0.074	0.139	0.175	0.140

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.29: Probit regression for all treatments. Dependent variable: Competition. Gender is measured on a continuous scale.

	(1)	(2)	(3)	(4)	(5)
BEM score: Feminine	0.010	0.014	0.017	0.013	0.017
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
BEM score: MASCULINE	0.022	0.018	0.014	0.014	0.012
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
FEMININE x BEM score: MASCULINE	0.002	0.003	-0.001	0.000	-0.001
	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
MASCULINE x BEM score: MASCULINE	0.001	0.004	0.003	0.007	0.004
	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
FEMININE x BEM score: Feminine	-0.017	-0.018	-0.026	-0.022	-0.027
	(0.014)	(0.014)	(0.015)	(0.016)	(0.016)
MASCULINE x BEM score: Feminine	-0.015	-0.021	-0.029	-0.029	-0.031
	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
Treatment: FEMININE	0.652	0.706	1.148	0.972	1.219
	(0.729)	(0.750)	(0.797)	(0.832)	(0.820)
Treatment: MASCULINE	0.671	0.771	1.166	1.061	1.210
	(0.783)	(0.810)	(0.884)	(0.913)	(0.895)
Perf. tournament		-0.003	-0.067 ***	-0.065 ***	-0.068 ***
		(0.013)	(0.016)	(0.016)	(0.017)
Delta perf.		0.132 ***	0.106 ***	0.112 ***	0.107 ***
		(0.019)	(0.021)	(0.022)	(0.021)
Belief tournament			-0.557 ***	-0.552 ***	-0.551 ***
			(0.071)	(0.071)	(0.071)
Risk			0.151 ***	0.152 ***	0.153 ***
			(0.045)	(0.046)	(0.045)
Const.	-1.677 **	-1.794 **	-0.241	-1.998	-0.521
	(0.585)	(0.628)	(0.719)	(1.227)	(0.767)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	780	780	780	780	780
Pseudo R2 (McFadden)	0.016	0.085	0.163	0.179	0.166
H_0: FEMININE on BEM score: Feminine	0.361	0.338	0.135	0.213	0.115
H_0: MASCULINE on BEM score: Feminine	0.381	0.329	0.160	0.206	0.145
H_0: FEMININE on BEM score: MASCULINE	0.347	0.324	0.127	0.203	0.108
H_0: MASCULINE on BEM score: MASCULINE	0.370	0.315	0.149	0.191	0.135

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on BEM score: Feminine" tests the effect of the treatment (FEMININE) on the subject group (BEM score: Feminine).

C.9.2 Risk

Table C.30: OLS regression for NEUTRAL. Dependent variable: Risk. Gender is measured on a continuous scale.

	(1)	(2)	(3)
BEM score: Feminine	-0.009	-0.008	-0.009
	(0.009)	(0.009)	(0.009)
BEM score: Masculine	0.006	0.006	0.010
	(0.009)	(0.010)	(0.009)
Age		-0.008	
		(0.009)	
Height		0.005	
		(0.007)	
Student status		0.045	
		(0.160)	
Income: < 20,000 GBP		-0.091	
		(0.161)	
Religion: Religious		-0.067	
		(0.167)	
Religion: Not say		-0.518	
		(0.419)	
Residence: US		0.091	
		(0.204)	
Residence: UK		0.167	
		(0.190)	
Residence: Other		0.769 *	
		(0.302)	
TCS			-0.173 *
			(0.078)
STT			-0.019
			(0.018)
Const.	2.011 ***	1.262	2.577 ***
	(0.440)	(1.339)	(0.568)
N	259	259	259
Adj. R2	-0.002	0.004	0.012

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.31: OLS regression for all treatments. Dependent variable: Risk. Gender is measured on a continuous scale.

	(1)	(2)	(3)
BEM score: Feminine	-0.009	-0.009	-0.010
	(0.009)	(0.009)	(0.009)
BEM score: Masculine	0.006	0.005	0.005
	(0.009)	(0.009)	(0.009)
FEMININE x BEM score: Masculine	-0.000	-0.000	-0.001
	(0.013)	(0.013)	(0.013)
MASCULINE x BEM score: Masculine	-0.001	-0.000	-0.001
	(0.014)	(0.014)	(0.014)
FEMININE x BEM score: Feminine	0.001	0.001	0.001
	(0.012)	(0.012)	(0.012)
MASCULINE x BEM score: Feminine	0.017	0.016	0.017
	(0.012)	(0.012)	(0.013)
Treatment: FEMININE	0.045	0.044	0.061
	(0.596)	(0.600)	(0.602)
Treatment: MASCULINE	-0.489	-0.503	-0.485
	(0.635)	(0.640)	(0.637)
Const.	2.011 ***	0.739	1.953 ***
	(0.440)	(0.861)	(0.482)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	-0.001	-0.001	-0.003
H_0: FEMININE on BEM score: Feminine	0.938	0.939	0.917
H_0: MASCULINE on BEM score: Feminine	0.452	0.441	0.457
H_0: FEMININE on BEM score: Masculine	0.940	0.941	0.919
H_0: MASCULINE on BEM score: Masculine	0.435	0.426	0.440

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on BEM score: Feminine" tests the effect of the treatment (FEMININE) on the subject group (BEM score: Feminine).

C.9.3 Altruism

Table C.32: OLS regression for NEUTRAL. Dependent variable: Donations. Gender is measured on a continuous scale.

	(1)	(2)	(3)
BEM score: Feminine	0.023	0.018	0.023
	(0.015)	(0.015)	(0.016)
BEM score: Masculine	0.002	-0.014	0.003
	(0.016)	(0.016)	(0.016)
Age		0.040 *	
		(0.017)	
Height		-0.002	
		(0.010)	
Student status		0.710 *	
		(0.278)	
Income: < 20,000 GBP		0.051	
		(0.275)	
Religion: Religious		0.797 **	
		(0.279)	
Religion: Not say		0.608	
		(0.722)	
Residence: US		-0.831 **	
		(0.309)	
Residence: UK		-0.480	
		(0.330)	
Residence: Other		-0.401	
		(0.499)	
TCS			-0.031
			(0.133)
STT			-0.002
			(0.031)
Const.	1.713 *	1.573	1.808
	(0.774)	(1.963)	(0.919)
N	259	259	259
Adj. R2	0.001	0.054	-0.007

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.33: OLS regression for all treatments. Dependent variable: Donations. Gender is measured on a continuous scale.

	(1)	(2)	(3)
BEM score: Feminine	0.023	0.022	0.023
	(0.015)	(0.015)	(0.015)
BEM score: Masculine	0.002	-0.006	0.004
	(0.016)	(0.016)	(0.016)
FEMININE x BEM score: Masculine	-0.003	0.003	-0.002
	(0.023)	(0.023)	(0.023)
MASCULINE x BEM score: Masculine	0.002	0.002	0.001
	(0.022)	(0.022)	(0.022)
FEMININE x BEM score: Feminine	0.013	0.009	0.015
	(0.020)	(0.020)	(0.020)
MASCULINE x BEM score: Feminine	0.019	0.018	0.021
	(0.021)	(0.020)	(0.021)
Treatment: FEMININE	-0.484	-0.446	-0.603
	(1.012)	(1.007)	(1.012)
Treatment: MASCULINE	-0.927	-0.841	-0.976
	(1.067)	(1.052)	(1.076)
Const.	1.713 *	2.827 *	2.100 *
	(0.774)	(1.351)	(0.821)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.013	0.034	0.014
H_0: FEMININE on BEM score: Feminine	0.638	0.660	0.556
H_0: MASCULINE on BEM score: Feminine	0.389	0.428	0.368
H_0: FEMININE on BEM score: Masculine	0.627	0.656	0.545
H_0: MASCULINE on BEM score: Masculine	0.381	0.420	0.359

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on BEM score: Feminine" tests the effect of the treatment (FEMININE) on the subject group (BEM score: Feminine).

C.10 Controlling for the strength of the primining intervention

C.10.1 Competitiveness

Table C.34: Probit regression for all treatments. Dependent variable: Competition.

	(1)	(2)	(3)	(4)	(5)
Ciswomen	-0.080	0.015	0.177	0.305	0.167
	(0.190)	(0.201)	(0.221)	(0.237)	(0.222)
Transmen	-0.094	-0.021	0.126	0.170	0.616 *
	(0.188)	(0.197)	(0.201)	(0.229)	(0.313)
Transwomen	-0.140	-0.051	-0.021	-0.020	0.509
	(0.216)	(0.224)	(0.233)	(0.243)	(0.324)
Rem. feminine words	0.028	0.009	0.009	0.018	0.006
	(0.039)	(0.038)	(0.037)	(0.039)	(0.038)
Rem. masculine words	0.042	0.054	0.047	0.070	0.048
	(0.044)	(0.045)	(0.046)	(0.048)	(0.046)
Rem. feminine words x Ciswomen	-0.058	-0.038	-0.042	-0.055	-0.040
	(0.053)	(0.054)	(0.057)	(0.058)	(0.057)
Rem. masculine words x Ciswomen	-0.063	-0.088	-0.087	-0.108	-0.086
	(0.061)	(0.062)	(0.063)	(0.064)	(0.062)
Rem. feminine words x Transmen	-0.063	-0.043	-0.036	-0.039	-0.024
	(0.053)	(0.052)	(0.052)	(0.053)	(0.052)
Rem. masculine words x Transmen	-0.085	-0.106	-0.082	-0.095	-0.072
	(0.058)	(0.058)	(0.061)	(0.065)	(0.061)
Rem. feminine words x Transwomen	0.021	0.038	0.056	0.041	0.056
	(0.060)	(0.059)	(0.059)	(0.059)	(0.059)
Rem. masculine words x Transwomen	-0.068	-0.092	-0.091	-0.109	-0.091
	(0.068)	(0.069)	(0.074)	(0.075)	(0.072)
Const.	-0.394 **	-0.463 *	0.954 **	-1.462	0.014
	(0.131)	(0.183)	(0.293)	(1.248)	(0.468)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	780	780	780	780	780
Pseudo R2 (McFadden)	0.013	0.081	0.160	0.176	0.167
H_0: Rem. feminine words on Cismen	0.472	0.819	0.830	0.663	0.880
H_0: Rem. masculine words on Cismen	0.337	0.242	0.317	0.145	0.312
H_0: Rem. feminine words on Ciswomen	0.415	0.446	0.402	0.364	0.395
H_0: Rem. masculine words on Ciswomen	0.600	0.402	0.352	0.388	0.373
H_0: Rem. feminine words on Transmen	0.322	0.349	0.476	0.588	0.638
H_0: Rem. masculine words on Transmen	0.273	0.198	0.415	0.571	0.587
H_0: Rem. feminine words on Transwomen	0.279	0.315	0.177	0.223	0.203
H_0: Rem. masculine words on Transwomen	0.613	0.462	0.420	0.469	0.429

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: Rem. feminine words on Ciswomen" tests the effect of Rem. feminine words on the subject group (Ciswomen).

C.10.2 Risk

Table C.35: OLS regression for all treatments. Dependent variable: Risk.

	(1)	(2)	(3)
Ciswomen	-0.063	-0.052	-0.065
	(0.155)	(0.170)	(0.156)
Transmen	-0.163	-0.164	-0.309
	(0.147)	(0.161)	(0.265)
Transwomen	0.270	0.235	0.162
	(0.205)	(0.205)	(0.290)
Rem. feminine words	0.058	0.062	0.060
	(0.037)	(0.037)	(0.037)
Rem. masculine words	0.101 *	0.103 *	0.100 *
	(0.041)	(0.041)	(0.041)
Rem. feminine words x Ciswomen	-0.031	-0.036	-0.031
	(0.046)	(0.046)	(0.046)
Rem. masculine words x Ciswomen	-0.064	-0.070	-0.062
	(0.052)	(0.052)	(0.052)
Rem. feminine words x Transmen	-0.019	-0.026	-0.022
	(0.046)	(0.046)	(0.046)
Rem. masculine words x Transmen	-0.092	-0.099 *	-0.092
	(0.049)	(0.049)	(0.049)
Rem. feminine words x Transwomen	-0.082	-0.084	-0.079
	(0.056)	(0.056)	(0.056)
Rem. masculine words x Transwomen	-0.056	-0.054	-0.057
	(0.061)	(0.062)	(0.061)
Const.	1.798 ***	1.552	1.752 ***
	(0.110)	(0.906)	(0.300)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.019	0.015	0.019
H_0: Rem. feminine words on Cismen	0.114	0.092	0.104
H_0: Rem. masculine words on Cismen	0.014	0.012	0.014
H_0: Rem. feminine words on Ciswomen	0.334	0.362	0.314
H_0: Rem. masculine words on Ciswomen	0.259	0.319	0.244
H_0: Rem. feminine words on Transmen	0.158	0.196	0.166
H_0: Rem. masculine words on Transmen	0.755	0.867	0.766
H_0: Rem. feminine words on Transwomen	0.558	0.587	0.640
H_0: Rem. masculine words on Transwomen	0.330	0.285	0.353

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: Rem. feminine words on Ciswomen" tests the effect of Rem. feminine words on the subject group (Ciswomen).

C.10.3 Altruism

Table C.36: OLS regression for all treatments. Dependent variable: Donations.

	(1)	(2)	(3)
Ciswomen	0.007	0.025	0.012
	(0.285)	(0.303)	(0.285)
Transmen	0.122	0.230	-0.014
	(0.284)	(0.298)	(0.420)
Transwomen	-0.225	-0.164	-0.403
	(0.324)	(0.324)	(0.438)
Rem. feminine words	-0.112	-0.115	-0.113
	(0.062)	(0.065)	(0.062)
Rem. masculine words	-0.054	-0.063	-0.055
	(0.071)	(0.073)	(0.071)
Rem. feminine words x Ciswomen	0.212 **	0.218 **	0.211 **
	(0.080)	(0.082)	(0.080)
Rem. masculine words x Ciswomen	0.100	0.123	0.098
	(0.090)	(0.092)	(0.090)
Rem. feminine words x Transmen	0.117	0.113	0.115
	(0.080)	(0.081)	(0.081)
Rem. masculine words x Transmen	0.032	0.041	0.030
	(0.090)	(0.090)	(0.090)
Rem. feminine words x Transwomen	0.141	0.164	0.139
	(0.096)	(0.097)	(0.096)
Rem. masculine words x Transwomen	0.084	0.093	0.085
	(0.102)	(0.102)	(0.103)
Const.	2.695 ***	1.848	3.150 ***
	(0.202)	(1.396)	(0.506)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	780	780	780
Adj. R2	0.005	0.030	0.004
H_0: Rem. feminine words on Cismen	0.072	0.076	0.071
H_0: Rem. masculine words on Cismen	0.450	0.390	0.441
H_0: Rem. feminine words on Ciswomen	0.049	0.041	0.051
H_0: Rem. masculine words on Ciswomen	0.412	0.279	0.433
H_0: Rem. feminine words on Transmen	0.921	0.975	0.971
H_0: Rem. masculine words on Transmen	0.697	0.684	0.659
H_0: Rem. feminine words on Transwomen	0.693	0.492	0.717
H_0: Rem. masculine words on Transwomen	0.685	0.673	0.681

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: Rem. feminine words on Ciswomen" tests the effect of Rem. feminine words on the subject group (Ciswomen).

C.11 Participants remembered at least 4 words

C.11.1 Competitiveness

Table C.37: Probit regression for all treatments. Dependent variable: Competition. The participants remembered at least 4 words in Part 1.

	(1)	(2)	(3)	(4)	(5)
Ciswomen	0.047	0.066	0.197	0.357	0.158
	(0.317)	(0.330)	(0.344)	(0.363)	(0.346)
Transmen	0.065	0.117	0.220	0.286	0.554
	(0.318)	(0.329)	(0.320)	(0.378)	(0.435)
Transwomen	0.122	0.223	0.208	0.264	0.650
	(0.377)	(0.392)	(0.416)	(0.413)	(0.516)
Treatment: FEMININE	0.361	0.189	0.158	0.212	0.136
	(0.306)	(0.309)	(0.302)	(0.326)	(0.305)
Treatment: MASCULINE	0.331	0.351	0.272	0.434	0.247
	(0.312)	(0.325)	(0.325)	(0.349)	(0.327)
FEMININE x Ciswomen	-0.378	-0.139	-0.094	-0.112	-0.055
	(0.404)	(0.421)	(0.446)	(0.465)	(0.448)
MASCULINE x Ciswomen	-0.413	-0.455	-0.383	-0.471	-0.327
	(0.412)	(0.432)	(0.449)	(0.459)	(0.450)
FEMININE x Transmen	-0.558	-0.375	-0.284	-0.314	-0.222
	(0.405)	(0.415)	(0.409)	(0.431)	(0.409)
MASCULINE x Transmen	-0.592	-0.660	-0.469	-0.496	-0.406
	(0.414)	(0.428)	(0.434)	(0.461)	(0.441)
FEMININE x Transwomen	-0.128	-0.021	0.139	0.053	0.132
	(0.466)	(0.476)	(0.494)	(0.492)	(0.507)
MASCULINE x Transwomen	-0.533	-0.707	-0.651	-0.779	-0.638
	(0.479)	(0.499)	(0.521)	(0.534)	(0.533)
Perf. tournament		0.006	-0.056 **	-0.059 **	-0.061 **
		(0.017)	(0.021)	(0.022)	(0.021)
Delta perf.		0.147 ***	0.114 ***	0.129 ***	0.118 ***
		(0.024)	(0.027)	(0.027)	(0.027)
Belief tournament			-0.602 ***	-0.612 ***	-0.604 ***
			(0.088)	(0.091)	(0.089)
Risk			0.165 **	0.166 **	0.164 **
			(0.054)	(0.055)	(0.054)
Const.	-0.595 *	-0.759 *	0.790	-2.259	-0.228
	(0.251)	(0.305)	(0.439)	(1.839)	(0.594)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	546	546	546	546	546
Pseudo R2 (McFadden)	0.017	0.104	0.189	0.231	0.200
H_0: FEMININE on Cismen	0.232	0.546	0.631	0.530	0.682
H_0: MASCULINE on Cismen	0.283	0.274	0.419	0.215	0.463
H_0: FEMININE on Ciswomen	0.947	0.853	0.822	0.732	0.778
H_0: MASCULINE on Ciswomen	0.758	0.710	0.704	0.902	0.786
H_0: FEMININE on Transmen	0.452	0.496	0.661	0.731	0.766
H_0: MASCULINE on Transmen	0.332	0.275	0.513	0.841	0.604
H_0: FEMININE on Transwomen	0.502	0.640	0.430	0.480	0.481
H_0: MASCULINE on Transwomen	0.574	0.336	0.327	0.373	0.317

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.11.2 Risk

Table C.38: OLS regression for all treatments. Dependent variable: Risk. The participants remembered at least 4 words in Part 1.

	(1)	(2)	(3)
Ciswomen	-0.148	-0.142	-0.154
	(0.284)	(0.297)	(0.285)
Transmen	-0.192	-0.217	-0.430
	(0.293)	(0.306)	(0.414)
Transwomen	0.083	0.069	-0.107
	(0.333)	(0.338)	(0.411)
Treatment: FEMININE	0.228	0.260	0.247
	(0.294)	(0.291)	(0.294)
Treatment: MASCULINE	0.525	0.577	0.517
	(0.299)	(0.300)	(0.299)
FEMININE x Ciswomen	-0.094	-0.120	-0.096
	(0.361)	(0.358)	(0.362)
MASCULINE x Ciswomen	-0.364	-0.417	-0.340
	(0.362)	(0.361)	(0.363)
FEMININE x Transmen	-0.081	-0.116	-0.093
	(0.364)	(0.362)	(0.365)
MASCULINE x Transmen	-0.469	-0.535	-0.459
	(0.369)	(0.370)	(0.370)
FEMININE x Transwomen	-0.289	-0.349	-0.262
	(0.415)	(0.418)	(0.411)
MASCULINE x Transwomen	-0.229	-0.236	-0.216
	(0.436)	(0.440)	(0.433)
Const.	1.883 ***	1.775	1.844 ***
	(0.240)	(1.280)	(0.416)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	546	546	546
Adj. R2	0.014	0.008	0.015
H_0: FEMININE on Cismen	0.437	0.373	0.402
H_0: MASCULINE on Cismen	0.080	0.055	0.084
H_0: FEMININE on Ciswomen	0.524	0.514	0.475
H_0: MASCULINE on Ciswomen	0.432	0.452	0.390
H_0: FEMININE on Transmen	0.494	0.512	0.481
H_0: MASCULINE on Transmen	0.798	0.854	0.790
H_0: FEMININE on Transwomen	0.836	0.764	0.956
H_0: MASCULINE on Transwomen	0.353	0.291	0.338

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.11.3 Altruism

Table C.39: OLS regression for all treatments. Dependent variable: Donations. The participants remembered at least 4 words in Part 1.

	(1)	(2)	(3)
Ciswomen	0.327	0.389	0.342
	(0.469)	(0.500)	(0.472)
Transmen	0.487	0.579	0.099
	(0.497)	(0.526)	(0.651)
Transwomen	-0.243	-0.116	-0.670
	(0.541)	(0.555)	(0.669)
Treatment: FEMININE	-0.349	-0.360	-0.335
	(0.477)	(0.490)	(0.478)
Treatment: MASCULINE	-0.079	-0.135	-0.082
	(0.484)	(0.499)	(0.484)
FEMININE x Ciswomen	0.745	0.769	0.726
	(0.604)	(0.609)	(0.605)
MASCULINE x Ciswomen	0.093	0.204	0.080
	(0.612)	(0.618)	(0.613)
FEMININE x Transmen	0.178	0.189	0.141
	(0.619)	(0.618)	(0.619)
MASCULINE x Transmen	-0.253	-0.149	-0.256
	(0.632)	(0.639)	(0.632)
FEMININE x Transwomen	0.591	0.669	0.603
	(0.704)	(0.716)	(0.705)
MASCULINE x Transwomen	0.204	0.199	0.218
	(0.701)	(0.718)	(0.704)
Const.	2.528 ***	1.844	3.301 ***
	(0.384)	(1.901)	(0.675)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	546	546	546
Adj. R2	0.004	0.015	0.004
H_0: FEMININE on Cismen	0.465	0.462	0.485
H_0: MASCULINE on Cismen	0.870	0.786	0.866
H_0: FEMININE on Ciswomen	0.286	0.266	0.293
H_0: MASCULINE on Ciswomen	0.971	0.853	0.996
H_0: FEMININE on Transmen	0.665	0.652	0.625
H_0: MASCULINE on Transmen	0.415	0.475	0.407
H_0: FEMININE on Transwomen	0.640	0.556	0.606
H_0: MASCULINE on Transwomen	0.805	0.902	0.790

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.12 Participants remembered less than 4 words

C.12.1 Competitiveness

Table C.40: Probit regression for all treatments. Dependent variable: Competition. The participants remembered less than 4 words in Part 1.

	(1)	(2)	(3)	(4)	(5)
Ciswomen	0.176	0.327	0.481	0.450	0.444
	(0.340)	(0.352)	(0.417)	(0.453)	(0.408)
Transmen	-0.254	-0.235	-0.094	-0.184	0.671
	(0.348)	(0.362)	(0.401)	(0.421)	(0.594)
Transwomen	-0.021	0.047	-0.051	-0.307	0.653
	(0.351)	(0.362)	(0.363)	(0.384)	(0.572)
Treatment: FEMININE	0.075	0.122	-0.034	-0.241	-0.076
	(0.380)	(0.425)	(0.440)	(0.475)	(0.451)
Treatment: MASCULINE	0.616	0.746 *	0.729 *	0.870 *	0.705 *
	(0.338)	(0.335)	(0.348)	(0.400)	(0.351)
FEMININE x Ciswomen	-0.590	-0.771	-0.844	-0.699	-0.721
	(0.606)	(0.655)	(0.740)	(0.800)	(0.745)
MASCULINE x Ciswomen	-0.882	-1.096 *	-1.124	-1.140	-1.050
	(0.545)	(0.551)	(0.606)	(0.687)	(0.612)
FEMININE x Transmen	0.421	0.431	0.554	0.639	0.631
	(0.603)	(0.681)	(0.672)	(0.708)	(0.689)
MASCULINE x Transmen	0.006	-0.116	-0.029	0.177	-0.002
	(0.552)	(0.543)	(0.543)	(0.616)	(0.561)
FEMININE x Transwomen	-0.338	-0.200	0.095	0.643	-0.206
	(0.617)	(0.641)	(0.630)	(0.637)	(0.647)
MASCULINE x Transwomen	-0.668	-0.857	-0.583	-0.741	-0.598
	(0.616)	(0.618)	(0.616)	(0.636)	(0.628)
Perf. tournament		-0.038	-0.097 ***	-0.109 ***	-0.088 **
		(0.024)	(0.029)	(0.033)	(0.030)
Delta perf.		0.118 **	0.105 **	0.107 **	0.104 **
		(0.036)	(0.038)	(0.041)	(0.039)
Belief tournament			-0.493 ***	-0.531 ***	-0.467 ***
			(0.129)	(0.148)	(0.132)
Risk			0.172	0.179	0.169
			(0.092)	(0.099)	(0.093)
Const.	-0.452 *	-0.273	0.975	0.666	0.490
	(0.204)	(0.291)	(0.509)	(2.158)	(0.944)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	234	234	234	234	234
Pseudo R2 (McFadden)	0.032	0.076	0.147	0.214	0.160
H_0: FEMININE on Cismen	0.840	0.750	0.931	0.551	0.849
H_0: MASCULINE on Cismen	0.061	0.029	0.036	0.024	0.043
H_0: FEMININE on Ciswomen	0.262	0.167	0.080	0.073	0.114
H_0: MASCULINE on Ciswomen	0.523	0.409	0.375	0.563	0.440
H_0: FEMININE on Transmen	0.278	0.232	0.273	0.432	0.252
H_0: MASCULINE on Transmen	0.143	0.144	0.115	0.032	0.118
H_0: FEMININE on Transwomen	0.579	0.873	0.910	0.483	0.627
H_0: MASCULINE on Transwomen	0.918	0.829	0.787	0.822	0.844

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.12.2 Risk

Table C.41: OLS regression for all treatments. Dependent variable: Risk. The participants remembered less than 4 words in Part 1.

	(1)	(2)	(3)
Ciswomen	-0.170	-0.141	-0.188
	(0.247)	(0.268)	(0.252)
Transmen	-0.128	-0.107	0.161
	(0.196)	(0.216)	(0.435)
Transwomen	0.755 *	0.651	1.044 *
	(0.370)	(0.392)	(0.510)
Treatment: FEMININE	-0.032	-0.053	-0.054
	(0.262)	(0.283)	(0.265)
Treatment: MASCULINE	0.026	0.036	0.014
	(0.236)	(0.249)	(0.234)
FEMININE x Ciswomen	0.220	0.258	0.275
	(0.472)	(0.482)	(0.479)
MASCULINE x Ciswomen	0.588	0.471	0.621
	(0.395)	(0.409)	(0.397)
FEMININE x Transmen	0.101	0.167	0.117
	(0.406)	(0.429)	(0.411)
MASCULINE x Transmen	-0.349	-0.461	-0.345
	(0.322)	(0.340)	(0.324)
FEMININE x Transwomen	-0.837	-0.734	-0.909
	(0.507)	(0.536)	(0.499)
MASCULINE x Transwomen	-0.449	-0.337	-0.462
	(0.676)	(0.662)	(0.671)
Const.	1.768 ***	2.141	1.537 **
	(0.157)	(1.294)	(0.542)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	234	234	234
Adj. R2	0.032	0.029	0.027
H_0: FEMININE on Cismen	0.902	0.851	0.840
H_0: MASCULINE on Cismen	0.913	0.886	0.951
H_0: FEMININE on Ciswomen	0.633	0.597	0.580
H_0: MASCULINE on Ciswomen	0.054	0.121	0.049
H_0: FEMININE on Transmen	0.826	0.730	0.843
H_0: MASCULINE on Transmen	0.142	0.072	0.145
H_0: FEMININE on Transwomen	0.047	0.081	0.024
H_0: MASCULINE on Transwomen	0.505	0.624	0.478

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.12.3 Altruism

Table C.42: OLS regression for all treatments. Dependent variable: Donations. The participants remembered less than 4 words in Part 1.

	(1)	(2)	(3)
Ciswomen	-0.095	0.081	-0.126
	(0.547)	(0.625)	(0.554)
Transmen	-0.505	-0.169	-0.034
	(0.479)	(0.482)	(0.695)
Transwomen	0.155	0.193	0.628
	(0.539)	(0.557)	(0.705)
Treatment: FEMININE	-0.252	-0.076	-0.287
	(0.500)	(0.545)	(0.506)
Treatment: MASCULINE	-0.421	0.009	-0.440
	(0.519)	(0.559)	(0.520)
FEMININE x Ciswomen	0.360	0.231	0.449
	(0.849)	(0.879)	(0.867)
MASCULINE x Ciswomen	0.578	0.133	0.632
	(0.818)	(0.873)	(0.820)
FEMININE x Transmen	0.675	0.462	0.700
	(0.810)	(0.797)	(0.807)
MASCULINE x Transmen	1.335	0.723	1.342
	(0.775)	(0.765)	(0.775)
FEMININE x Transwomen	-0.055	-0.093	-0.171
	(0.828)	(0.857)	(0.816)
MASCULINE x Transwomen	0.416	0.291	0.394
	(1.016)	(0.969)	(1.014)
Const.	2.791 ***	0.629	2.396 **
	(0.322)	(2.381)	(0.915)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	234	234	234
Adj. R2	-0.033	0.006	-0.039
H_0: FEMININE on Cismen	0.614	0.889	0.571
H_0: MASCULINE on Cismen	0.418	0.987	0.399
H_0: FEMININE on Ciswomen	0.875	0.825	0.817
H_0: MASCULINE on Ciswomen	0.804	0.831	0.762
H_0: FEMININE on Transmen	0.508	0.514	0.515
H_0: MASCULINE on Transmen	0.114	0.184	0.118
H_0: FEMININE on Transwomen	0.642	0.796	0.480
H_0: MASCULINE on Transwomen	0.995	0.710	0.959

Note: Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious cisman, who earns more than 20K GBP, and lives in continental Europe. In the last column from the right, the baseline is a cisman. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H_0: FEMININE on Ciswomen" tests the effect of the treatment (FEMININE) on the subject group (Ciswomen).

C.13 Gender congruent upbringing

C.13.1 Competitiveness

Table C.43: Probit regression for NEUTRAL. Dependent variable: Competition. Only transgender participants are considered.

	(1)	(2)	(3)	(4)	(5)
Gender congruent upbringing	0.054	0.004	-0.118	-0.279	-0.187
	(0.295)	(0.301)	(0.353)	(0.427)	(0.375)
Perf. tournament		0.017	-0.103	-0.095	-0.100
		(0.039)	(0.056)	(0.062)	(0.058)
Delta perf.		0.101	0.060	0.033	0.056
		(0.056)	(0.066)	(0.080)	(0.068)
Belief tournament			-0.819 ***	-0.958 ***	-0.793 ***
			(0.220)	(0.272)	(0.228)
Risk			0.252 *	0.298 *	0.269 *
			(0.111)	(0.127)	(0.108)
Age				0.022	
				(0.018)	
Height				-0.001	
				(0.012)	
Student status				-0.292	
				(0.356)	
Income: < 20,000 GBP				0.230	
				(0.422)	
Religion: Religious				0.327	
				(0.361)	
Religion: Not say				6.768 ***	
				(0.499)	
Residence: US				0.027	
				(0.501)	
Residence: UK				0.104	
				(0.450)	
Residence: Other				0.226	
				(0.519)	
TCS					0.168
					(0.200)
STT					-0.028
					(0.042)
Const.	-0.557 ***	-0.735 *	1.713	1.172	1.385
	(0.141)	(0.349)	(0.938)	(2.643)	(1.063)
N	116	116	116	116	116
Pseudo R2 (McFadden)	0.000	0.048	0.196	0.306	0.202

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.44: Probit regression for all treatments. Dependent variable: Competition. Only transgender participants are considered.

	(1)	(2)	(3)	(4)	(5)
Gender congruent upbringing	0.054	0.005	-0.108	-0.120	-0.279
	(0.295)	(0.318)	(0.344)	(0.381)	(0.358)
Treatment: FEMININE	0.035	0.056	0.182	0.204	0.143
	(0.197)	(0.209)	(0.217)	(0.217)	(0.221)
Treatment: MASCULINE	-0.099	-0.161	-0.083	-0.081	-0.056
	(0.207)	(0.212)	(0.233)	(0.248)	(0.237)
FEMININE x Gender congruent upbringing	-0.028	0.021	-0.102	-0.087	0.056
	(0.408)	(0.431)	(0.471)	(0.505)	(0.482)
MASCULINE x Gender congruent upbringing	0.035	0.075	0.155	0.192	0.161
	(0.402)	(0.421)	(0.450)	(0.490)	(0.468)
Perf. tournament		-0.022	-0.114 ***	-0.117 ***	-0.111 ***
		(0.022)	(0.029)	(0.029)	(0.030)
Delta perf.		0.162 ***	0.132 ***	0.141 ***	0.135 ***
		(0.032)	(0.036)	(0.038)	(0.035)
Belief tournament			-0.727 ***	-0.730 ***	-0.705 ***
			(0.112)	(0.120)	(0.114)
Risk			0.196 **	0.222 **	0.194 **
			(0.065)	(0.069)	(0.065)
Const.	-0.557 ***	-0.430	1.669 ***	-0.365	1.139 *
	(0.141)	(0.229)	(0.482)	(1.433)	(0.549)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	–	–	Yes	–
Controls (TCS, STT)	–	–	–	–	Yes
N	355	355	355	355	355
Pseudo R2 (McFadden)	0.001	0.079	0.198	0.233	0.209
H_0: FEMININE on GCU	0.983	0.833	0.840	0.778	0.624
H_0: MASCULINE on GCU	0.852	0.809	0.852	0.781	0.791

Note: Competition is a binary variable equal to 1 if the participant enters the tournament in Part 4 and 0 otherwise. Delta perf. is the difference in performance between Part 3 (tournament) and Part 2 (piece-rate). Belief tournament is the participants' belief of their performance rank within their group in Part 3, where the value 1 represents the rank with the highest performance. Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H0: FEMININE on GCU" tests the effect of the treatment (FEMININE) on the group of participants that were brought up gender congruently (Gender congruent upbringing=1).

C.13.2 Risk

Table C.45: OLS regression for NEUTRAL. Dependent variable: Risk. Only transgender participants are considered.

	(1)	(2)	(3)
Gender congruent upbringing	-0.529 *	-0.485 *	-0.445
	(0.236)	(0.236)	(0.285)
Age		-0.006	
		(0.016)	
Height		0.000	
		(0.014)	
Student status		0.121	
		(0.260)	
Income: < 20,000 GBP		-0.008	
		(0.266)	
Religion: Religious		0.244	
		(0.228)	
Religion: Not say		-0.597	
		(0.559)	
Residence: US		-0.091	
		(0.341)	
Residence: UK		0.026	
		(0.312)	
Residence: Other		0.845 *	
		(0.425)	
TCS			-0.066
			(0.150)
STT			-0.019
			(0.037)
Const.	2.008 ***	1.894	2.343 ***
	(0.126)	(2.359)	(0.415)
N	116	116	116
Adj. R2	0.027	0.032	0.019

Note: Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.46: OLS regression for all treatments. Dependent variable: Risk. Only transgender participants are considered.

	(1)	(2)	(3)
Gender congruent upbringing	-0.529 *	-0.535 *	-0.588 *
	(0.236)	(0.236)	(0.255)
Treatment: FEMININE	-0.229	-0.201	-0.239
	(0.166)	(0.163)	(0.169)
Treatment: MASCULINE	-0.144	-0.116	-0.138
	(0.184)	(0.184)	(0.184)
FEMININE x Gender congruent upbringing	0.726 *	0.737 *	0.773 *
	(0.323)	(0.321)	(0.335)
MASCULINE x Gender congruent upbringing	0.583	0.556	0.582
	(0.333)	(0.334)	(0.335)
Const.	2.008 ***	0.453	1.825 ***
	(0.126)	(1.201)	(0.238)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	355	355	355
Adj. R2	0.001	0.014	-0.001
H_0: FEMININE on GCU	0.074	0.055	0.063
H_0: MASCULINE on GCU	0.116	0.115	0.115

Note: Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H0: FEMININE on GCU" tests the effect of the treatment (FEMININE) on the group of participants that were brought up gender congruently (Gender congruent upbringing=1).

C.13.3 Altruism

Table C.47: OLS regression for NEUTRAL. Dependent variable: Donations. Only transgender participants are considered.

	(1)	(2)	(3)
Gender congruent upbringing	0.141	0.019	0.248
	(0.428)	(0.423)	(0.486)
Age		0.056 *	
		(0.024)	
Height		-0.012	
		(0.016)	
Student status		0.917 *	
		(0.392)	
Income: < 20,000 GBP		0.540	
		(0.419)	
Religion: Religious		1.100 **	
		(0.362)	
Religion: Not say		1.344	
		(1.498)	
Residence: US		-0.735	
		(0.538)	
Residence: UK		-0.335	
		(0.553)	
Residence: Other		-0.303	
		(0.663)	
TCS			-0.174
			(0.259)
STT			0.013
			(0.054)
Const.	2.675 ***	2.480	3.013 ***
	(0.214)	(2.900)	(0.668)
N	116	116	116
Adj. R2	-0.008	0.087	-0.021

Note: Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. In the second last column from the right, the baseline is a non-student, non-religious person, who earns more than 20K GBP, and lives in continental Europe. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table C.48: OLS regression for all treatments. Dependent variable: Donations. Only transgender participants are considered.

	(1)	(2)	(3)
Gender congruent upbringing	0.141	0.069	0.229
	(0.428)	(0.410)	(0.449)
Treatment: FEMININE	-0.046	-0.004	-0.074
	(0.293)	(0.284)	(0.293)
Treatment: MASCULINE	-0.337	-0.343	-0.355
	(0.304)	(0.301)	(0.305)
FEMININE x Gender congruent upbringing	0.213	0.203	0.197
	(0.587)	(0.551)	(0.598)
MASCULINE x Gender congruent upbringing	1.029	1.158 *	1.038
	(0.568)	(0.547)	(0.568)
Const.	2.675 ***	2.464	3.038 ***
	(0.214)	(1.690)	(0.395)
Controls (Age, Height, Student status, Income, Religion, Residence)	–	Yes	–
Controls (TCS, STT)	–	–	Yes
N	355	355	355
Adj. R2	0.013	0.076	0.012
H_0: FEMININE on GCU	0.742	0.677	0.812
H_0: MASCULINE on GCU	0.150	0.076	0.155

Note: Gender congruent upbringing is a binary variable equal to 1 if the way the participant's parents treated the participant matches the reported gender of the participant or the parents treated their child neutrally. Standard errors in parentheses are heteroskedasticity robust. Rows starting with H_0 report the p-values of a joint coefficient test that the coefficients' sum equals 0. For example, "H0: FEMININE on GCU" tests the effect of the treatment (FEMININE) on the group of participants that were brought up gender congruently (Gender congruent upbringing=1).

C.14 Additional information

C.14.1 Study sample

We recruited a total of 798 participants. Please note that due to a technical problem with how the participant's performance was shown on their screen, we exclude $n = 3$ cisgender and $n = 6$ transgender observations. We tested with our debriefing questionnaire whether the participants had an idea about the aim of the study, the study topic, etc. Eight cisgender and one transgender participant(s) wrote to think that s/he were primed. These $n = 9$ observations are also excluded from our analysis. Thus, the final number of subjects by subject groups and treatment is 780, as summarized in Table C.49.

To have comparable transgender and cisgender observations, we first collected the majority of the transgender observations, including their main demographic characteristics (age, student status, education, income, religious affiliation, and residence). We then used Prolific's sorting tool to recruit a similar cisgender sample based on those criteria.

Table C.49: Distribution of subject groups across treatments

	Treatment			
	NEUTRAL	FEMININE	MASCULINE	Total
Cismen	72	71	71	214
Ciswomen	71	70	70	211
Transmen	72	72	71	215
Transwomen	44	50	46	140
Total	259	263	258	780

C.14.2 Datasets

gender_data.csv is the main data set. The file contains n = 798 observations and 103 variables. Details on the variables can be found in the second dataset codebook.csv.

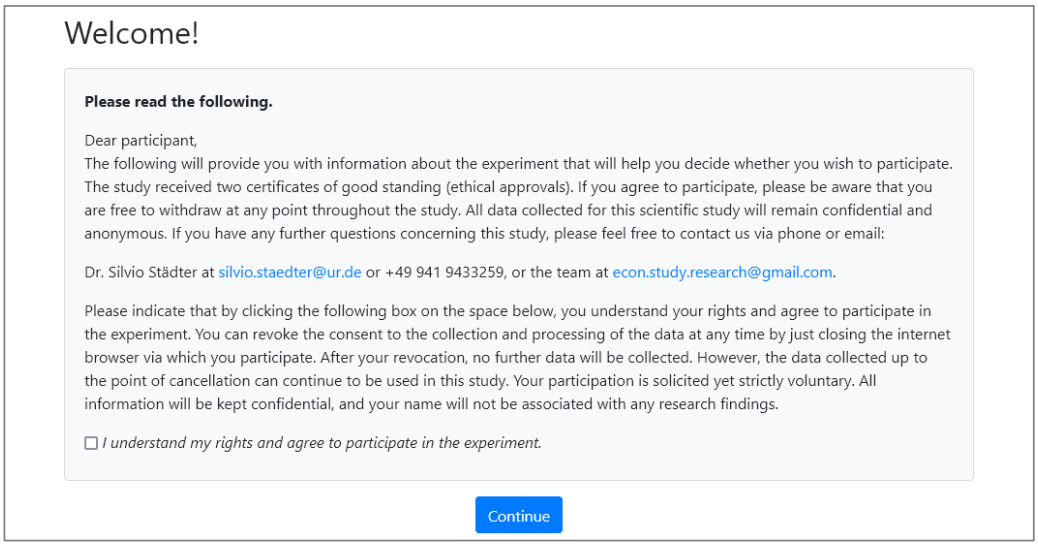
codebook.csv This file provides details on the variables of the main data set. Each row includes explanations for one of the 103 variables. Additionally, the third column summarizes the response options the subjects had.

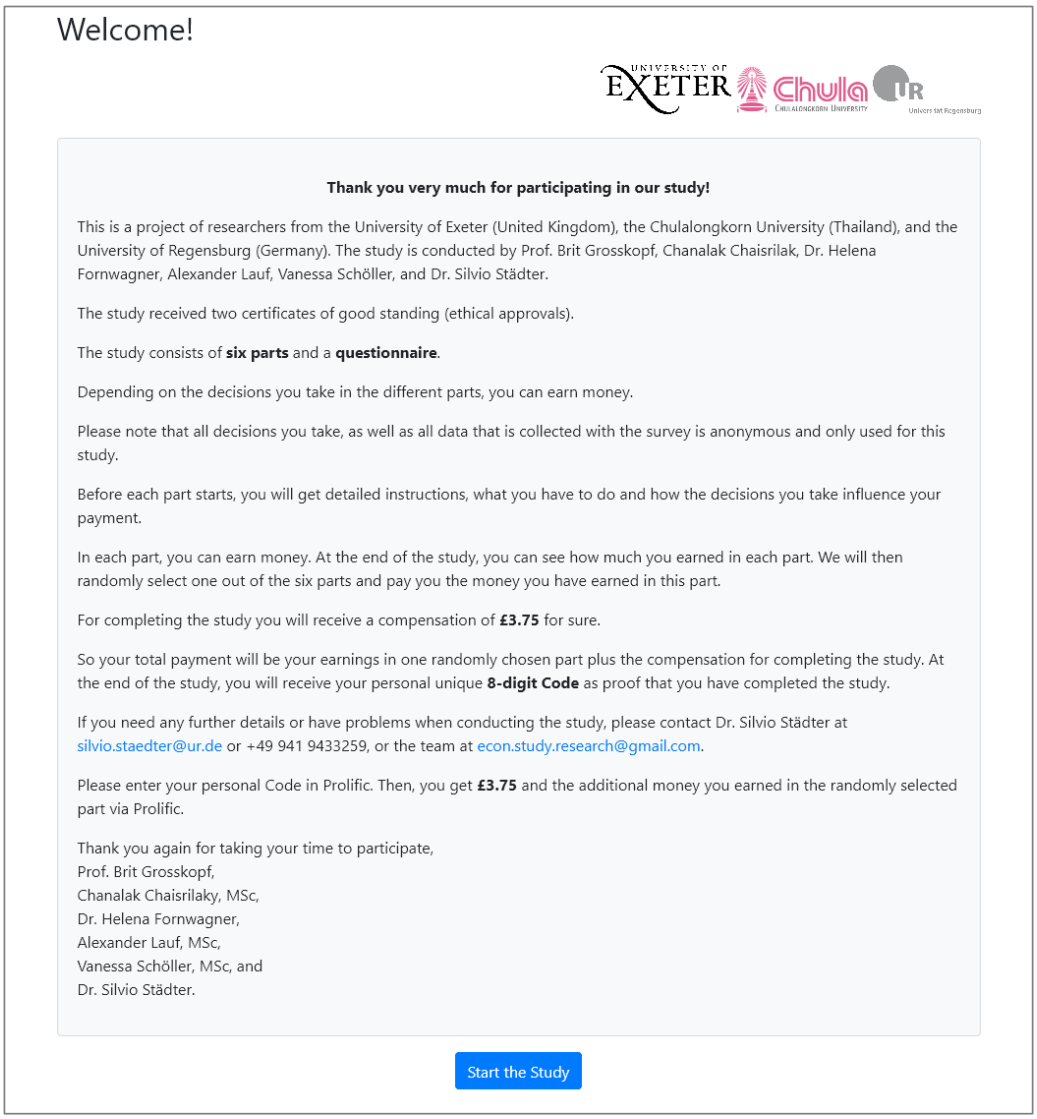
The collected data and additional material is available at OSF (https://osf.io/tyzjh/?view_only=66a8abca5f6a4aeead68f6fef19a0ee9).

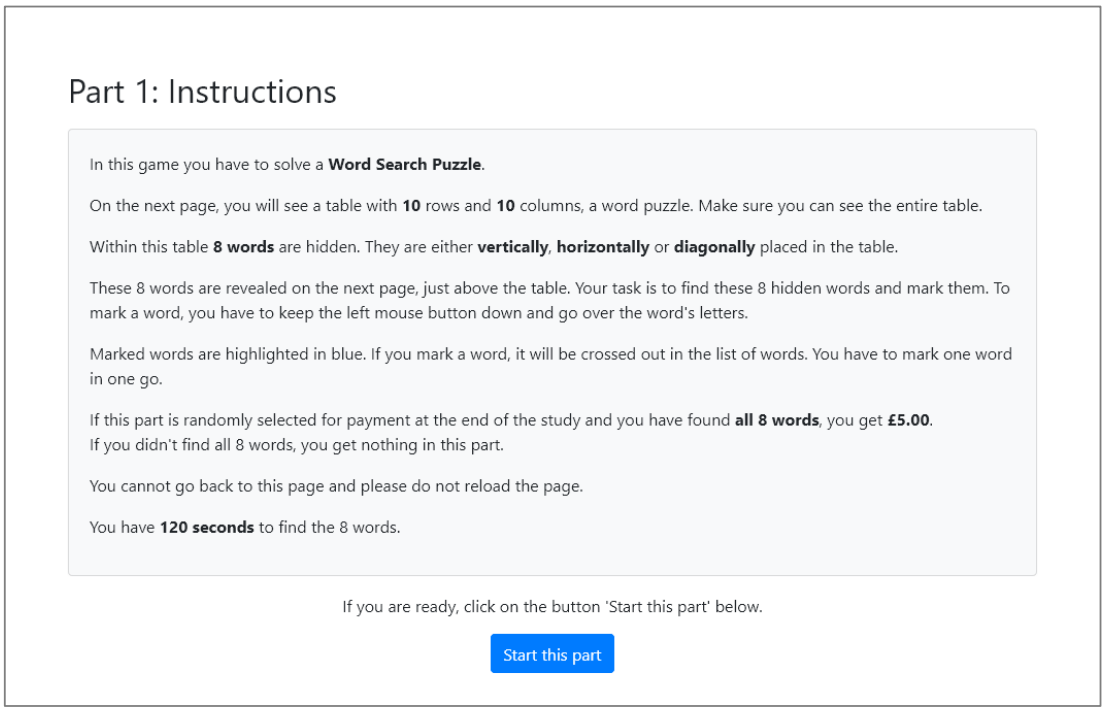
C.14.3 Instructions

The following pages contain screenshots of the online study conducted on the platform Prolific. Please note that one participant was randomly allocated to one treatment. Thus, one participant saw one out of the three different treatment pages. In addition, depending on the choice made

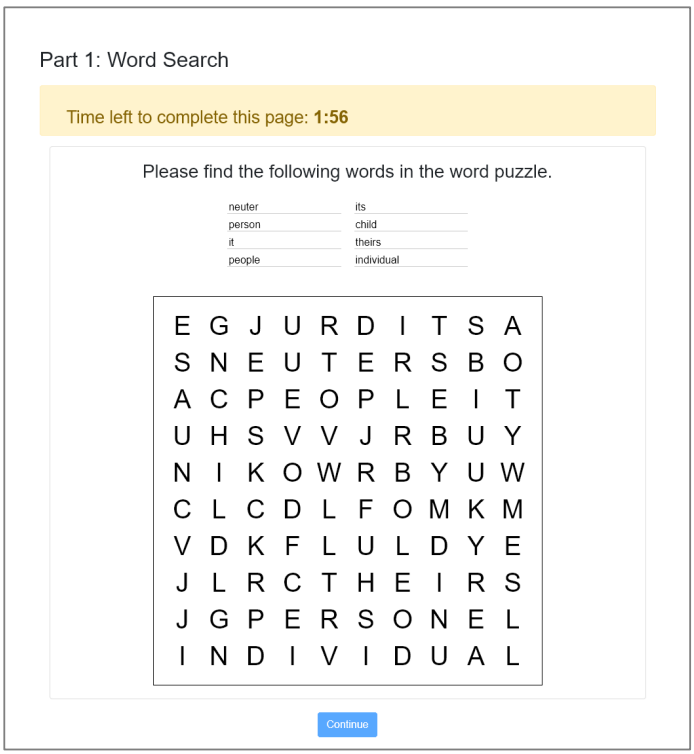
in Part 4, the system showed either the pages marked as Option A or Option B. A blue headline marks the varying screens. All other pages were identical.



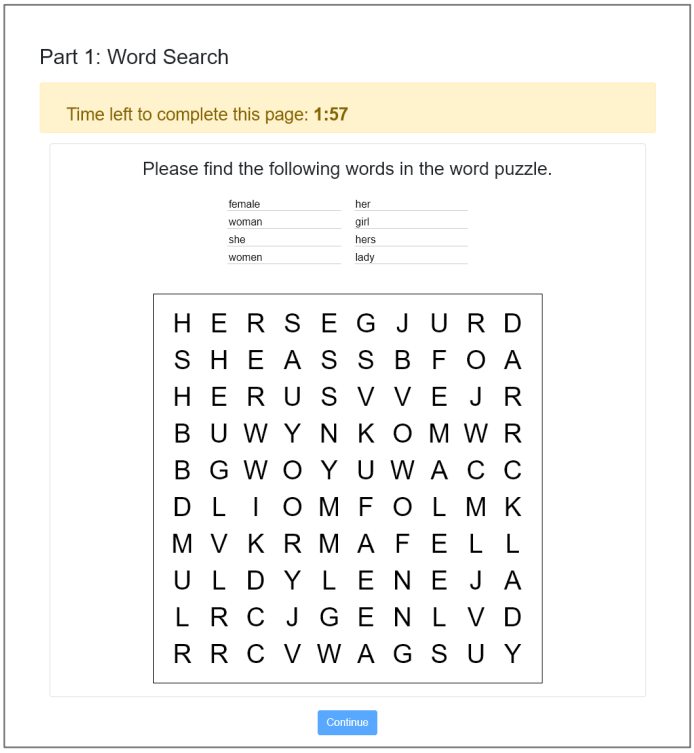




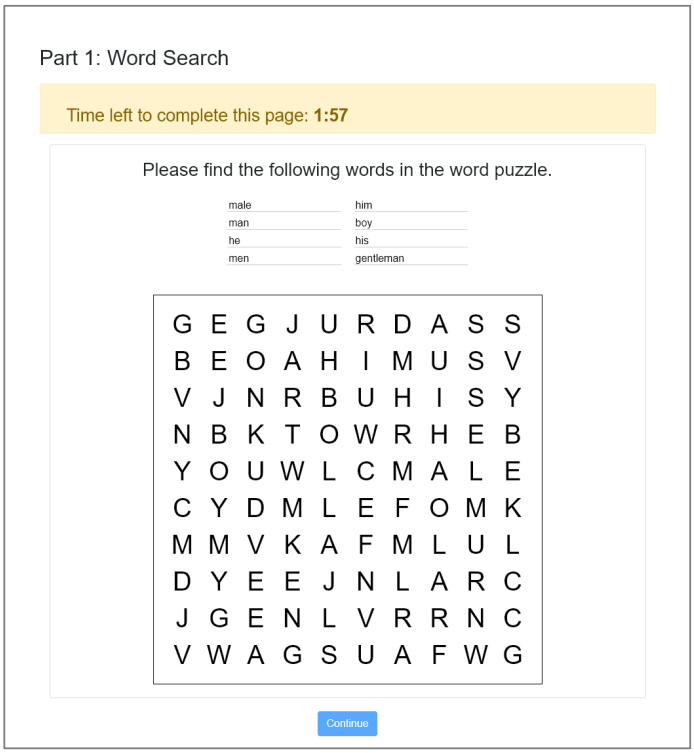
Treatment: NEUTRAL

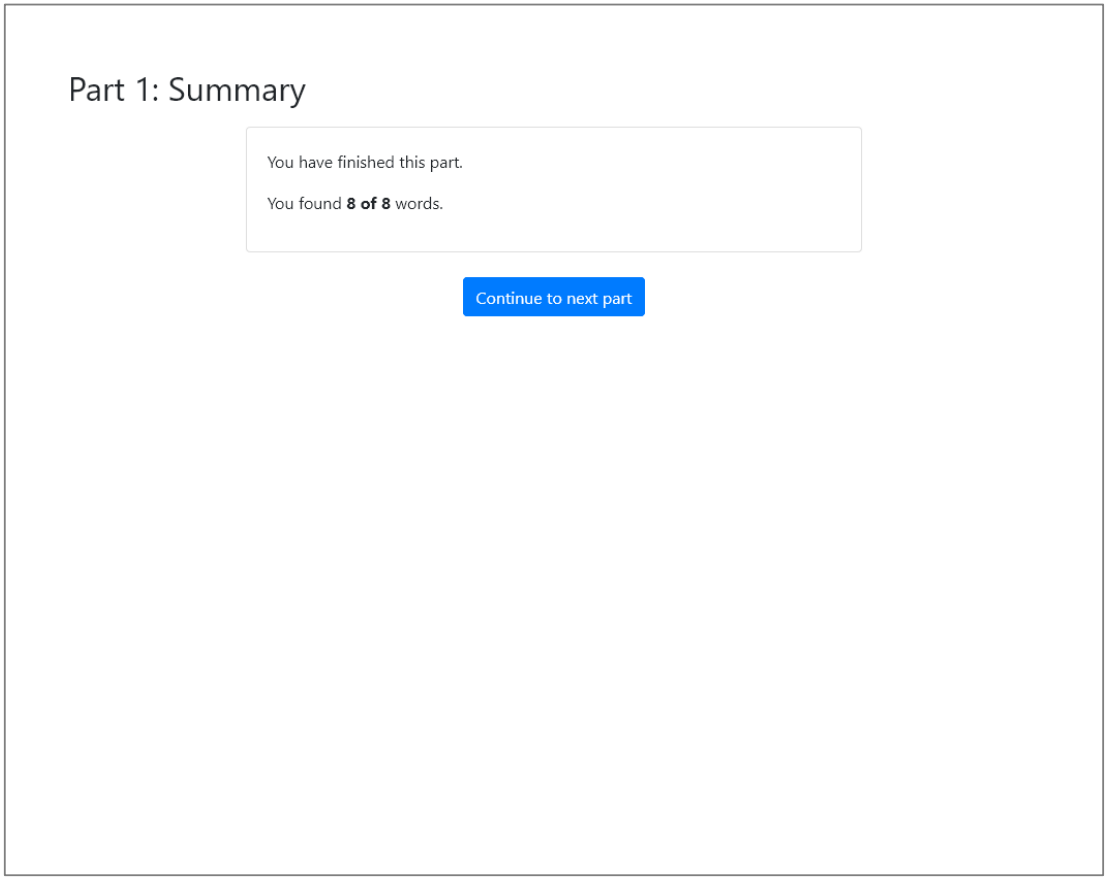


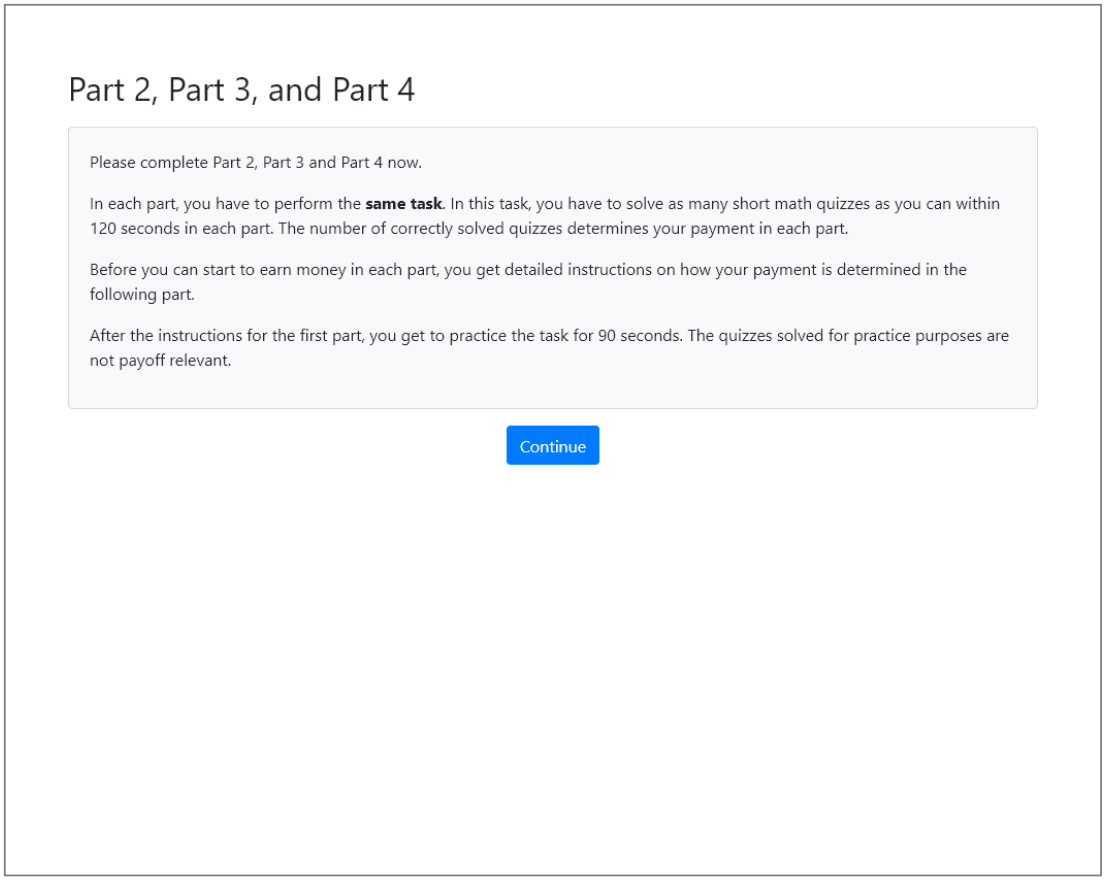
Treatment: FEMININE

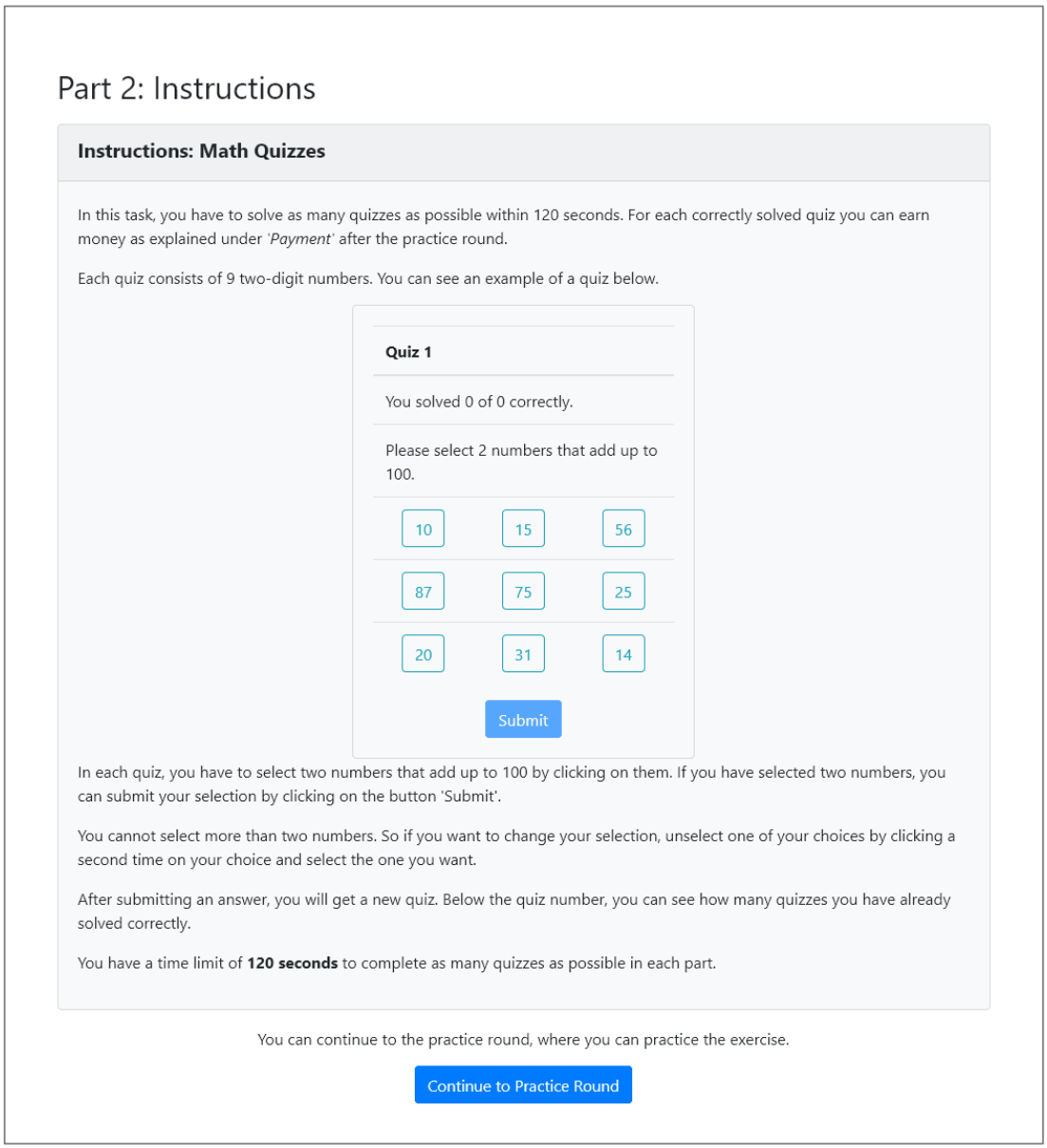


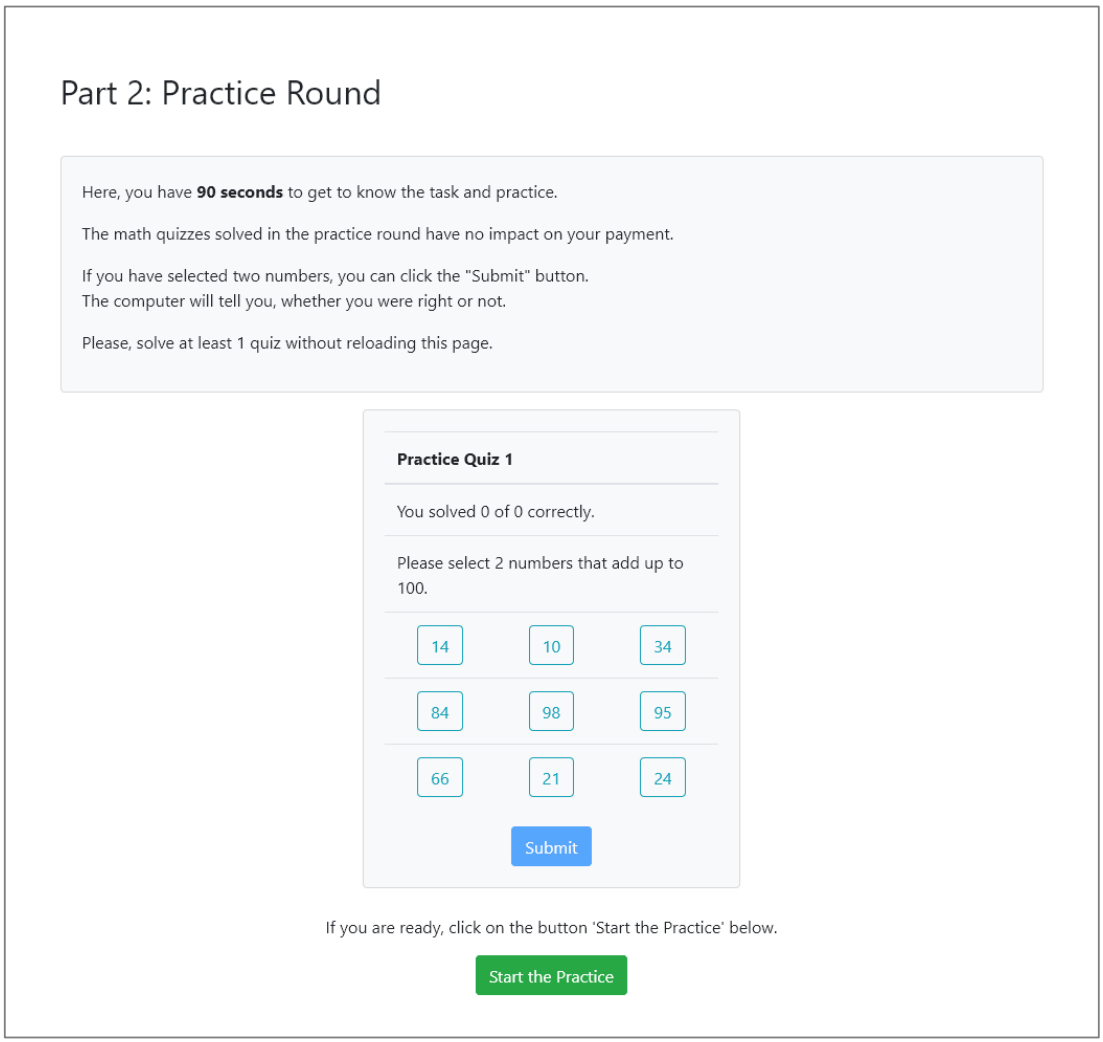
Treatment: MASCULINE

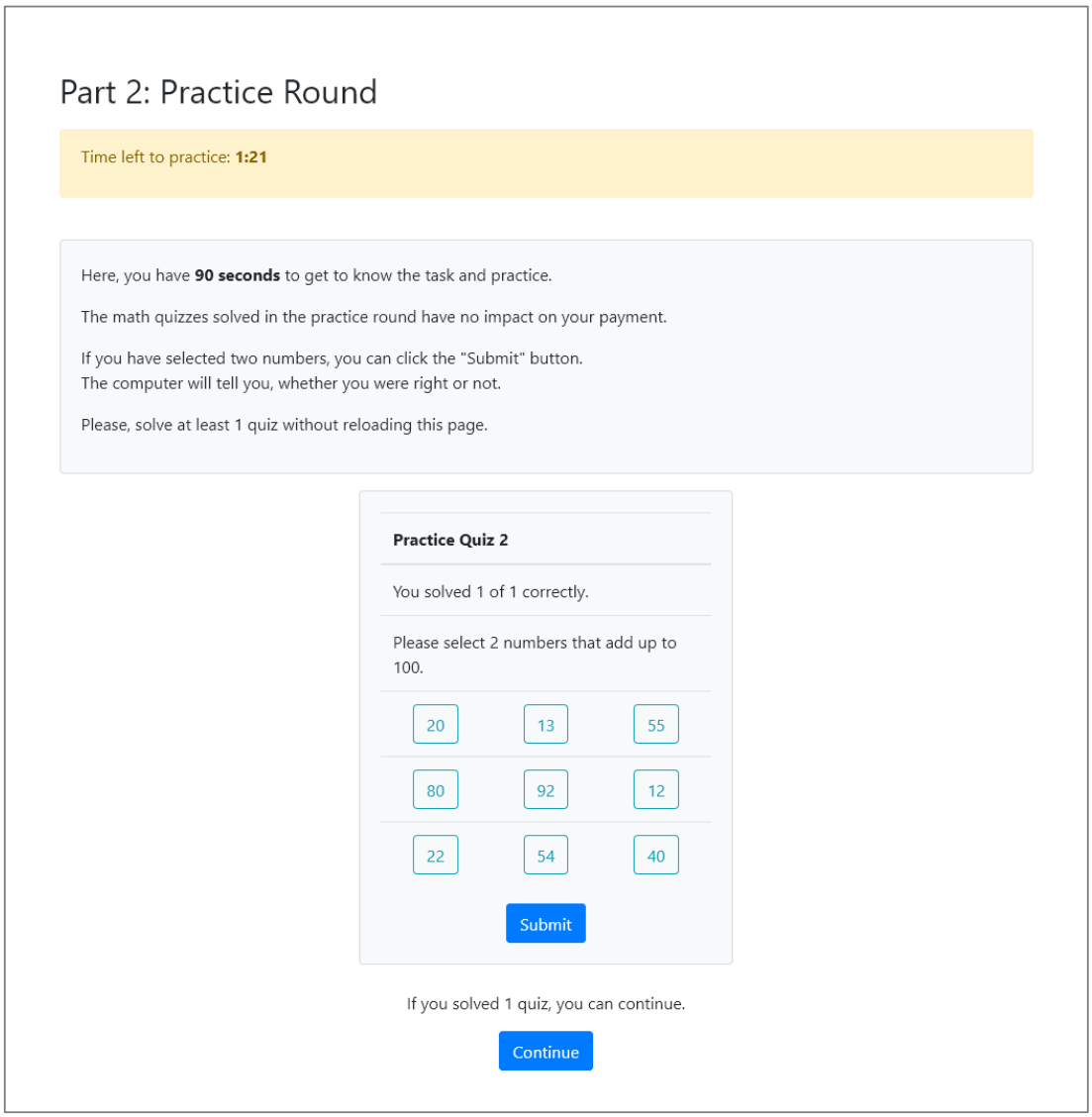


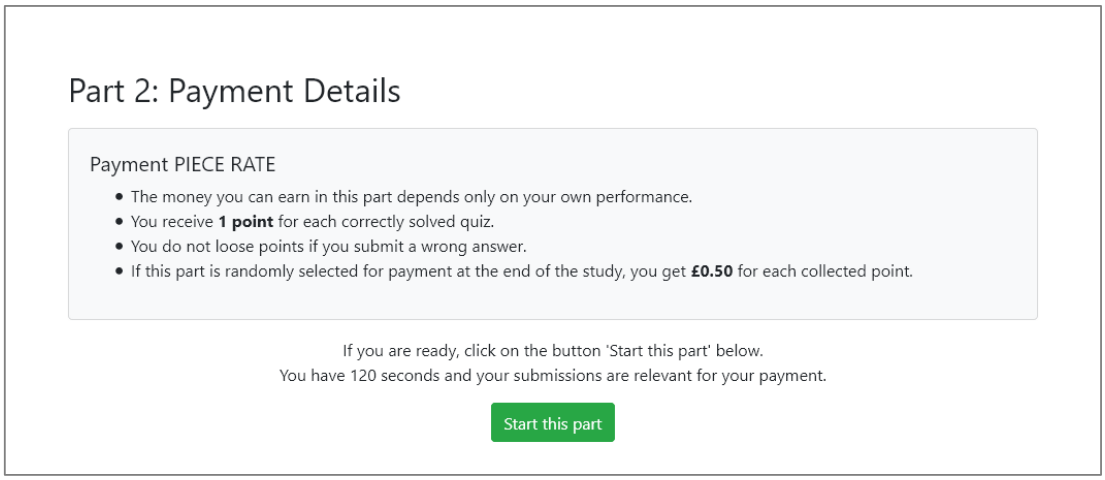

Part 1: Summary
You have finished this part.
You found 8 of 8 words.
Continue to next part

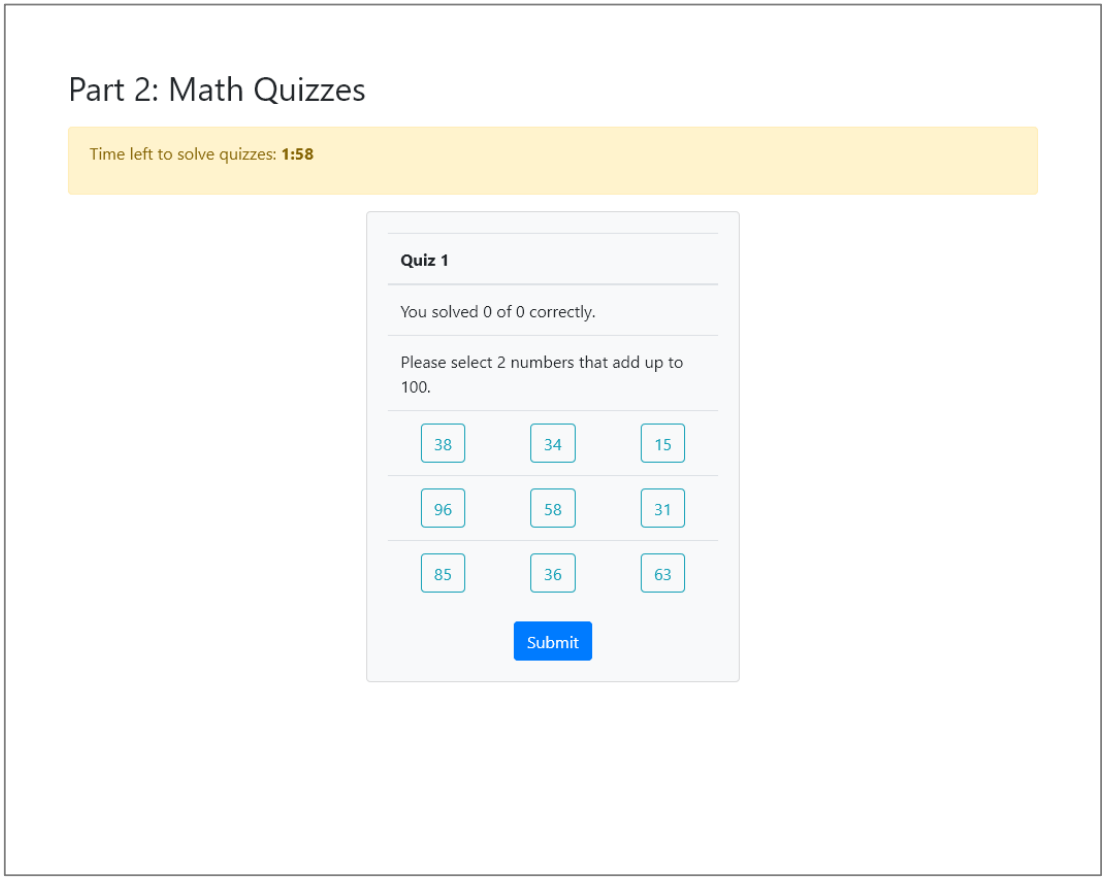

Part 2, Part 3, and Part 4
Please complete Part 2, Part 3 and Part 4 now.
In each part, you have to perform the same task. In this task, you have to solve as many short math quizzes as you can within 120 seconds in each part. The number of correctly solved quizzes determines your payment in each part.
Before you can start to earn money in each part, you get detailed instructions on how your payment is determined in the following part.
After the instructions for the first part, you get to practice the task for 90 seconds. The quizzes solved for practice purposes are not payoff relevant.
Continue

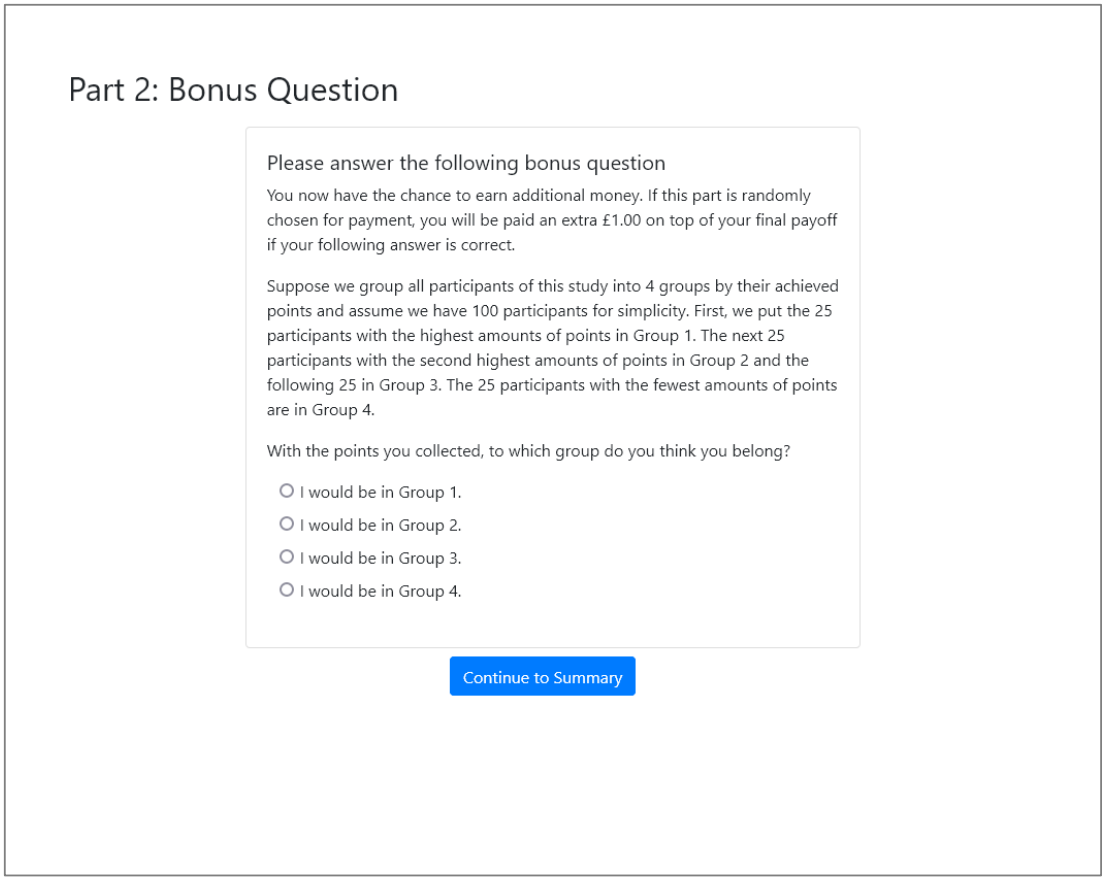

Part 2: Instructions
Instructions: Math Quizzes
In this task, you have to solve as many quizzes as possible within 120 seconds. For each correctly solved quiz you can earn money as explained under 'Payment' after the practice round.
Each quiz consists of 9 two-digit numbers. You can see an example of a quiz below.
Quiz 1
You solved 0 of 0 correctly.
Please select 2 numbers that add up to 100.
10
15
56
87
75
25
20
31
14
Submit
In each quiz, you have to select two numbers that add up to 100 by clicking on them. If you have selected two numbers, you can submit your selection by clicking on the button 'Submit'.
You cannot select more than two numbers. So if you want to change your selection, unselect one of your choices by clicking a second time on your choice and select the one you want.
After submitting an answer, you will get a new quiz. Below the quiz number, you can see how many quizzes you have already solved correctly.
You have a time limit of 120 seconds to complete as many quizzes as possible in each part.
You can continue to the practice round, where you can practice the exercise.
Continue to Practice Round

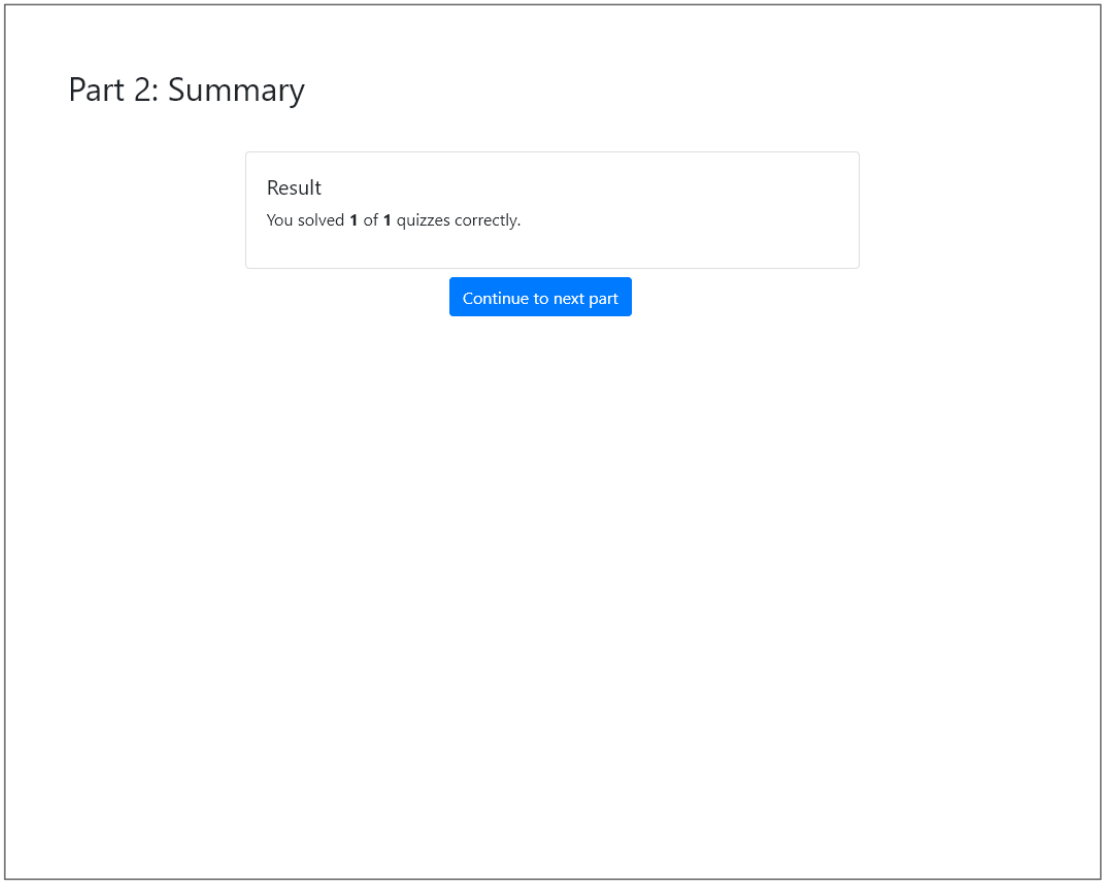

Part 2: Practice Round
Here, you have 90 seconds to get to know the task and practice.
The math quizzes solved in the practice round have no impact on your payment.
If you have selected two numbers, you can click the "Submit" button.
The computer will tell you, whether you were right or not.
Please, solve at least 1 quiz without reloading this page.
Practice Quiz 1
You solved 0 of 0 correctly.
Please select 2 numbers that add up to 100.
14
10
34
84
98
95
66
21
24
Submit
If you are ready, click on the button 'Start the Practice' below.
Start the Practice

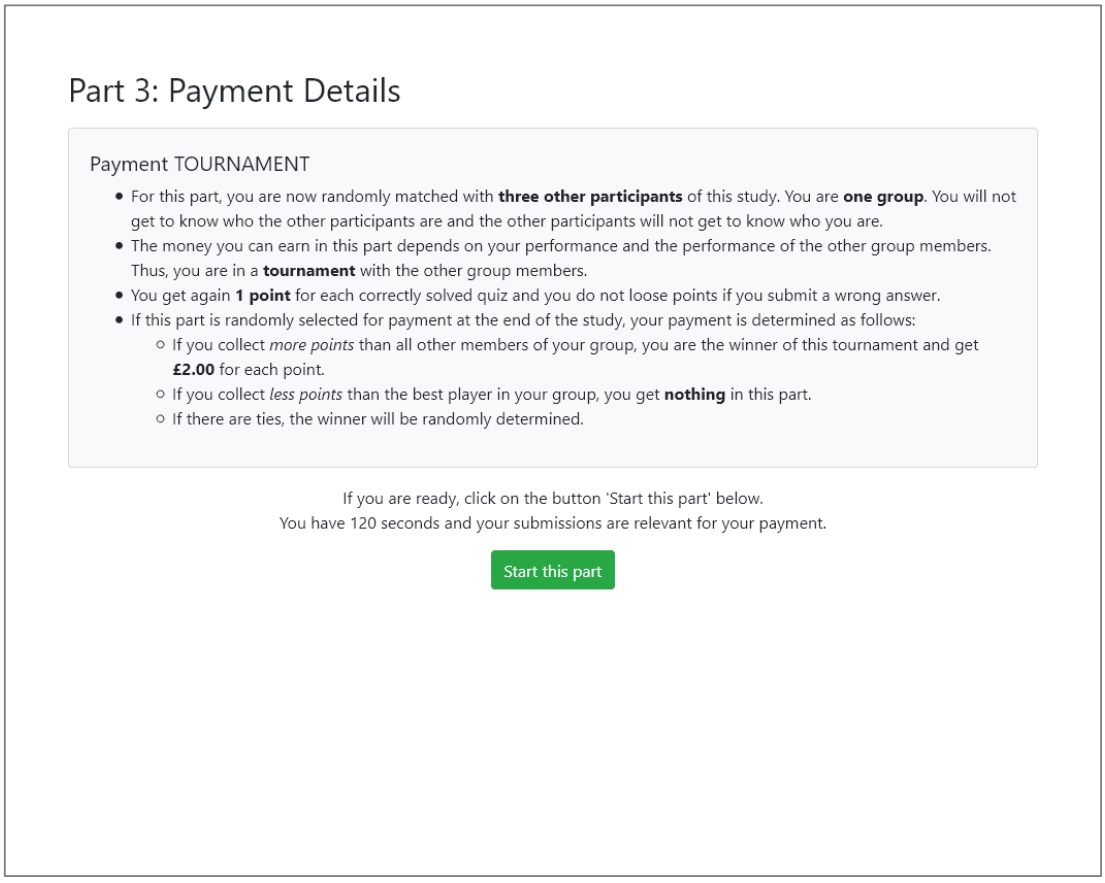

Part 2: Practice Round
Time left to practice: 1:21
Here, you have 90 seconds to get to know the task and practice.
The math quizzes solved in the practice round have no impact on your payment.
If you have selected two numbers, you can click the "Submit" button.
The computer will tell you, whether you were right or not.
Please, solve at least 1 quiz without reloading this page.
Practice Quiz 2
You solved 1 of 1 correctly.
Please select 2 numbers that add up to 100.
20
13
55
80
92
12
22
54
40
Submit
If you solved 1 quiz, you can continue.
Continue

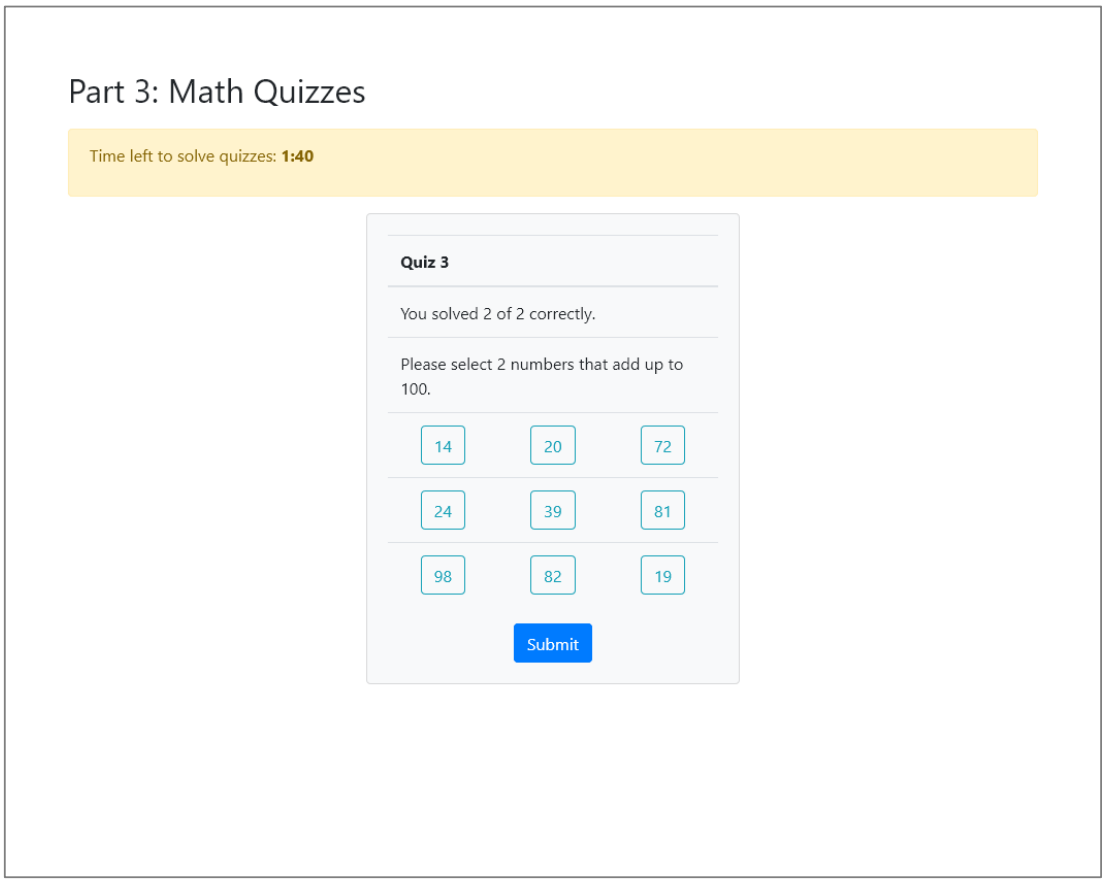

Part 2: Payment Details
Payment PIECE RATE
The money you can earn in this part depends only on your own performance.
You receive 1 point for each correctly solved quiz.
You do not loose points if you submit a wrong answer.
If this part is randomly selected for payment at the end of the study, you get £0.50 for each collected point.
If you are ready, click on the button 'Start this part' below.
You have 120 seconds and your submissions are relevant for your payment.
Start this part

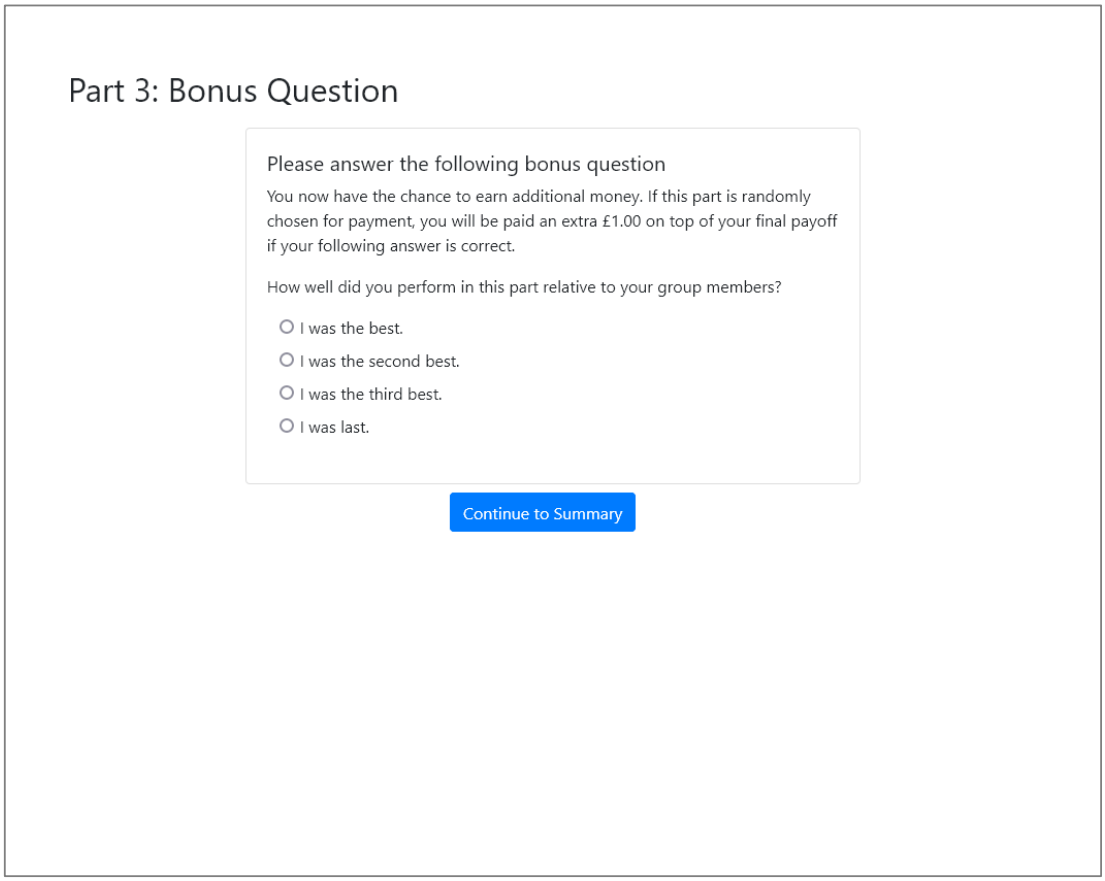

Part 2: Math Quizzes
Time left to solve quizzes: 1:58
Quiz 1
You solved 0 of 0 correctly.
Please select 2 numbers that add up to 100.
38
34
15
96
58
31
85
36
63
Submit

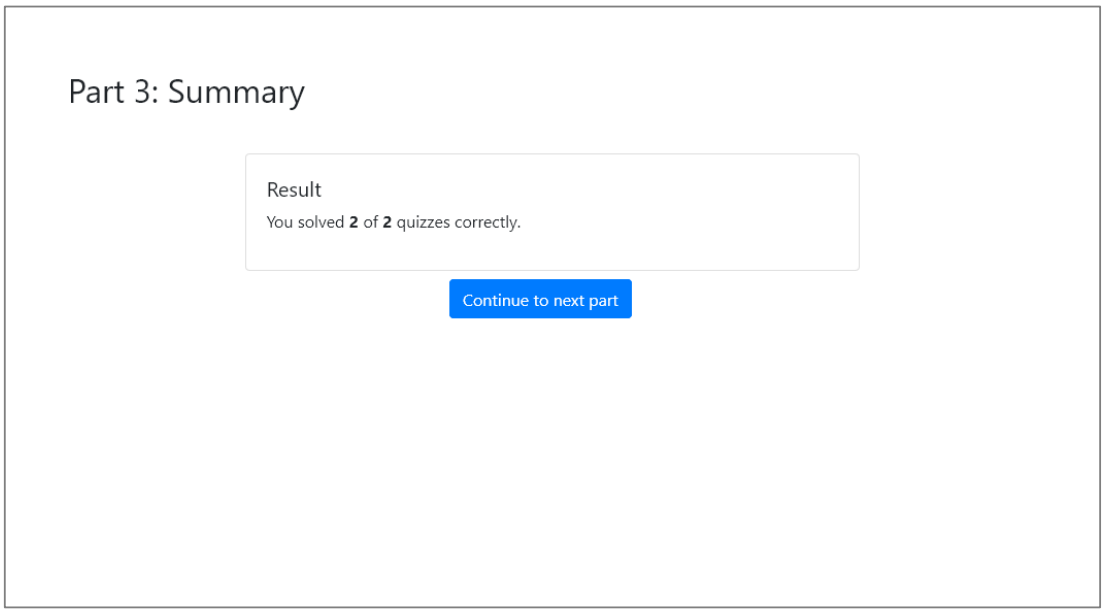

Part 2: Bonus Question
Please answer the following bonus question
You now have the chance to earn additional money. If this part is randomly chosen for payment, you will be paid an extra £1.00 on top of your final payoff if your following answer is correct.
Suppose we group all participants of this study into 4 groups by their achieved points and assume we have 100 participants for simplicity. First, we put the 25 participants with the highest amounts of points in Group 1. The next 25 participants with the second highest amounts of points in Group 2 and the following 25 in Group 3. The 25 participants with the fewest amounts of points are in Group 4.
With the points you collected, to which group do you think you belong?
I would be in Group 1.
I would be in Group 2.
I would be in Group 3.
I would be in Group 4.
Continue to Summary

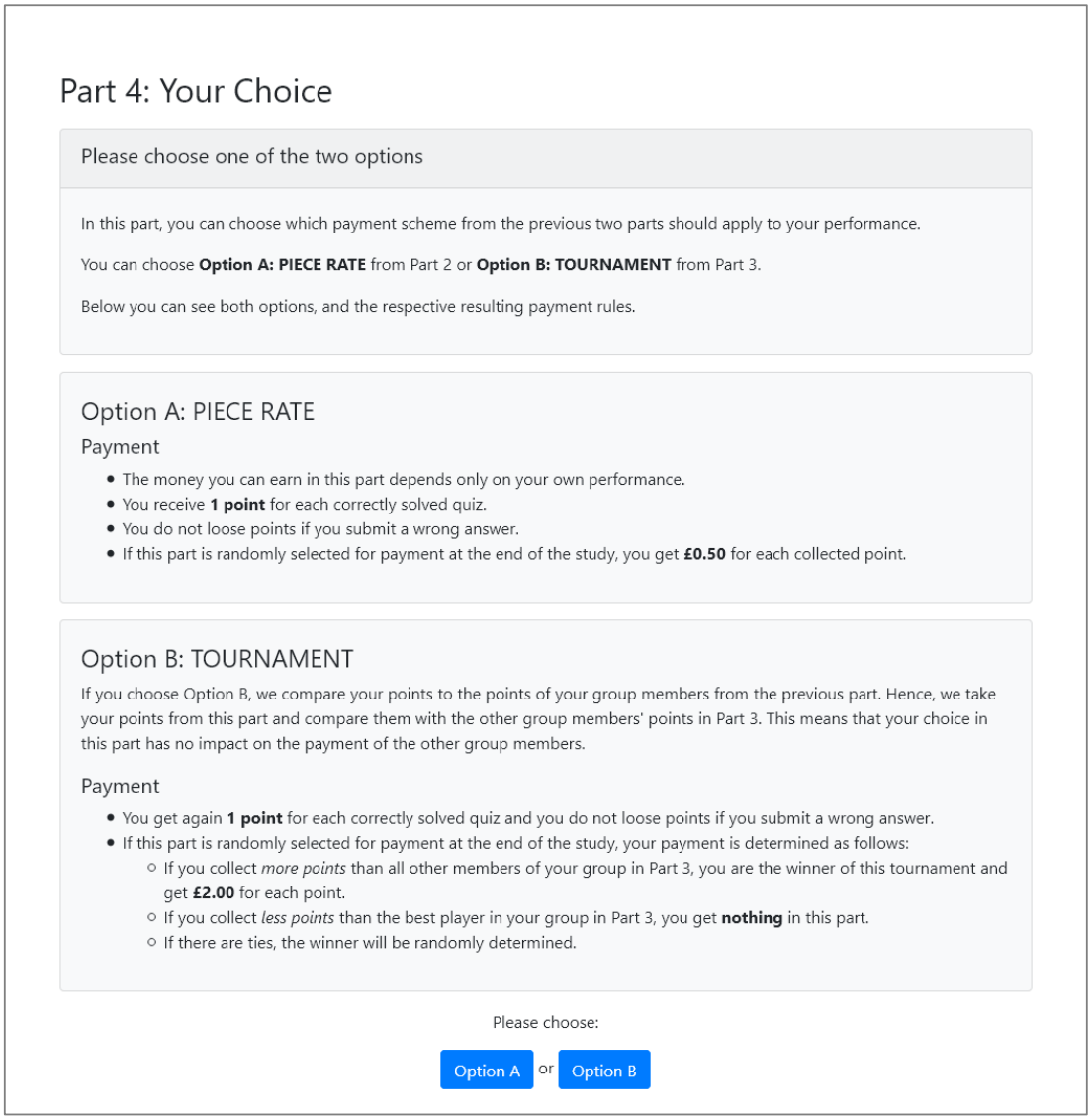

Part 2: Summary
Result
You solved 1 of 1 quizzes correctly.
Continue to next part


Part 3: Payment Details
Payment TOURNAMENT
For this part, you are now randomly matched with three other participants of this study. You are one group. You will not get to know who the other participants are and the other participants will not get to know who you are.
The money you can earn in this part depends on your performance and the performance of the other group members. Thus, you are in a tournament with the other group members.
You get again 1 point for each correctly solved quiz and you do not loose points if you submit a wrong answer.
If this part is randomly selected for payment at the end of the study, your payment is determined as follows:
If you collect more points than all other members of your group, you are the winner of this tournament and get £2.00 for each point.
If you collect less points than the best player in your group, you get nothing in this part.
If there are ties, the winner will be randomly determined.
If you are ready, click on the button 'Start this part' below.
You have 120 seconds and your submissions are relevant for your payment.
Start this part

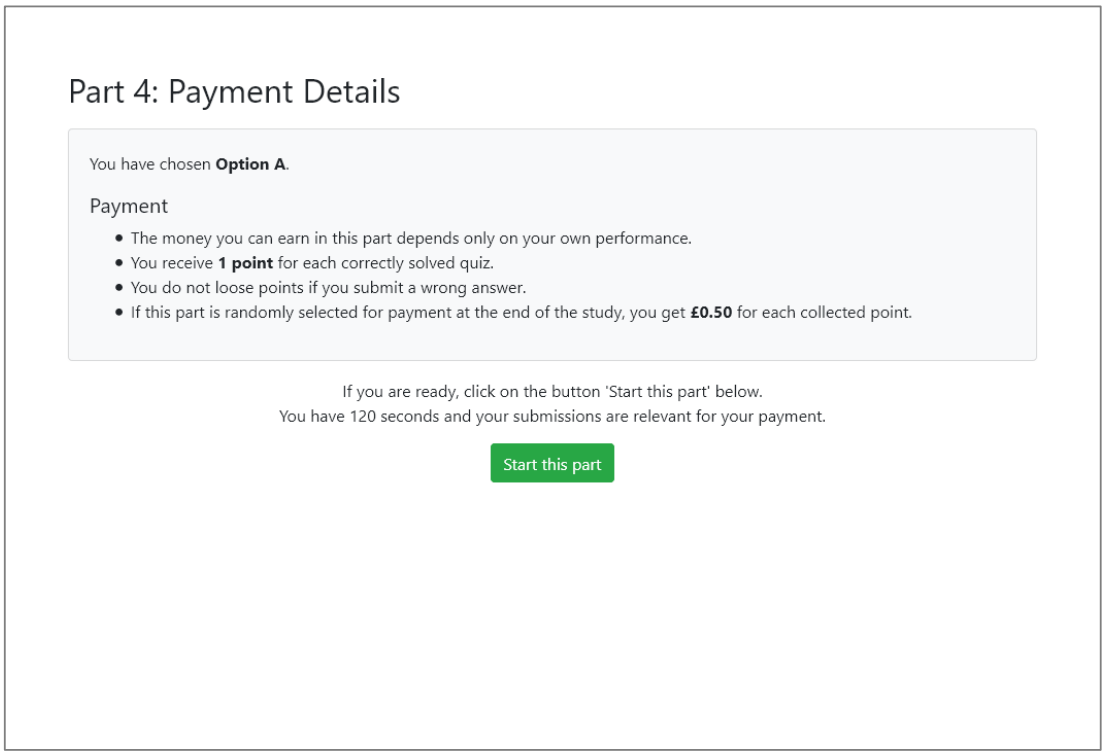

Part 3: Math Quizzes
Time left to solve quizzes: 1:40
Quiz 3
You solved 2 of 2 correctly.
Please select 2 numbers that add up to 100.
14
20
72
24
39
81
98
82
19
Submit


Part 3: Bonus Question
Please answer the following bonus question
You now have the chance to earn additional money. If this part is randomly chosen for payment, you will be paid an extra £1.00 on top of your final payoff if your following answer is correct.
How well did you perform in this part relative to your group members?
I was the best.
I was the second best.
I was the third best.
I was last.
Continue to Summary

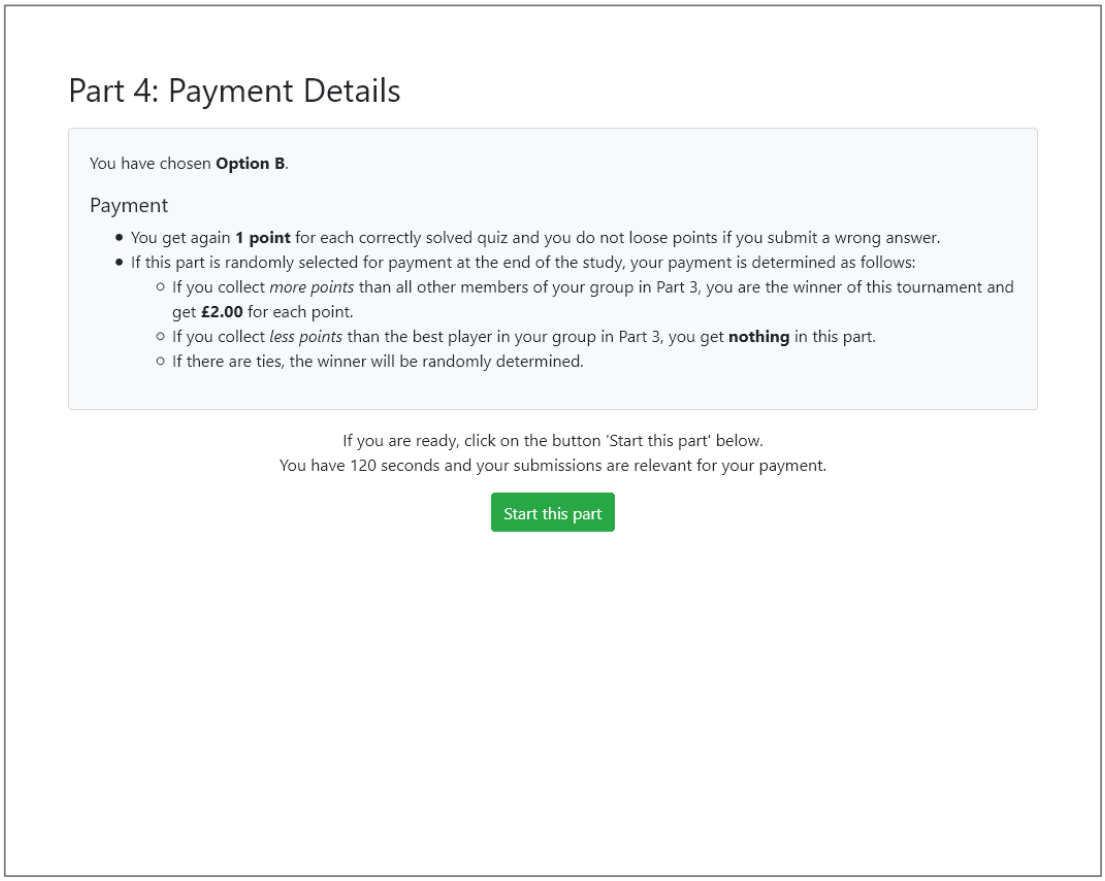

Part 3: Summary
Result
You solved 2 of 2 quizzes correctly.
Continue to next part

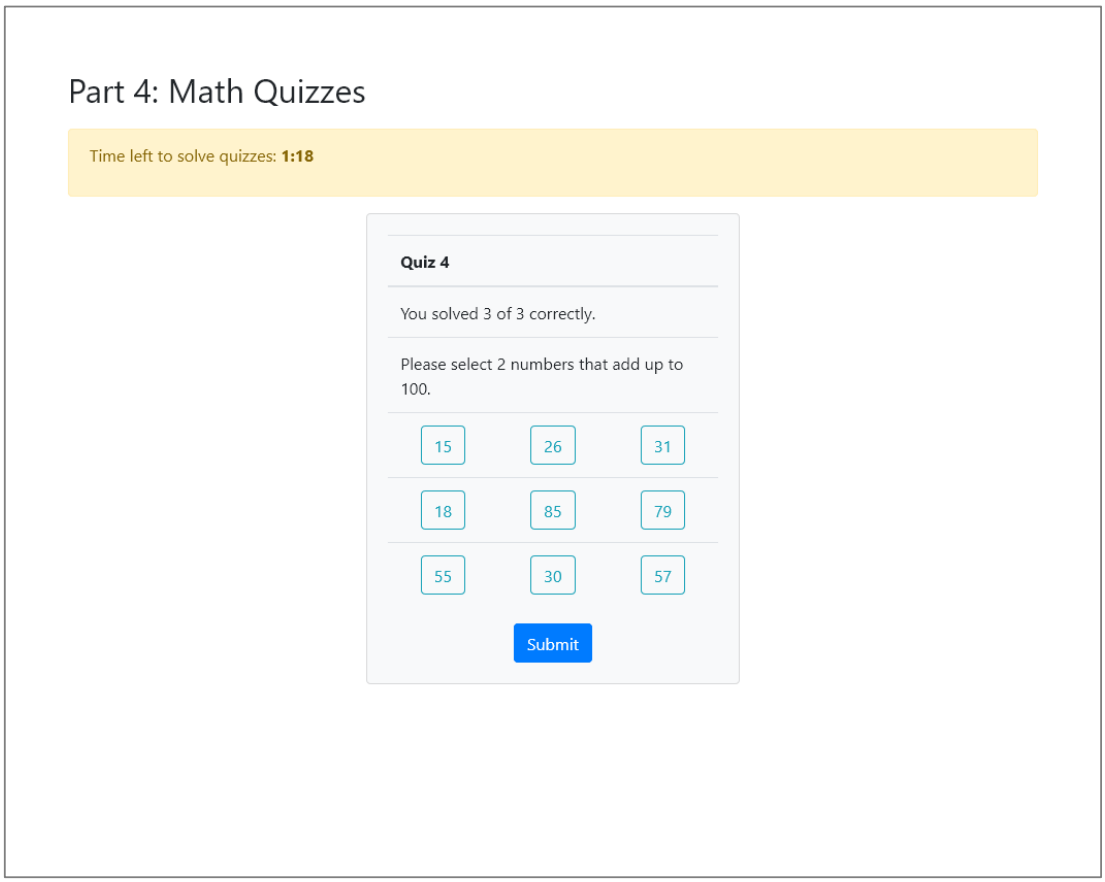

Part 4: Your Choice
Please choose one of the two options
In this part, you can choose which payment scheme from the previous two parts should apply to your performance.
You can choose Option A: PIECE RATE from Part 2 or Option B: TOURNAMENT from Part 3.
Below you can see both options, and the respective resulting payment rules.
Option A: PIECE RATE
Payment
The money you can earn in this part depends only on your own performance.
You receive 1 point for each correctly solved quiz.
You do not loose points if you submit a wrong answer.
If this part is randomly selected for payment at the end of the study, you get £0.50 for each collected point.
Option B: TOURNAMENT
If you choose Option B, we compare your points to the points of your group members from the previous part. Hence, we take your points from this part and compare them with the other group members' points in Part 3. This means that your choice in this part has no impact on the payment of the other group members.
Payment
You get again 1 point for each correctly solved quiz and you do not loose points if you submit a wrong answer.
If this part is randomly selected for payment at the end of the study, your payment is determined as follows:
If you collect more points than all other members of your group in Part 3, you are the winner of this tournament and get £2.00 for each point.
If you collect less points than the best player in your group in Part 3, you get nothing in this part.
If there are ties, the winner will be randomly determined.
Please choose:
Option A
or
Option B

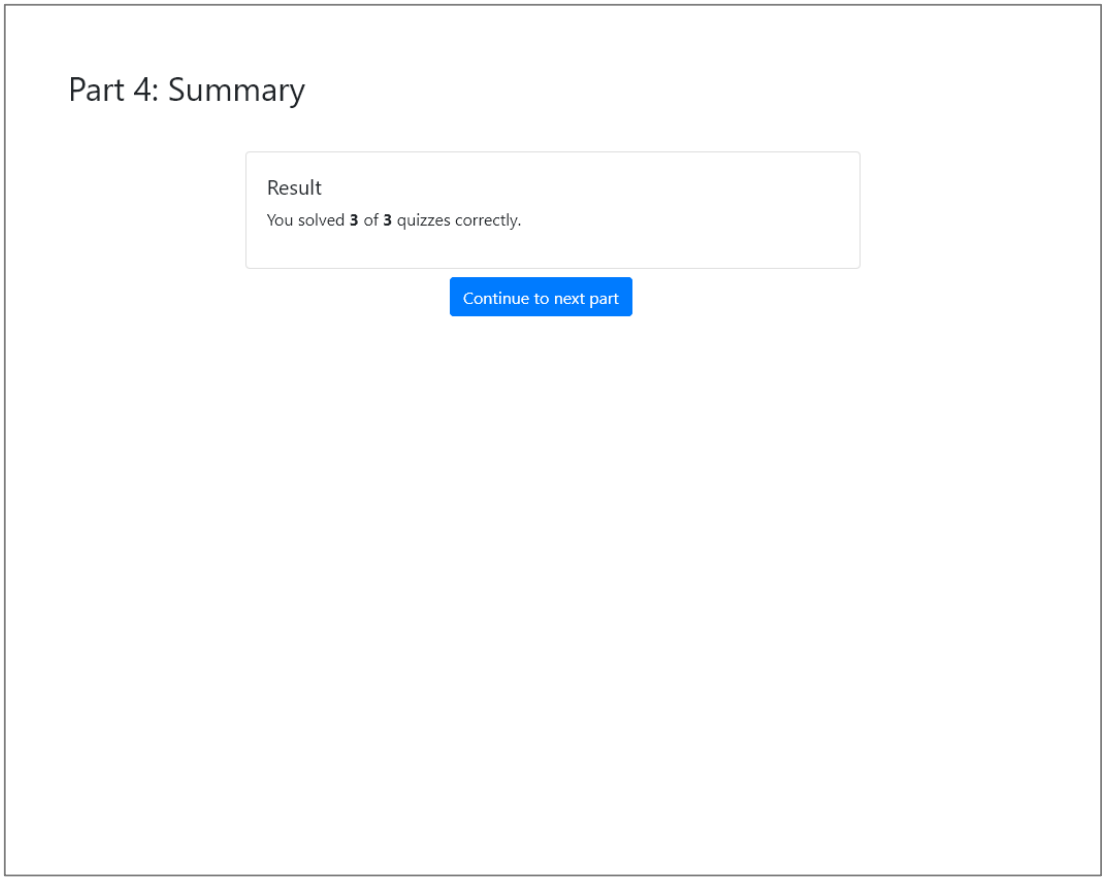
Choice: Option A



Choice: Option B




Part 4: Math Quizzes
Time left to solve quizzes: 1:18
Quiz 4
You solved 3 of 3 correctly.
Please select 2 numbers that add up to 100.
15
26
31
18
85
79
55
30
57
Submit


Part 4: Summary
Result
You solved 3 of 3 quizzes correctly.
Continue to next part

Part 5: Instructions

In this part, you get a start capital of **£4.00** and you have to decide how much of it you want to invest into a risky lottery.

The success of this investment is decided by *flipping a virtual coin*. You can get 'heads' or 'tails' with an equal probability. That means you have a 50% chance of success.

If you get **'tails'**, your investment was successful. In this case, the amount you have invested is multiplied by **2.5**. Your earnings are then the leftover of your start capital and the money from the successful investment.

But if you get **'heads'**, your investment was not successful, and you have only the leftover of your start capital.

There are no restrictions: you can put everything of your start capital in the investment, or nothing, or anything in between.

On the next page, you can decide how much to invest and see the potential outcomes of your investment.

Continue

Part 5: Your Decision

Your Decision

How much of your £4.00 do you want to invest?

4 £

If you have set the amount you want to invest, click on the button "Invest" below.

Under "Potential Outcome" to the right, you can see what can happen when you click "Invest".

Potential Outcome

You want to invest: £4.00

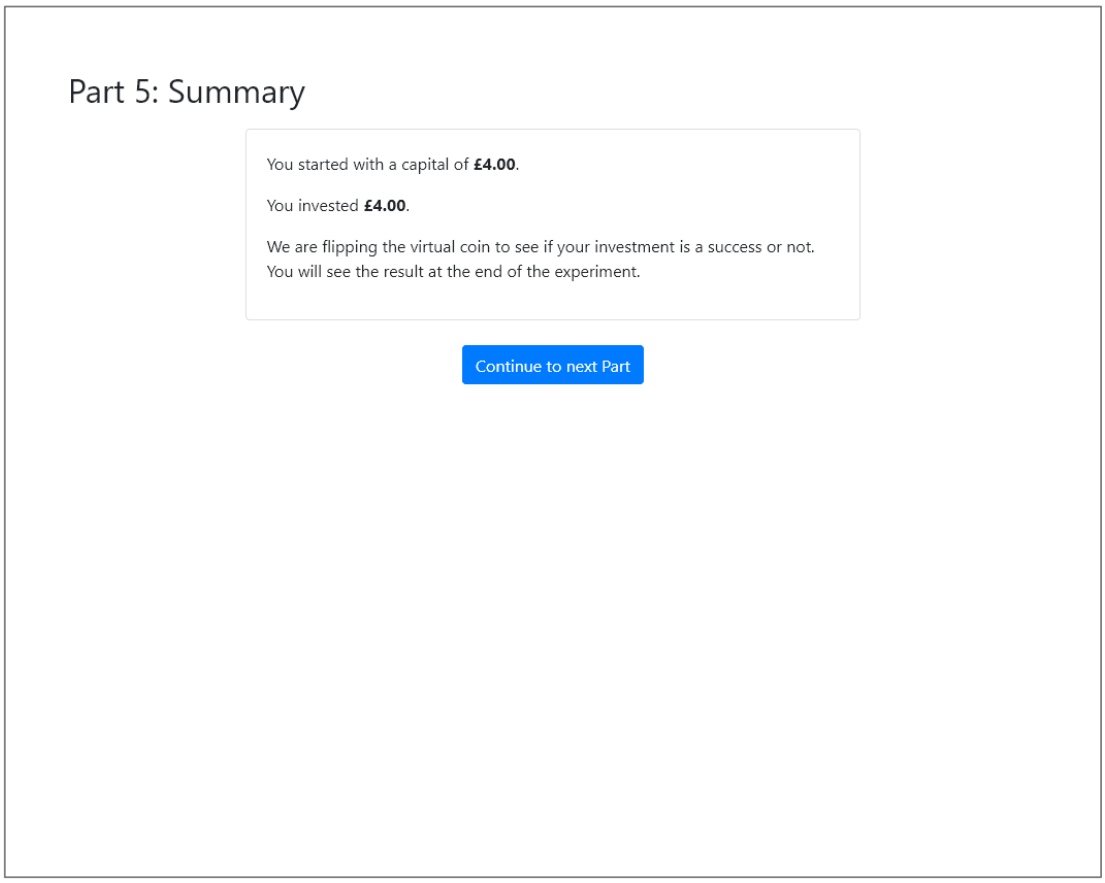
You want to keep: £0.00

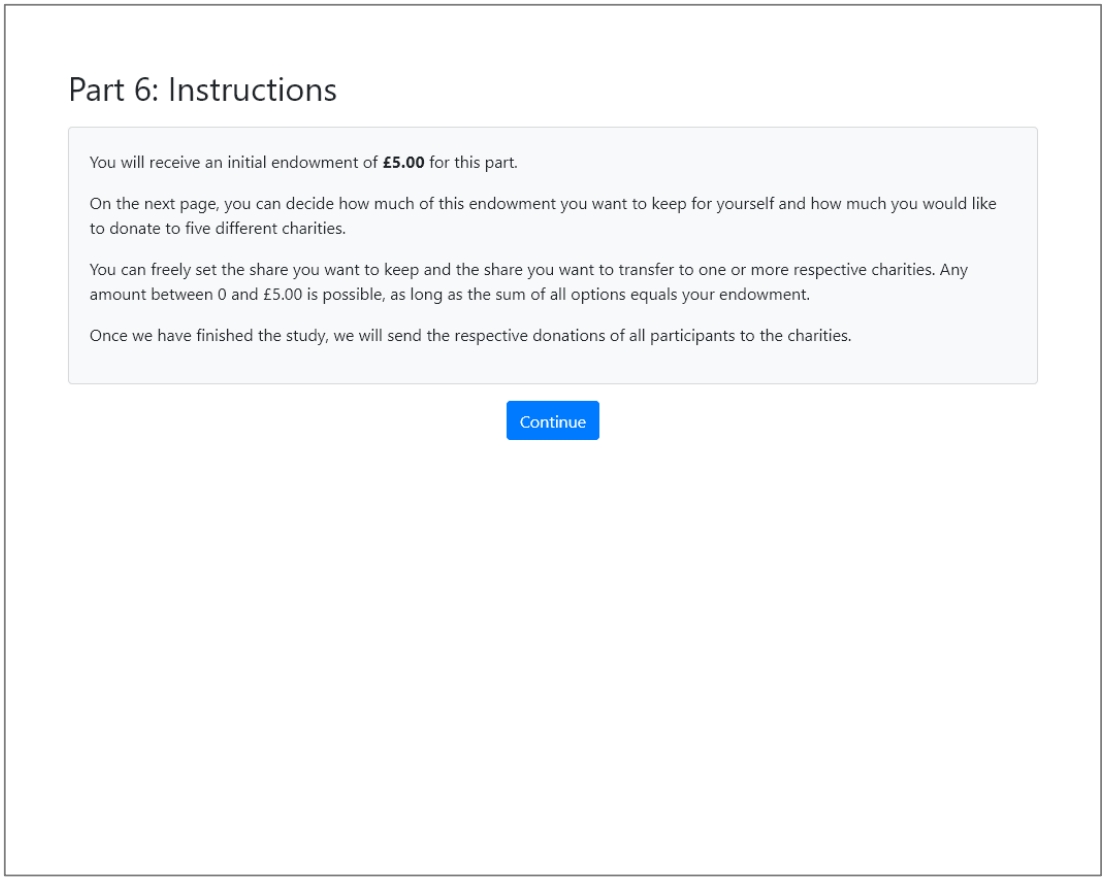
If the coin shows 'tails', you additionally get £10.00.
So in total you get: £0.00 + £10.00 = **£10.00**

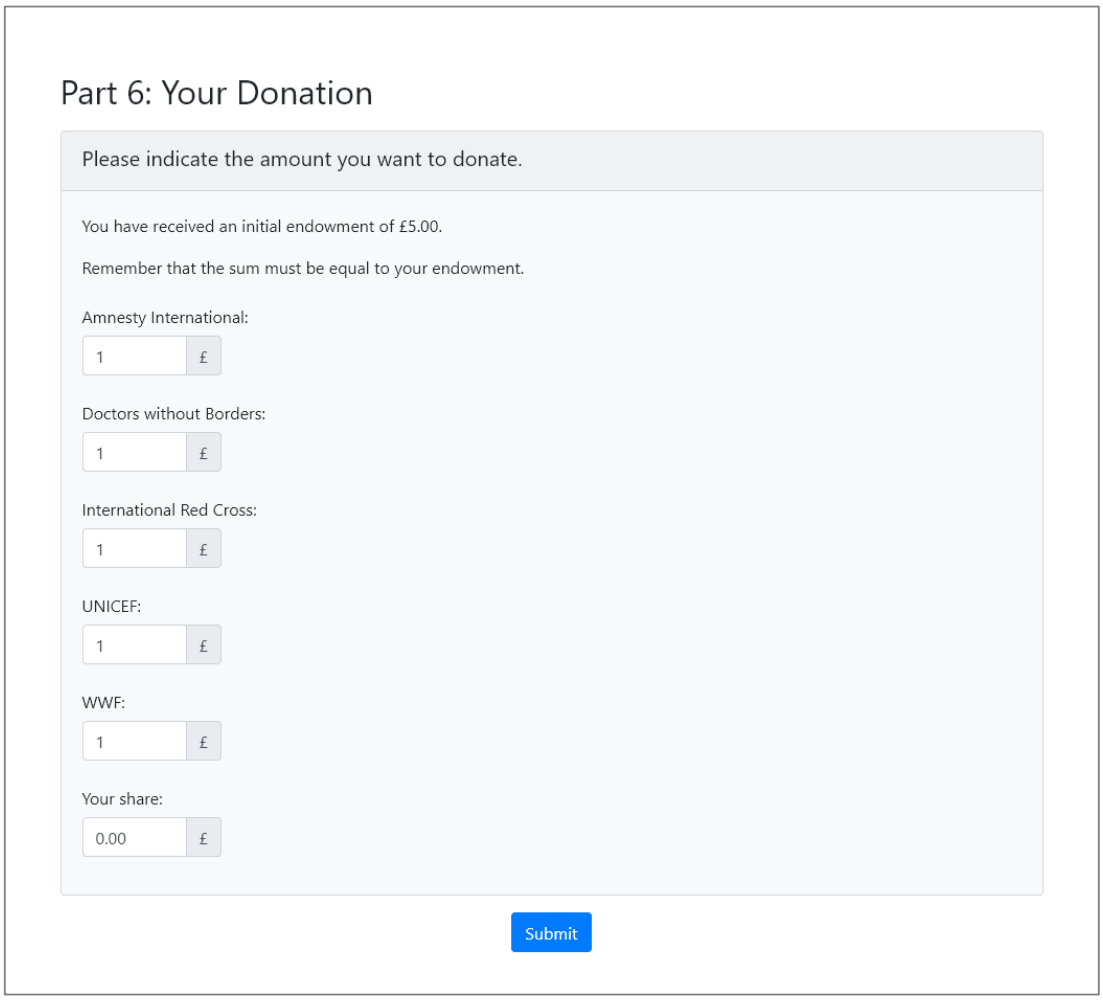
If the coin shows 'heads', you get £0.
So in total you get: £0.00 + £0 = **£0.00**

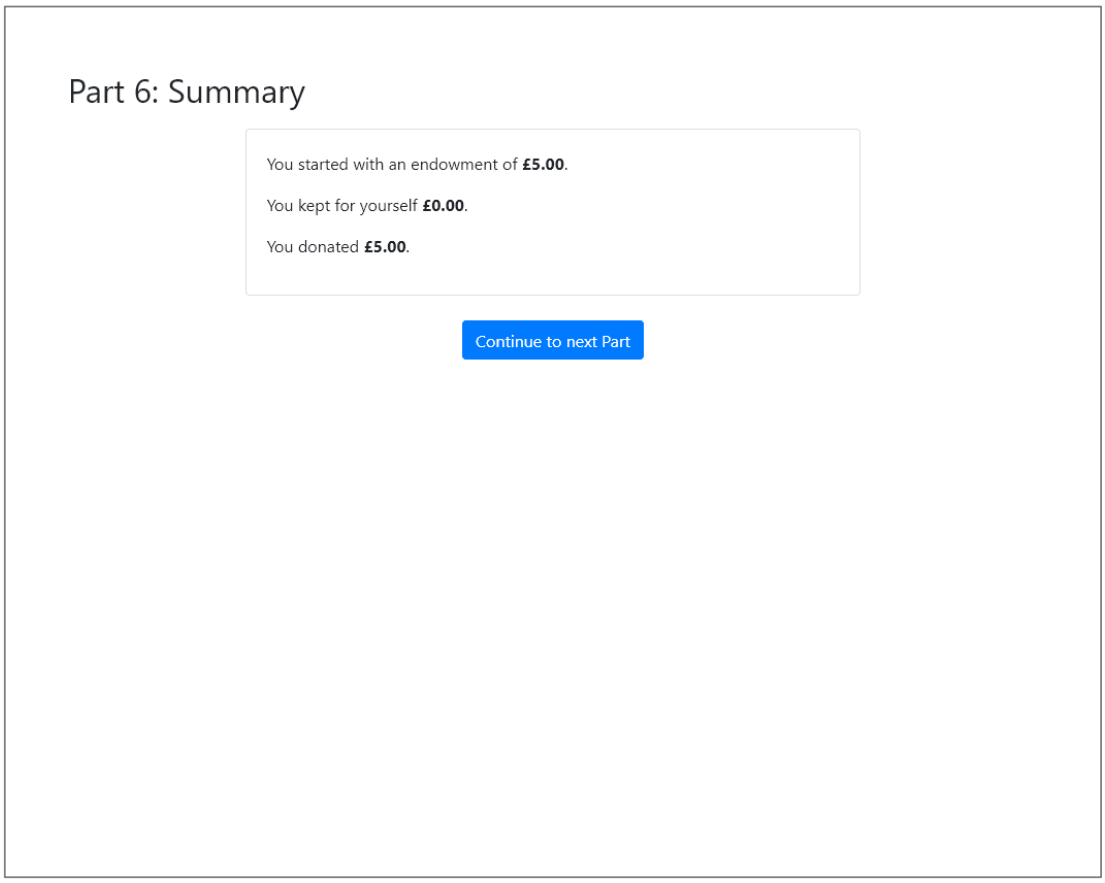
If you have made your decision, please click on the button 'Invest' below.

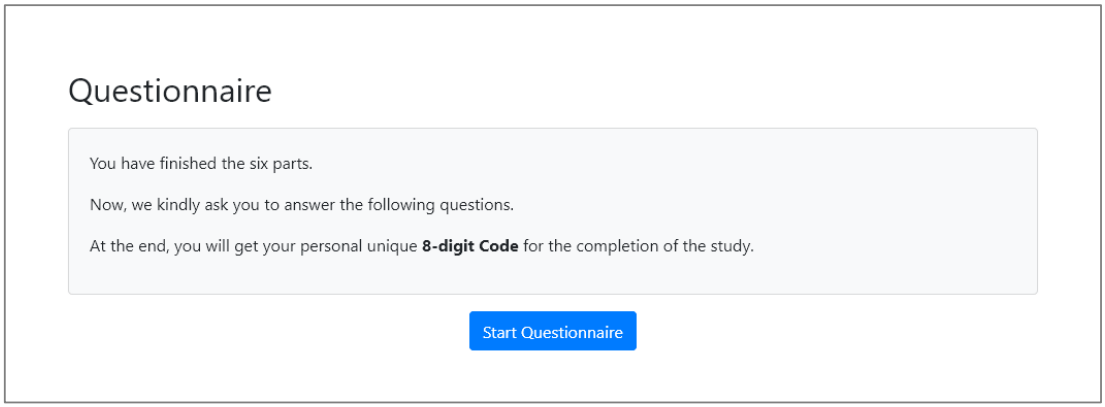
Invest


Part 5: Summary
You started with a capital of £4.00.
You invested £4.00.
We are flipping the virtual coin to see if your investment is a success or not.
You will see the result at the end of the experiment.
Continue to next Part


Part 6: Instructions
You will receive an initial endowment of £5.00 for this part.
On the next page, you can decide how much of this endowment you want to keep for yourself and how much you would like to donate to five different charities.
You can freely set the share you want to keep and the share you want to transfer to one or more respective charities. Any amount between 0 and £5.00 is possible, as long as the sum of all options equals your endowment.
Once we have finished the study, we will send the respective donations of all participants to the charities.
Continue


Part 6: Your Donation
Please indicate the amount you want to donate.
You have received an initial endowment of £5.00.
Remember that the sum must be equal to your endowment.
Amnesty International:
1
£
Doctors without Borders:
1
£
International Red Cross:
1
£
UNICEF:
1
£
WWF:
1
£
Your share:
0.00
£
Submit


Part 6: Summary
You started with an endowment of £5.00.
You kept for yourself £0.00.
You donated £5.00.
Continue to next Part


Questionnaire
You have finished the six parts.
Now, we kindly ask you to answer the following questions.
At the end, you will get your personal unique 8-digit Code for the completion of the study.
Start Questionnaire

Questionnaire

Below you find several personality traits that may or may not apply to you.

Please indicate for each trait the extent to which it applies to you on a scale from 1 (*Never true*) to 6 (*Always true*).

Trait	*Never true*	*Sometimes true*	*Occasionally true*	*Often true*	*Usually true*	*Always true*
Defends own beliefs	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Tender	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Conscientious	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Independent	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Sympathetic	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Moody	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Assertive	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Sensitive to needs of other	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Reliable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Strong personality	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Understanding	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Jealous	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Forceful	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Compassionate	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Truthful	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

Continue

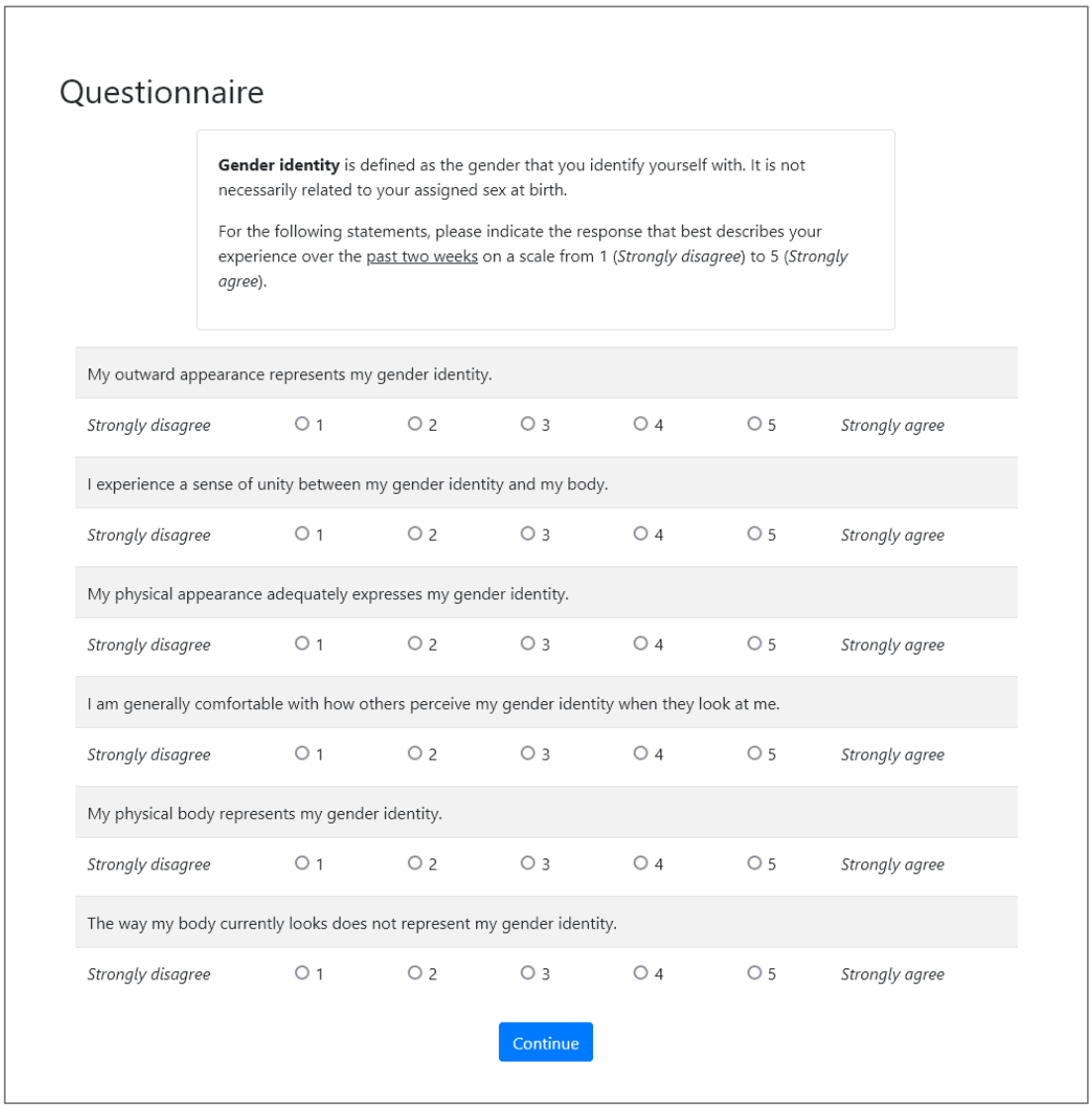
Questionnaire

Below you find several personality traits that may or may not apply to you.

Please indicate for each trait the extent to which it applies to you on a scale from 1 (*Never true*) to 6 (*Always true*).

Trait	*Never true*	*Sometimes true*	*Occasionally true*	*Often true*	*Usually true*	*Always true*
Has leader abilities	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Eager to soothe hurt feelings	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Secretive	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Willing to take risk	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Warm	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Adaptable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Dominant	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Affectionate	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Conceited	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Willing to take a stand	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Loves children	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Tactful	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Aggressive	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Gentle	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
Conventional	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

Continue



Questionnaire

Gender identity is defined as the gender that you identify yourself with. It is not necessarily related to your assigned sex at birth.

For the following statements, please indicate the response that best describes your experience over the past two weeks on a scale from 1 (*Strongly disagree*) to 5 (*Strongly agree*).

My outward appearance represents my gender identity.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

I experience a sense of unity between my gender identity and my body.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

My physical appearance adequately expresses my gender identity.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

I am generally comfortable with how others perceive my gender identity when they look at me.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

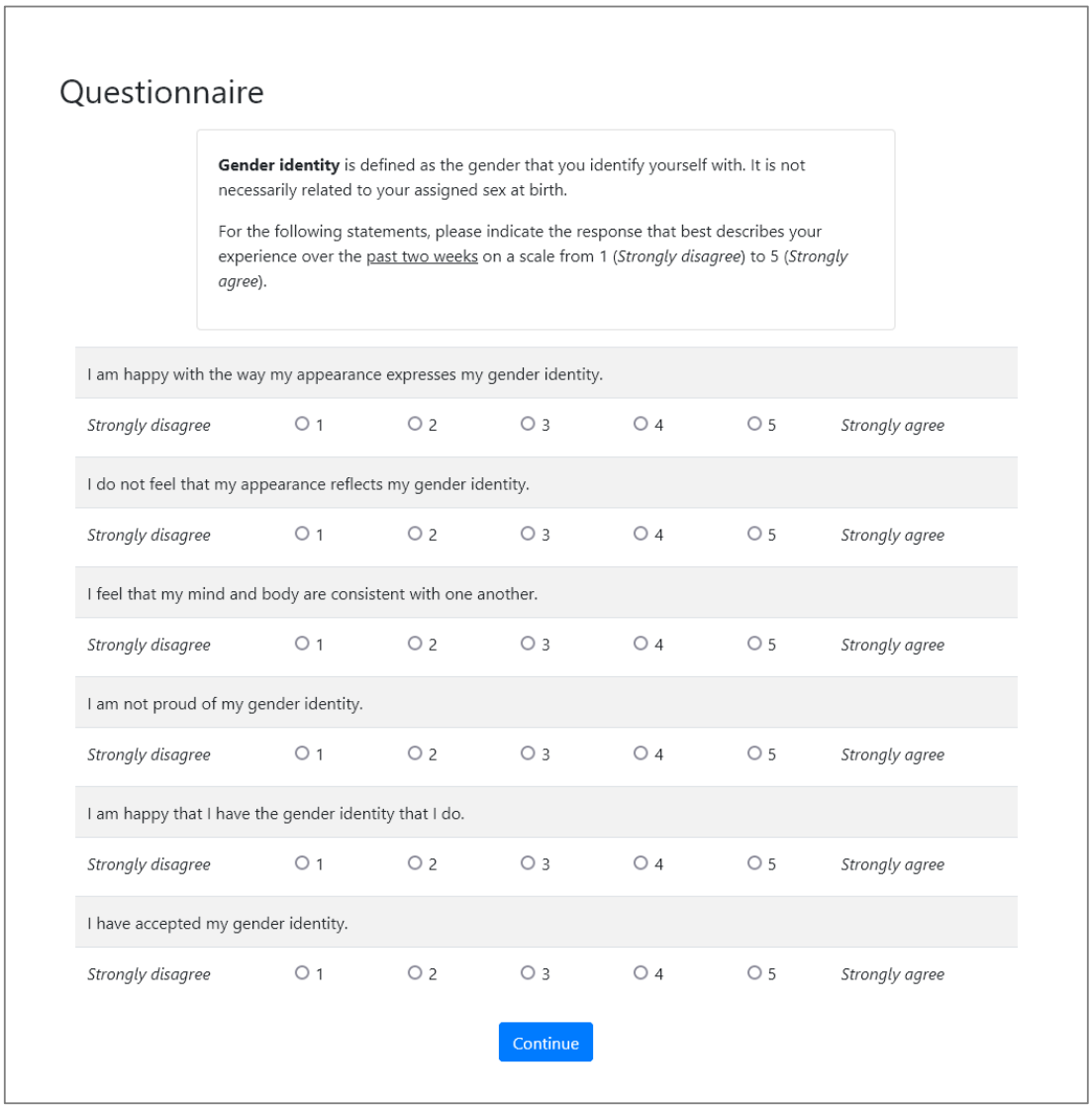
My physical body represents my gender identity.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

The way my body currently looks does not represent my gender identity.

Strongly disagree ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 *Strongly agree*

Continue



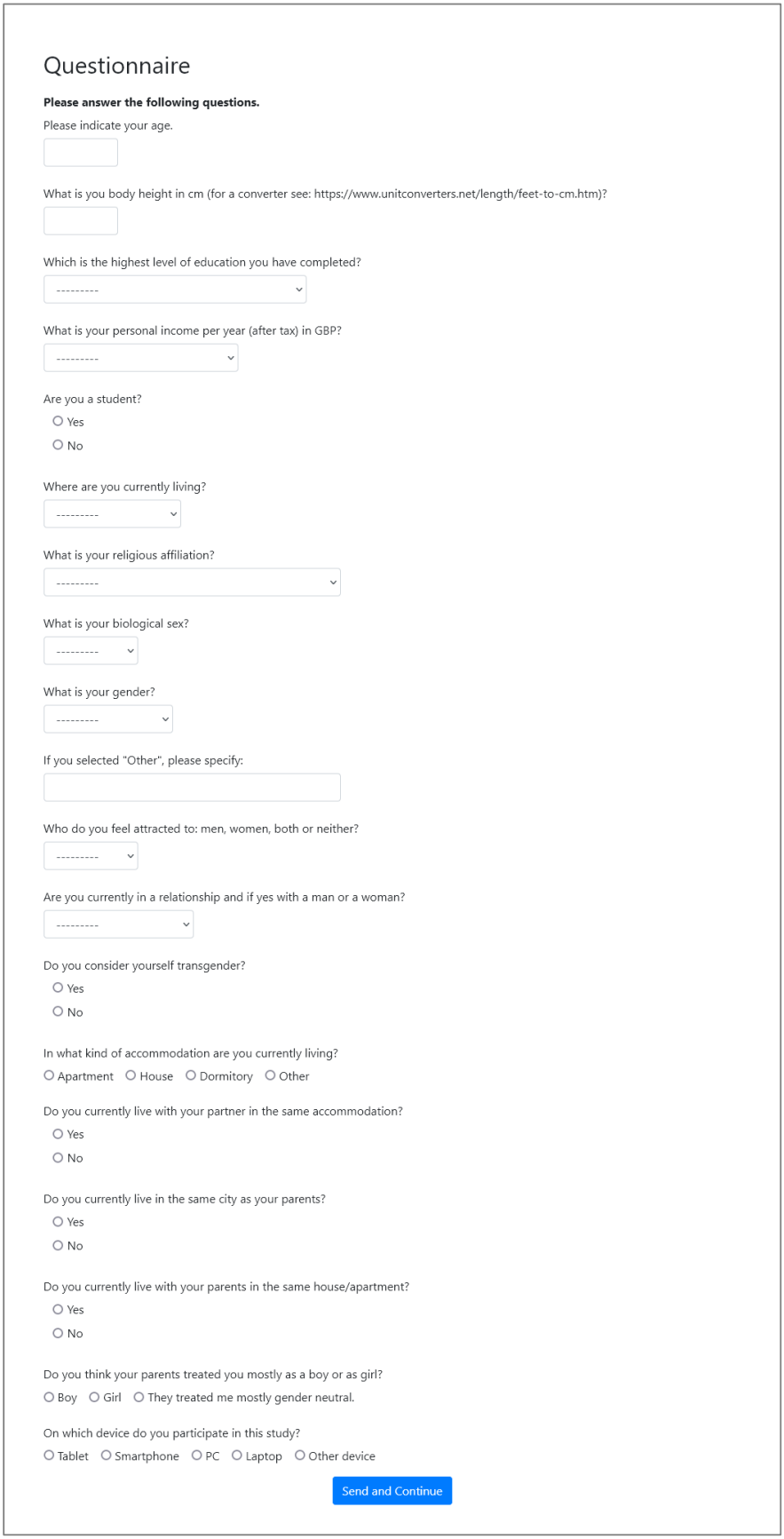
Questionnaire

Gender identity is defined as the gender that you identify yourself with. It is not necessarily related to your assigned sex at birth.

For the following statements, please indicate the response that best describes your experience over the past two weeks on a scale from 1 (*Strongly disagree*) to 5 (*Strongly agree*).

I am happy with the way my appearance expresses my gender identity.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*
I do not feel that my appearance reflects my gender identity.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*
I feel that my mind and body are consistent with one another.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*
I am not proud of my gender identity.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*
I am happy that I have the gender identity that I do.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*
I have accepted my gender identity.						
Strongly disagree	○ 1	○ 2	○ 3	○ 4	○ 5	*Strongly agree*

Continue



Questionnaire

Please answer the following questions.

Please indicate your age.

What is you body height in cm (for a converter see: https://www.unitconverters.net/length/feet-to-cm.htm)?

Which is the highest level of education you have completed?

What is your personal income per year (after tax) in GBP?

Are you a student?

- Yes
- No

Where are you currently living?

What is your religious affiliation?

What is your biological sex?

What is your gender?

If you selected "Other", please specify:

Who do you feel attracted to: men, women, both or neither?

Are you currently in a relationship and if yes with a man or a woman?

Do you consider yourself transgender?

- Yes
- No

In what kind of accommodation are you currently living?

- Apartment
- House
- Dormitory
- Other

Do you currently live with your partner in the same accommodation?

- Yes
- No

Do you currently live in the same city as your parents?

- Yes
- No

Do you currently live with your parents in the same house/apartment?

- Yes
- No

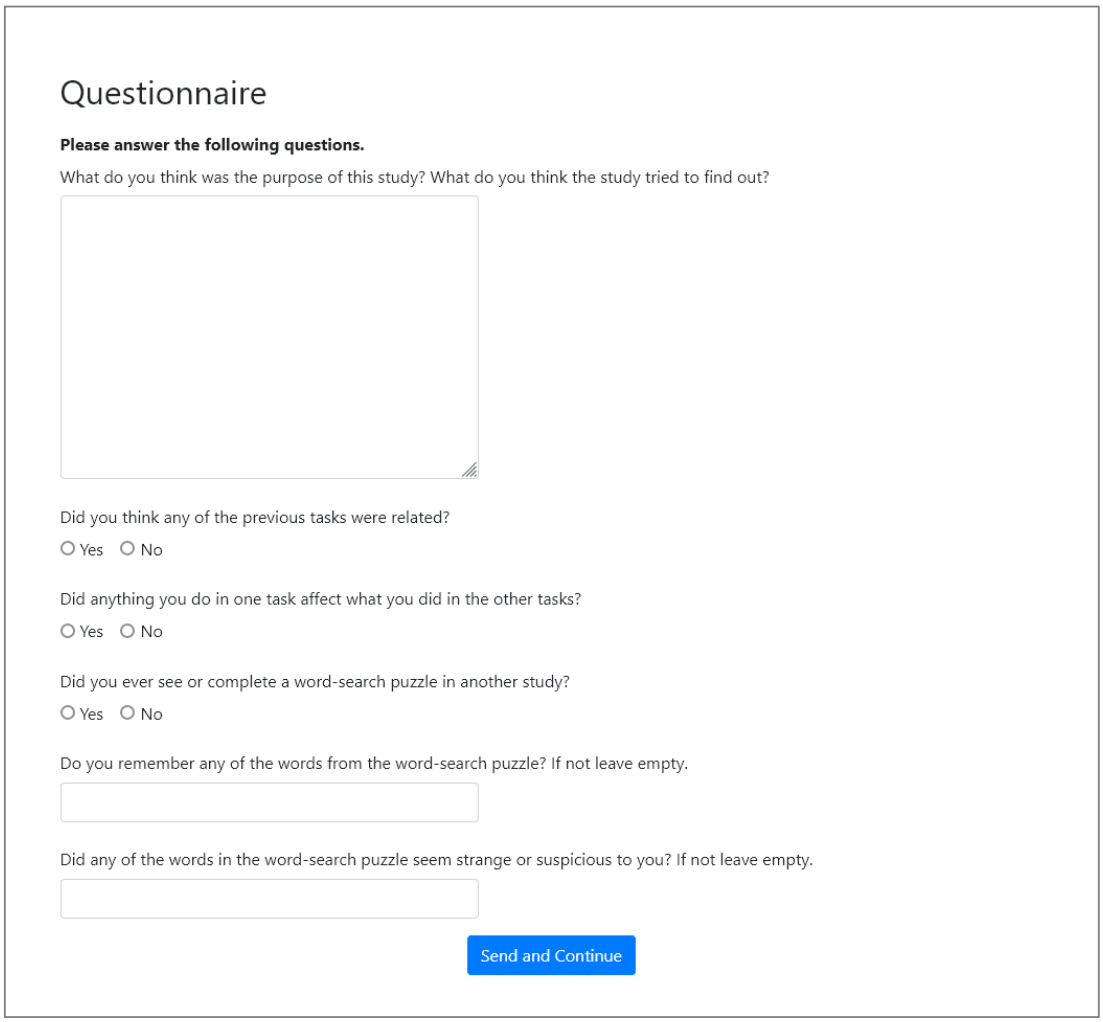
Do you think your parents treated you mostly as a boy or as girl?

- Boy
- Girl
- They treated me mostly gender neutral.

On which device do you participate in this study?

- Tablet
- Smartphone
- PC
- Laptop
- Other device

Send and Continue


Questionnaire
Please answer the following questions.
What do you think was the purpose of this study? What do you think the study tried to find out?
Did you think any of the previous tasks were related?
Yes
No
Did anything you do in one task affect what you did in the other tasks?
Yes
No
Did you ever see or complete a word-search puzzle in another study?
Yes
No
Do you remember any of the words from the word-search puzzle? If not leave empty.
Did any of the words in the word-search puzzle seem strange or suspicious to you? If not leave empty.
Send and Continue

Bibliography

Ackerman, F. and Heinzerling, L. (2002), 'Pricing the priceless: Cost-benefit analysis of environmental protection', *University of Pennsylvania Law Review* **150**(5), 1553–1584.

Adams, R. B. and Funk, P. (2012), 'Beyond the glass ceiling: Does gender matter?', *Management Science* **58**(2), 219–235.

Ahmed, A. M. and Salas, O. (2011), 'Implicit influences of christian religious representations on dictator and prisoner's dilemma game decisions', *The Journal of Socio-Economics* **40**(3), 242–246.

Akerlof, G. A. and Kranton, R. E. (2000), 'Economics and identity', *The Quarterly Journal of Economics* **115**(3), 715–753.

Akerlof, G. A. and Kranton, R. E. (2005), 'Identity and the economics of organizations', *Journal of Economic Perspectives* **19**(1), 9–32.

Alempaki, D., Starmer, C. and Tufano, F. (2019), 'On the priming of risk preferences: The role of fear and general affect', *Journal of Economic Psychology* **75**, 102137.

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics* **95**(9-10), 1082–1095.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø. and Tungodden, B. (2016*a*), 'What explains the gender gap in college track dropout? Experimental and administrative evidence', *American Economic Review* **106**(5), 296–302.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø. and Tungodden, B. (2016*b*), 'Willingness to compete: Family matters', *Management Science* **62**(8), 2149–2162.

Almenberg, J. and Dreber, A. (2015), 'Gender, stock market participation and financial literacy', *Economics Letters* **137**, 140–142.

Alt, M. and Gallier, C. (2022), 'Incentives and intertemporal behavioral spillovers: A two-period experiment on charitable giving', *Journal of Economic Behavior & Organization* **200**, 959–972.

Altemeyer-Bartscher, M., Rübbelke, D. T. and Sheshinski, E. (2010), 'Environmental protection and the private provision of international public goods', *Economica* **77**(308), 775–784.

Altonji, J. G. and Blank, R. M. (1999), 'Race and gender in the labor market', *Handbook of Labor Economics* **3**, 3143–3259.

American Psychological Association (2015), 'Guidelines for psychological practice with transgender and gender nonconforming people', *American Psychologist* **70**(9), 832–864.

Amienyo, D. and Azapagic, A. (2016), 'Life cycle environmental impacts and costs of beer production and consumption in the UK', *The International Journal of Life Cycle Assessment* **21**(4), 492–509.

Andersen, S., Ertac, S., Gneezy, U., List, J. A. and Maximiano, S. (2013), 'Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society', *Review of Economics and Statistics* **95**(4), 1438–1443.

Anderson, A., Dreber, A. and Vestman, R. (2015), 'Risk taking, behavioral biases and genes: Results from 149 active investors', *Journal of Behavioral and Experimental Finance* **6**, 93–100.

Anderson, L. R. and Mellor, J. M. (2008), 'Predicting health behaviors with an experimental measure of risk preference', *Journal of Health Economics* **27**(5), 1260–1274.

Andersson, J. J. (2019), 'Carbon taxes and CO_2 emissions: Sweden as a case study', *American Economic Journal: Economic Policy* **11**(4), 1–30.

Andreoni, J. and Vesterlund, L. (2001), 'Which is the fair sex? Gender differences in altruism', *The Quarterly Journal of Economics* **116**(1), 293–312.

Angner, E. and Loewenstein, G. (2007), 'Behavioral economics', *Handbook of the Philosophy of Science: Philosophy of Economic* pp. 641–690.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W. and Wilson, A. J. (2016), 'The slider task: An example of restricted inference on incentive effects', *Journal of the Economic Science Association* **2**, 1–12.

Asensio, O. I. and Delmas, M. A. (2015), 'Nonprice incentives and energy conservation', *Proceedings of the National Academy of Sciences* **112**(6), E510–E515.

Asensio, O. I. and Delmas, M. A. (2016), 'The dynamics of behavior change: Evidence from energy conservation', *Journal of Economic Behavior & Organization* **126**, 196–212.

Atkinson, A. B. and Stern, N. H. (1974), 'Pigou, taxation and public goods', *The Review of Economic Studies* **41**(1), 119–128.

Atkinson, G. (2000), 'Measuring corporate sustainability', *Journal of Environmental Planning and Management* **43**(2), 235–252.

Atkinson, G. and Mourato, S. (2008), 'Environmental Cost-Benefit Analysis', *Annual Review of Environment and Resources* **33**(1), 317–344.

Atkinson, S. M., Baird, S. B. and Frye, M. B. (2003), 'Do female mutual fund managers manage differently?', *Journal of Financial Research* **26**(1), 1–18.

Auger, P. and Devinney, T. M. (2007), 'Do what consumers way matter? The misalignment of preferences with unconstrained ethical intentions', *Journal of Business Ethics* **76**(4), 361–383.

Azrieli, Y., Chambers, C. P. and Healy, P. J. (2018), 'Incentives in experiments: A theoretical analysis', *Journal of Political Economy* **126**(4), 1472–1503.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A. and Nelson, L. D. (2013), 'Commitment and behavior change: Evidence from the field', *Journal of Consumer Research* **39**(5), 1070–1084.

Balafoutas, L., Fornwagner, H. and Sutter, M. (2018), 'Closing the gender gap in competitiveness through priming', *Nature Communications* **9**(1), 1–6.

Balafoutas, L. and Sutter, M. (2012), 'Affirmative action policies promote women and do not harm efficiency in the laboratory', *Science* **335**(6068), 579–582.

Balafoutas, L. and Sutter, M. (2019), 'How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior', *European Economic Review* **118**, 1–13.

Baldiga, N. R. and Coffman, K. B. (2018), 'Laboratory evidence on the effects of sponsorship on the competitive preferences of men and women', *Management Science* **64**(2), 888–901.

Bank of England (2021), 'Inflation calculator', Available Online: https://www.bankofengland.co.uk/monetary-policy/inflation/inflation-calculator. Accessed: March 2021.

Baram, M. S. (1979), 'Cost-benefit analysis: An inadequate basis for health, safety, and environmental regulatory decisionmaking', *Ecology LQ* **8**, 473.

Barby, C., Barker, R., Cohen, R., Eccles, R. G., Heller, C., Mayer, C., Roche, B., Serafeim, G., Stroehle, J. C., Younger, R. et al. (2021), 'Measuring purpose: An integrated framework', *Available at SSRN 3771892* .

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K. and Trötschel, R. (2001), 'The automated will: Nonconscious activation and pursuit of behavioral goals.', *Journal of Personality and Social Psychology* **81**(6), 1014.

Barreda-Tarazona, I., García-Gallego, A., Georgantzís, N., Jaramillo-Gutiérrez, A. and Sabater-Grande, G. (2011), 'An experiment on the motivation and the effects of socially responsible market behavior', *International Journal of Ecological Economics and Statistics* **23**, 105–128.

Barsky, R. B., Juster, F. T., Kimball, M. S. and Shapiro, M. D. (1997), 'Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study', *The Quarterly Journal of Economics* **112**(2), 537–579.

Bateson, M., Nettle, D. and Roberts, G. (2006), 'Cues of being watched enhance cooperation in a real-world setting', *Biology Letters* **2**(3), 412–414.

Bebbington, J., Brown, J. and Frame, B. (2007), 'Accounting technologies and sustainability assessment models', *Ecological Economics* **61**(2), 224 – 236.

Beblo, M. and Markowsky, E. (2022), 'When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature', *Journal of Economic Behavior and Organization* **198**, 139–163.

Bem, D. J. (1967), 'Self-perception: An alternative interpretation of cognitive dissonance phenomena.', *Psychological Review* **74**(3), 183.

Bénabou, R. and Tirole, J. (2011), 'Identity, morals, and taboos: Beliefs as assets', *The Quarterly Journal of Economics* **126**(2), 805–855.

Benjamin, D. J., Choi, J. J. and Fisher, G. (2016), 'Religious identity and economic behavior', *Review of Economics and Statistics* **98**(4), 617–637.

Benjamin, D. J., Choi, J. J. and Strickland, A. J. (2010), 'Social identity and preferences', *American Economic Review* **100**(4), 1913–28.

Benz, M. and Meier, S. (2008), 'Do people behave in experiments as in the field?—evidence from donations', *Experimental Economics* **11**(3), 268–281.

Berge, L. I. O., Bjorvatn, K., Pires, A. J. G. and Tungodden, B. (2015), 'Competitive in the lab, successful in the field?', *Journal of Economic Behavior & Organization* **118**, 303–317.

Big Room Inc. (2021), 'Ecolabel index', Available Online: http://www.ecolabelindex.com/. Accessed: August 2021.

Bilén, D., Dreber, A. and Johannesson, M. (2021), 'Are women more generous than men? A meta-analysis', *Journal of the Economic Science Association* **7**(1), 1–18.

Bittner, A. and Goodyear-Grant, E. (2017), 'Sex isn't gender: Reforming concepts and measurements in the study of public opinion', *Political Behavior* **39**(4), 1019–1041.

Blake, J. (1999), 'Overcoming the 'value-action gap' in environmental policy: Tensions between national policy and local experience', *Local Environment* **4**(3), 257–278.

Blanken, I., van de Ven, N. and Zeelenberg, M. (2015), 'A meta-analytic review of moral licensing', *Personality and Social Psychology Bulletin* **41**(4), 540–558.

Blanken, I., van de Ven, N., Zeelenberg, M. and Meijers, M. H. (2014), 'Three attempts to replicate the moral licensing effect', *Social Psychology* .

Boardman, A. E., Greenberg, D. H., Vining, A. R. and Weimer, D. L. (2017), *Cost-benefit analysis: concepts and practice*, Cambridge University Press.

Boksem, M. A., Mehta, P. H., van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., Smidts, A. and Sanfey, A. G. (2013), 'Testosterone inhibits trust but promotes reciprocity', *Psychological Science* **24**(11), 2306–2314.

Bolderdijk, J. W., Steg, L., Geller, E. S., Lehman, P. K. and Postmes, T. (2013), 'Comparing the effectiveness of monetary versus moral motives in environmental campaigning', *Nature Climate Change* **3**(4), 413–416.

Bonin, H., Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2007), 'Cross-sectional earnings risk and occupational sorting: The role of risk attitudes', *Labour Economics* **14**(6), 926–937.

Bonnefon, J.-F., Landier, A., Sastry, P. R. and Thesmar, D. (2022), The moral preferences of investors: Experimental evidence, Working paper, National Bureau of Economic Research.

Booth, A. and Nolen, P. (2012), 'Choosing to compete: How different are girls and boys?', *Journal of Economic Behavior & Organization* **81**(2), 542–555.

Boschini, A., Dreber, A., von Essen, E., Muren, A. and Ranehill, E. (2018), 'Gender and altruism in a random sample', *Journal of Behavioral and Experimental Economics* **77**, 72–77.

Brañas-Garza, P., Bucheli, M., Espinosa, M. P. and García-Muñoz, T. (2013), 'Moral cleansing and moral licenses: Experimental evidence', *Economics & Philosophy* **29**(2), 199–212.

Brandts, J., Groenert, V. and Rott, C. (2015), 'The impact of advice on women's and men's selection into competition', *Management Science* **61**(5), 1018–1035.

Brenøe, A. A., Heursen, L., Ranehill, E. and Weber, R. A. (2022), 'Continuous gender identity and economics', *AEA Papers and Proceedings* **112**, 573–77.

Brodback, D., Guenster, N. and Pouget, S. (2020), 'The valuation of corporate social responsibility: A willingness-to-pay experiment', Available Online: http://uleef.business.utah.edu/newsite/pubs/Pouget.pdf.

Bucciol, A. and Miniaci, R. (2011), 'Household portfolios and implicit risk preference', *Review of Economics and Statistics* **93**(4), 1235–1250.

Burger, A. M., Schuler, J. and Eberling, E. (2022), 'Guilty pleasures: Moral licensing in climate-related behavior', *Global Environmental Change* **72**, 102415.

Buser, T., Geijtenbeek, L. and Plug, E. (2018), 'Sexual orientation, competitiveness and income', *Journal of Economic Behavior & Organization* **151**, 191–198.

Buser, T., Niederle, M. and Oosterbeek, H. (2014), 'Gender, competitiveness, and career choices', *The Quarterly Journal of Economics* **129**(3), 1409–1447.

Buser, T., Niederle, M. and Oosterbeek, H. (2021), Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures, Working Paper 28916, National Bureau of Economic Research.

Buser, T., Peter, N. and Wolter, S. C. (2017), 'Gender, willingness to compete and career choices along the whole ability distribution'.

Buskens, V., Raub, W., van Miltenburg, N., Montoya, E. R. and van Honk, J. (2016), 'Testosterone administration moderates effect of social environment on trust in women depending on second-to-fourth digit ratio', *Scientific Reports* **6**(1), 1–8.

Butera, L. and List, J. A. (2017), An economic approach to alleviate the crises of confidence in science: With an application to the public goods game, Technical report, National Bureau of Economic Research.

Cadsby, C. B., Servátka, M. and Song, F. (2013), 'How competitive are female professionals? A tale of identity conflict', *Journal of Economic Behavior & Organization* **92**, 284–303.

Callen, M., Isaqzadeh, M., Long, J. D. and Sprenger, C. (2014), 'Violence and risk preference: Experimental evidence from afghanistan', *American Economic Review* **104**(1), 123–48.

Cambridge Dictionary (n.d.), eco-label, *in* 'https://dictionary.cambridge.org/dictionary/english/', Cambridge University Press. Accessed: November 2021.

Camerer, C. (1999), 'Behavioral economics: Reunifying psychology and economics', *Proceedings of the National Academy of Sciences* **96**(19), 10575–10577.

Camerer, C. F., Loewenstein, G. and Rabin, M. (2004), *Advances in behavioral economics*, Princeton university press.

Camilleri, A. R., Larrick, R. P., Hossain, S. and Patino-Echeverri, D. (2019), 'Consumers underestimate the emissions associated with food but are aided by labels', *Nature Climate Change* **9**(1), 53–58.

Capstick, S., Whitmarsh, L., Nash, N., Haggar, P. and Lord, J. (2019), 'Compensatory and catalyzing beliefs: Their relationship to pro-environmental behavior and behavioral spillover in seven countries', *Frontiers in Psychology* **10**, 963.

Carbon Trust (2019), 'Research reveals consumer demand for climate change labelling', Available Online: https://www.carbontrust.com/de/node/51. Accessed: September 2021.

Card, D., Cardoso, A. R. and Kline, P. (2016), 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women', *The Quarterly Journal of Economics* **131**(2), 633–686.

Cárdenas, J. C., Dreber, A., von Essen, E. and Ranehill, E. (2015), 'Cooperativeness and competitiveness in children', *Journal of Behavioral and Experimental Economics* **59**, 32–41.

Carlsson, F., Jaime, M. and Villegas, C. (2021), 'Behavioral spillover effects from a social information campaign', *Journal of Environmental Economics and Management* **109**, 102325.

Carpenter, J., Connolly, C. and Myers, C. K. (2008), 'Altruistic behavior in a representative dictator experiment', *Experimental Economics* **11**(3), 282–298.

Cassar, A. and Rigdon, M. L. (2021), 'Prosocial option increases women's entry into competition', *Proceedings of the National Academy of Sciences* **118**(45).

Cassar, A., Wordofa, F. and Zhang, Y. J. (2016), 'Competing for the benefit of offspring eliminates the gender gap in competitiveness', *Proceedings of the National Academy of Sciences* **113**(19), 5201–5205.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P. and Wallace, B. (2009), 'Genetic variation in preferences for giving and risk taking', *The Quarterly Journal of Economics* **124**(2), 809–842.

Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, Ö. and Wallace, B. (2010), 'Genetic variation in financial decision-making', *The Journal of Finance* **65**(5), 1725–1754.

Cesarini, D., Johannesson, M., Magnusson, P. K. and Wallace, B. (2012), 'The behavioral genetics of behavioral anomalies', *Management Science* **58**(1), 21–34.

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

Charness, G., Gneezy, U. and Halladay, B. (2016), 'Experimental methods: Pay one or pay all', *Journal of Economic Behavior & Organization* **131**, 141–150.

Charness, G., Gneezy, U. and Kuhn, M. A. (2013), 'Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics', *Journal of Economic Behavior & Organization* **91**, 93–100.

Charness, G. and Holder, P. (2019), 'Charity in the laboratory: Matching, competition, and group identity', *Management Science* **65**(3), 1398–1407.

Charness, G. and Villeval, M.-C. (2009), 'Cooperation and competition in intergenerational experiments in the field and the laboratory', *American Economic Review* **99**(3), 956–78.

Chartrand, T. L. and Bargh, J. A. (1996), 'Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions.', *Journal of Personality and Social Psychology* **71**(3), 464.

Chatelain, G., Hille, S. L., Sander, D., Patel, M., Hahnel, U. J. J. and Brosch, T. (2018), 'Feel good, stay green: Positive affect promotes pro-environmental behaviors and mitigates compensatory "mental bookkeeping" effects', *Journal of Environmental Psychology* **56**, 3–11.

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

Chen, V. L., Delmas, M. A., Locke, S. L. and Singh, A. (2017), 'Information strategies for energy conservation: A field experiment in India', *Energy Economics* **68**, 215–227.

Chichilnisky, G. and Rezai, A. (2020), *Handbook on the economics of climate change*, Edward Elgar Publishing.

Chryst, B., Marlon, J., van der Linden, S., Leiserowitz, A., Maibach, E. and Roser-Renouf, C. (2018), 'Global warming's "Six Americas Short Survey": Audience segmentation of climate change views using a four question instrument', *Environmental Communication* **12**(8), 1109–1122.

Clot, S., Della Giusta, M. and Jewell, S. (2022), 'Once good, always good? Testing nudge's spillovers on pro environmental behavior', *Environment and Behavior* **54**(3), 655–669.

Clot, S., Grolleau, G. and Ibanez, L. (2013), 'Self-licensing and financial rewards: Is morality for sale?', *Economics Bulletin* **33**, 2298–2306.

Clot, S., Grolleau, G. and Ibanez, L. (2016), 'Do good deeds make bad people?', *European Journal of Law and Economics* **42**, 491–513.

Cohen, D. (2021), 'Unilever: Breakthrough as food industry giant introduces carbon footprint labels on food', Available Online: https://www.independent.co.uk/climate-change/news/unilever-carbon-footprint-labels-food-b1882697.html. Accessed: September 2021.

Cohen, J. (2013), *Statistical power analysis for the behavioral sciences*, Academic press.

Cohen, M. A. and Viscusi, W. K. (2012), 'The role of information disclosure in climate mitigation policy', *Climate Change Economics* **3**(04), 1250020.

Cohn, A., Engelmann, J., Fehr, E. and Maréchal, M. A. (2015), 'Evidence for countercyclical risk aversion: An experiment with financial professionals', *American Economic Review* **105**(2), 860–85.

Cohn, A., Fehr, E. and Maréchal, M. A. (2017), 'Do professional norms in the banking industry favor risk-taking?', *The Review of Financial Studies* **30**(11), 3801–3823.

Cojoc, D. and Stoian, A. (2014), 'Dishonesty and charitable behavior', *Experimental Economics* **17**(4), 717–732.

Compton, N. B. (2021), 'Google Flights will now show you the environmental impact of your travel', Available Online: https://www.washingtonpost.com/travel/2021/10/06/google-flights-carbon-footprint-travel/. Accessed: December 2021.

Cortés, P., Pan, J., Pilossoph, L. and Zafar, B. (2021), Gender differences in job search and the earnings gap: Evidence from business majors, Working paper, IZA Discussion Paper.

Croson, R. and Gneezy, U. (2009), 'Gender differences in preferences', *Journal of Economic Literature* **47**(2), 448–474.

Croson, R. and Treich, N. (2014), 'Behavioral environmental economics: promises and challenges', *Environmental and Resource Economics* **58**, 335–351.

Crump, M. J., McDonnell, J. V. and Gureckis, T. M. (2013), 'Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research', *PloS one* **8**(3), e57410.

Czibor, E., Claussen, J. and van Praag, M. (2019), 'Women in a men's world: Risk taking in an online card game community', *Journal of Economic Behavior & Organization* **158**, 62–89.

Datta Gupta, N., Poulsen, A. and Villeval, M. C. (2013), 'Gender matching and competitiveness: Experimental evidence', *Economic Inquiry* **51**(1), 816–835.

de Quidt, J., Haushofer, J. and Roth, C. (2018), 'Measuring and bounding experimenter demand', *American Economic Review* **108**(11), 3266–3302.

de Quidt, J., Vesterlund, L. and Wilson, A. J. (2019), Experimenter demand effects, *in* 'Handbook of Research Methods and Applications in Experimental Economics', Edward Elgar Publishing.

Delmas, M. A., Fischlein, M. and Asensio, O. I. (2013), 'Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012', *Energy Policy* **61**, 729 – 739.

Delmas, M. A. and Grant, L. E. (2014), 'Eco-Labeling strategies and price-premium: The wine industry puzzle', *Business & Society* **53**(1), 6–44.

Dhaliwal, D. S., Li, O. Z., Tsang, A. and Yang, Y. G. (2011), 'Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting', *The Accounting Review* **86**(1), 59–100.

Dietz, T., Gardner, G. T., Gilligan, J., Stern, P. C. and Vandenbergh, M. P. (2009), 'Household actions can provide a behavioral wedge to rapidly reduce us carbon emissions', *Proceedings of the National Academy of Sciences* **106**(44), 18452–18456.

Dolan, P. and Galizzi, M. M. (2015), 'Like ripples on a pond: Behavioral spillovers and their implications for research and policy', *Journal of Economic Psychology* **47**, 1–16.

Downar, B., Ernstberger, J., Reichelstein, S., Schwenen, S. and Zaklan, A. (2021), 'The impact of carbon disclosure mandates on emissions and financial operating performance', *Review of Accounting Studies* pp. 1–39.

Dreber, A. and Johannesson, M. (2018), 'Sex hormones and economic decision making in the lab: A review of the causal evidence', *Routledge international handbook of social neuroendocrinology* pp. 391–402.

Drupp, M. A., Khadjavi, M., Riekhof, M.-C. and Voss, R. (2020), 'Professional identity and the gender gap in risk-taking. Evidence from field experiments with scientists', *Journal of Economic Behavior & Organization* **170**, 418–432.

Echeverría, R., Moreira, V. H., Sepúlveda, C. and Wittwer, C. (2014), 'Willingness to pay for carbon footprint on foods', *British Food Journal* **116 No. 2**, 186–196.

Eckel, C. C. and Grossman, P. J. (1996), 'Altruism in anonymous dictator games', *Games and Economic Behavior* **16**(2), 181–191.

Effron, D. A. (2016), '3 beyond "being good frees us to be bad"', *Cheating, Corruption, and Concealment: The Roots of Dishonesty* p. 33.

Effron, D. A., Cameron, J. S. and Monin, B. (2009), 'Endorsing Obama licenses favoring Whites', *Journal of experimental social psychology* **45**(3), 590–593.

Egan, C. and Waide, P. (2005), 'A multi-country comparative evaluation of labelling research', *European Council for an Energy Efficient Economy (ECEEE)* pp. 811–822.

Elkington, J. (1997), *Cannibals with forks – Triple bottom line of 21st century business*, Oxford: Capstone.

Ellemers, N. (2018), 'Gender stereotypes', *Annual Review of Psychology* **69**, 275–298.

Engel, J. and Szech, N. (2020), 'A little good is good enough: Ethical consumption, cheap excuses, and moral self-licensing', *PloS one* **15**(1), e0227036.

Epley, N. and Gilovich, T. (2016), 'The mechanics of motivated reasoning', *Journal of Economic perspectives* **30**(3), 133–40.

Erb, H.-P., Bioy, A. and Hilton, D. J. (2002), 'Choice preferences without inferences: Subconscious priming of risk attitudes', *Journal of Behavioral Decision Making* **15**(3), 251–262.

Eurofound (2016), *The gender employment gap: Challenges and solutions*, Publications Office of the European Union, Luxembourg.

European Commission (2019), 'Publication of the total number of allowances in circulation in 2018 for the purposes of the Market Stability Reserve under the EU Emissions Trading System established by Directive 2003/87/EC', Available Online: https://ec.europa.eu/clima/sites/clima/files/ets/reform/docs/c_2019_3288_en.pdf. Accessed: July 2021.

European Commission (2020), 'Legislative proposal on substantiating green claims. Inception Impact Assessment', Available Online: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12511-Environmental-performance-of-products-&-businesses-substantiating-claims_en. Accessed: September 2021.

European Commission (2022), 'About the competence centre on behavioural insights', Available Online: https://knowledge4policy.ec.europa.eu/behavioural-insights/about_en. Accessed: July 2023.

European Institute for Gender Equality (2023), 'Economic benefits of gender equality in the european union', Available Online: https://eige.europa.eu/newsroom/economic-benefits-gender-equality. Accessed: July 2023.

Eurostat (2023), 'Gender statistics', Available Online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_statistics. Accessed: July 2023.

Exchange Rates UK (2021), 'Exchange Rates.org.uk. The UK's favourite Currency Site', Available Online: https://www.exchangerates.org.uk/. Accessed: March 2021.

Exley, C. L. (2016), 'Excusing selfishness in charitable giving: The role of risk', *The Review of Economic Studies* **83**(2), 587–628.

Falk, A., Andre, P., Boneva, T. and Chopra, F. (2021), 'Fighting climate change: The role of norms, preferences, and moral values'.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018), 'Global evidence on economic preferences', *The Quarterly Journal of Economics* **133**(4), 1645–1692.

Federici, S., Lepri, A., Bacci, S. and Bartolucci, F. (2022), 'Male recognition bias in sex assignment based on visual stimuli', *Scientific Reports* **12**(1), 1–9.

Ferraro, P. J. and Price, M. K. (2013), 'Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment', *Review of Economics and Statistics* **95**(1), 64–73.

Festinger, L. (1957), *A theory of cognitive dissonance*, Vol. 2, Stanford university press.

Flory, J. A., Gneezy, U., Leonard, K. L. and List, J. A. (2018), 'Gender, age, and competition: A disappearing gap?', *Journal of Economic Behavior & Organization* **150**, 256–276.

Forest Trends' Ecosystem Marketplace (2020), 'Voluntary carbon and the post-pandemic recovery. State of voluntary carbon markets report, special climate week NYC 2020 installment', Washington DC: Forest Trends Association, Available Online: https://waconservationaction.org/wp-content/uploads/2020/11/EM-Voluntary-Carbon-and-Post-Pandemic-Recovery-2020.pdf. 21 September 2020.

Fornwagner, H., Grosskopf, B., Lauf, A., Schöller, V. and Städter, S. (2022), 'On the robustness of gender differences in economic behavior', *Scientific Reports* **12**(1), 21549.

Fornwagner, H., Pompeo, M. and Serdarevic, N. (2023), 'Choosing competition on behalf of someone else', *Management Science* **69**(3), 1555–1574.

Forsythe, R., Horowitz, J. L., Savin, N. E. and Sefton, M. (1994), 'Fairness in simple bargaining experiments', *Games and Economic Behavior* **6**(3), 347–369.

Franzen, A. and Pointner, S. (2013), 'The external validity of giving in the dictator game', *Experimental Economics* **16**(2), 155–169.

Freedman, J. L. and Fraser, S. C. (1966), 'Compliance without pressure: The foot-in-the-door technique.', *Journal of personality and social psychology* **4**(2), 195.

Galizzi, M. M. and Navarro-Martinez, D. (2019), 'On the external validity of social preference games: A systematic lab-field study', *Management Science* **65**(3), 976–1002.

Garvey, A. M. and Bolton, L. E. (2017), 'Eco-product choice cuts both ways: How proenvironmental licensing versus reinforcement is contingent on environmental consciousness', *Journal of Public Policy & Marketing* **36**(2), 284–298.

Gee, L. K. and Meer, J. (2019), The altruism budget: Measuring and encouraging charitable giving, Working paper, National Bureau of Economic Research.

Geldenhuys, M. and Bosch, A. (2020), 'A Rasch adapted version of the 30-Item Bem Sex Role Inventory (BSRI)', *Journal of Personality Assessment* **102**(3), 428–439.

Geng, L., Cheng, X., Tang, Z., Zhou, K. and Ye, L. (2016), 'Can previous pro-environmental behaviours influence subsequent environmental behaviours? The licensing effect of pro-environmental behaviours', *Journal of Pacific Rim Psychology* **10**, e9.

Gholamzadehmir, M., Sparks, P. and Farsides, T. (2019), 'Moral licensing, moral cleansing and pro-environmental behaviour: The moderating role of pro-environmental attitudes', *Journal of Environmental Psychology* **65**, 101334.

Gifford, R., Kormos, C. and McIntyre, A. (2011), 'Behavioral dimensions of climate change: Drivers, responses, barriers, and interventions', *Wiley Interdisciplinary Reviews: Climate Change* **2**(6), 801–827.

Gilad, D. and Kliger, D. (2008), 'Priming the risk attitudes of professionals in financial decision making', *Review of Finance* **12**(3), 567–586.

Gilg, A., Barr, S. and Ford, N. (2005), 'Green consumption or sustainable lifestyles? Identifying the sustainable consumer', *Futures* **37**(6), 481–504.

Gill, D. and Prowse, V. (2012), 'A structural analysis of disappointment aversion in a real effort competition', *American Economic Review* **102**(1), 469–503.

Gillingham, K., Kotchen, M. J., Rapson, D. S. and Wagner, G. (2013), 'The rebound effect is overplayed', *Nature* **493**(7433), 475–476.

Gleue, M., Harrs, S., Feldhaus, C. and Löschel, A. (2022), 'Identity and voluntary efforts for climate protection', *Available at SSRN 4068486* .

Global Reporting Initiative (2020*a*), 'Consolidated set of GRI Sustainability Reporting Standards 2020', Available Online: https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/. Accessed: October 2020.

Global Reporting Initiative (2020*b*), 'The GRI Standards: The global standards for sustainability reporting', Available Online: https://www.globalreporting.org/standards/media/2458/gri_standards_brochure.pdf. Accessed: October 2020.

Gneezy, A., Imas, A., Brown, A., Nelson, L. D. and Norton, M. I. (2012), 'Paying to be nice: Consistency and costly prosocial behavior', *Management Science* **58**(1), 179–187.

Gneezy, U., Imas, A. and Madarász, K. (2014), 'Conscience accounting: Emotion dynamics and social behavior', *Management Science* **60**(11), 2645–2658.

Gneezy, U., Leonard, K. L. and List, J. A. (2009), 'Gender differences in competition: Evidence from a matrilineal and a patriarchal society', *Econometrica* **77**(5), 1637–1664.

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

Gollier, C. (2001), *The economics of risk and time*, MIT press.

Gong, B. and Yang, C.-L. (2012), 'Gender differences in risk attitudes: Field experiments on the matrilineal Mosuo and the patriarchal Yi', *Journal of Economic Behavior & Organization* **83**(1), 59–65.

GOV.UK (2021*a*), 'National Statistics: Road fuel prices: 2 August 2021. Weekly prices of unleaded petrol and diesel', Available Online: https://www.gov.uk/government/statistics/road-fuel-prices-2-august-2021. Accessed: September 2021.

GOV.UK (2021*b*), 'Official Statistics. UK's carbon footprint 1997- 2018', Available Online: https://www.gov.uk/government/statistics/uks-carbon-footprint. Accessed: September 2021.

Gray, R. and Milne, M. (2002), 'Sustainability reporting: Who's kidding whom?', *Chartered Accountants Journal of New Zealand* **81**(6), 66–70.

Grewal, J., Riedl, E. J. and Serafeim, G. (2019), 'Market reaction to mandatory nonfinancial disclosure', *Management Science* **65**(7), 3061–3084.

Grieder, M., Kistler, D. and Schmitz, J. (2020), 'The hidden benefits of Corporate Social Responsibility', *Available at SSRN 3744900* .

Grieder, M., Schmitz, J. and Schubert, R. (2021), 'Asking to give: Moral licensing and prosocial behavior in the aggregate', *Available at SSRN 3920355* .

Grinstein, A., Kodra, E., Chen, S., Sheldon, S. and Zik, O. (2018), 'Carbon innumeracy', *Plos one* **13**(5), e0196282.

Grubb, M. A., Tymula, A., Gilaie-Dotan, S., Glimcher, P. W. and Levy, I. (2016), 'Neuroanatomy accounts for age-related changes in risk preferences', *Nature Communications* **7**(1), 1–5.

Gsottbauer, E., Logar, I. and van den Bergh, J. (2015), 'Towards a fair, constructive and consistent criticism of all valuation languages: Comment on Kallis et al. (2013)', *Ecological Economics* **112**, 164 – 169.

Guiso, L. and Paiella, M. (2008), 'Risk aversion, wealth, and background risk', *Journal of the European Economic Association* **6**(6), 1109–1150.

Guiso, L., Sapienza, P. and Zingales, L. (2018), 'Time varying risk aversion', *Journal of Financial Economics* **128**(3), 403–421.

Gutiérrez-Roig, M., Gracia-Lázaro, C., Perelló, J., Moreno, Y. and Sánchez, A. (2014), 'Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence', *Nature Communications* **5**(1), 1–7.

Habtezion, S. (2016), 'Gender and climate change: Overview of linkages between gender and climate change', *New York: United Nations Development Programme* .

Hahnel, U. J., Arnold, O., Waschto, M., Korcaj, L., Hillmann, K., Roser, D. and Spada, H. (2015), 'The power of putting a label on it: Green labels weigh heavier than contradicting product information for consumers' purchase decisions and post-purchase behavior', *Frontiers in Psychology* **6**, 1392.

Haley, K. J. and Fessler, D. M. (2005), 'Nobody's watching?: Subtle cues affect generosity in an anonymous economic game', *Evolution and Human Behavior* **26**(3), 245–256.

Hardies, K., Breesch, D. and Branson, J. (2013), 'Gender differences in overconfidence and risk taking: Do self-selection and socialization matter?', *Economics Letters* **118**(3), 442–444.

Harrison, G. W. (2006), 'Experimental evidence on alternative environmental valuation methods', *Environmental and Resource Economics* **34**(1), 125–162.

Hartmann, P., Marcos, A. and Barrutia, J. M. (2023), 'Carbon tax salience counteracts price effects through moral licensing', *Global Environmental Change* **78**, 102635.

He, J. C., Kang, S. K. and Lacetera, N. (2021), 'Opt-out choice framing attenuates gender differences in the decision to compete in the laboratory and in the field', *Proceedings of the National Academy of Sciences* **118**(42).

Herbohn, K. (2005), 'A full cost environmental accounting experiment', *Accounting, Organizations and Society* **30**(6), 519 – 536.

Hertwich, E. G. and Peters, G. P. (2009), 'Carbon footprint of nations: A global, trade-linked analysis', *Environmental science & technology* **43**(16), 6414–6420.

Herzig, C. and Schaltegger, S. (2011), 'Corporate sustainability reporting', *Sustainability communication: Interdisciplinary perspectives and theoretical foundation* pp. 151–169.

HM Government (2018), 'A green future: Our 25 year plan to improve the environment', Available Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf. Accessed: March 2023.

Hofmann, W., Wisneski, D. C., Brandt, M. J. and Skitka, L. J. (2014), 'Morality in everyday life', *Science* **345**(6202), 1340–1343.

Hogarth, R. M., Karelaia, N. and Trujillo, C. A. (2012), 'When should I quit? Gender differences in exiting competitions', *Journal of Economic Behavior & Organization* **83**(1), 136–150.

Holt, C. A. and Laury, S. K. (2002), 'Risk aversion and incentive effects', *American Economic Review* **92**(5), 1644–1655.

Horton, J. J., Rand, D. G. and Zeckhauser, R. J. (2011), 'The online laboratory: Conducting experiments in a real labor market', *Experimental economics* **14**(3), 399–425.

Howe, P. D., Mildenberger, M., Marlon, J. R. and Leiserowitz, A. (2015), 'Geographic variation in opinions on climate change at state and local scales in the USA', *Nature Climate Change* **5**(6), 596–603.

Huang, S. K., Kuo, L. and Chou, K.-L. (2016), 'The applicability of marginal abatement cost approach: A comprehensive review', *Journal of Cleaner Production* **127**, 59–71.

Hummel, P. and Hörisch, J. (2020), '"It's not what you say, but how you say it": How the provision of qualitative, quantitative and monetary environmental information influences companies' internal decision making', *Journal of Cleaner Production* **268**, 122247.

Hyde, J. S. (2005), 'The gender similarities hypothesis.', *American psychologist* **60**(6), 581.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C. and van Anders, S. M. (2019), 'The future of sex and gender in psychology: Five challenges to the gender binary.', *American Psychologist* **74**(2), 171.

Ibanez, L. and Roussel, S. (2021), 'The effects of induced emotions on environmental preferences and behavior: An experimental study', *Plos one* **16**(9), e0258045.

Imai, T., Pace, D. D., Schwardmann, P. and van der Weele, J. J. (2022), 'Correcting consumer misperceptions about co 2 emissions'.

Inter-Parliamentary Union & United Nations Entity for Gender Equality and the Empowerment of Women (2023), 'Women in politics: 2023', Available Online: https://www.unwomen.org/sites/default/files/2023-03/Women-in-politics-2023-en.pdf. Accessed: July 2023.

Intergovernmental Panel on Climate Change (2022), Summary for Policymakers [H.-O. Pörtner, D.C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem (eds.)], *in* 'Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]', Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jordan, J., Mullen, E. and Murnighan, J. K. (2011), 'Striving for the moral self: The effects of recalling past moral actions on future moral behavior', *Personality & Social Psychology Bulletin* **37**(5), 701–713.

Jouvenot, V. and Krueger, P. (2019), 'Mandatory corporate carbon disclosure: Evidence from a natural experiment', *Available at SSRN 3434490* .

Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1986), 'Fairness and the assumptions of economics', *Journal of Business* **59**(4), S285–S300.

Kaklamanou, D., Jones, C. R., Webb, T. L. and Walker, S. R. (2015), 'Using public transport can make up for flying abroad on holiday: Compensatory green beliefs and environmentally significant behavior', *Environment and Behavior* **47**(2), 184–204.

Kallbekken, S. and Sælen, H. (2013), ''Nudging' hotel guests to reduce food waste as a win–win environmental measure', *Economics Letters* **119**(3), 325–327.

Kallis, G., Gómez-Baggethun, E. and Zografos, C. (2015), 'The limits of monetization in valuing the environment', *Ecological Economics* **112**(112), 170–173.

Kallis, G., Gómez-Baggethun, E. and Zografos, C. (2013), 'To value or not to value? That is not the question', *Ecological Economics* **94**, 97 – 105.

Kamas, L. and Preston, A. (2015), 'Can social preferences explain gender differences in economic behavior?', *Journal of Economic Behavior & Organization* **116**, 525–539.

Kappes, A., Nussberger, A.-M., Faber, N. S., Kahane, G., Savulescu, J. and Crockett, M. J. (2018), 'Uncertainty about the impact of social decisions increases prosocial behaviour', *Nature Human Behaviour* **2**(8), 573–580.

Kastlunger, B., Dressler, S. G., Kirchler, E., Mittone, L. and Voracek, M. (2010), 'Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2d: 4d)', *Journal of Economic Psychology* **31**(4), 542–552.

Keaveney, A., Peters, E. and Way, B. (2020), 'Effects of acetaminophen on risk taking', *Social Cognitive and Affective Neuroscience* **15**(7), 725–732.

Kessel, D., Mollerstrom, J. and van Veldhuizen, R. (2021), 'Can simple advice eliminate the gender gap in willingness to compete?', *European Economic Review* **138**, 103777.

Kessler, J. B. and Milkman, K. L. (2018), 'Identity in charitable giving', *Management Science* **64**(2), 845–859.

Kessler, S. J. and McKenna, W. (1985), *Gender: An ethnomethodological approach*, University of Chicago Press.

Kessler, S. J. and McKenna, W. (2000), 'Gender construction in everyday life: Transsexualism (abridged)', *Feminism & Psychology* **10**(1), 11–29.

Kesternich, M., Reif, C. and Rübbelke, D. (2017), 'Recent trends in behavioral environmental economics', *Environmental and Resource Economics* **67**, 403–411.

Klarna Bank AB (2021), 'Klarna launches carbon footprint insights for 90 million consumers', Available Online: https://www.klarna.com/international/press/klarna-launches-carbon-footprint-insights-for-90-million-consumers/. Accessed: December 2021.

Kleven, H., Landais, C. and Søgaard, J. E. (2019), 'Children and gender inequality: Evidence from Denmark', *American Economic Journal: Applied Economics* **11**(4), 181–209.

Kollmuss, A. and Agyeman, J. (2002), 'Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior?', *Environmental Education Research* **8**(3), 239–260.

König-Kersting, C. and Trautmann, S. T. (2018), 'Countercyclical risk aversion: Beyond financial professionals', *Journal of Behavioral and Experimental Finance* **18**, 94–101.

Kozee, H. B., Tylka, T. L. and Bauerband, L. A. (2012), 'Measuring transgender individuals' comfort with gender identity and appearance: Development and validation of the transgender congruence scale', *Psychology of Women Quarterly* **36**(2), 179–196.

KPMG (2017), 'The road ahead. The KPMG survey of corporate responsibility reporting 2017', Available Online: https://assets.kpmg/content/dam/kpmg/xx/pdf/2017/10/kpmg-survey-of-corporate-responsibility-reporting-2017.pdf. Accessed: October 2020.

Kuper, N. and Bott, A. (2019), 'Has the evidence for moral licensing been inflated by publication bias?', *Meta-Psychology* **3**.

Lacasse, K. (2016), 'Don't be satisfied, identify! Strengthening positive spillover by connecting pro-environmental behaviors to an "environmentalist" label', *Journal of Environmental Psychology* **48**, 149–158.

Lalot, F., Falomir-Pichastor, J. M. and Quiamzade, A. (2022), 'Regulatory focus and self-licensing dynamics: A motivational account of behavioural consistency and balancing', *Journal of Environmental Psychology* **79**, 101731.

Lanzini, P. and Thøgersen, J. (2014), 'Behavioural spillover in the environmental domain: an intervention study', *Journal of Environmental Psychology* **40**, 381–390.

Lehner, M., Mont, O. and Heiskanen, E. (2016), 'Nudging – A promising tool for sustainable consumption behaviour?', *Journal of Cleaner Production* **134**, 166 – 177.

Leibbrandt, A. and List, J. A. (2015), 'Do women avoid salary negotiations? Evidence from a large-scale natural field experiment', *Management Science* **61**(9), 2016–2024.

Leibbrandt, A., Wang, L. C. and Foo, C. (2018), 'Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women', *Management Science* **64**(8), 3501–3516.

Leire, C. and Åke Thidell (2005), 'Product-related environmental information to guide consumer purchases – a review and analysis of research on perceptions, understanding and use among Nordic consumers', *Journal of Cleaner Production* **13**(10), 1061 – 1070.

Lemaster, P. and Strough, J. (2014), 'Beyond Mars and Venus: Understanding gender differences in financial risk tolerance', *Journal of Economic Psychology* **42**, 148–160.

Leythienne, D. and Pérez-Julián, M. (2021), *Gender pay gaps in the European Union: A statistical analysis*, Publications Office of the European Union.

Li, Q., Long, R. and Chen, H. (2017), 'Empirical study of the willingness of consumers to purchase low-carbon products by considering carbon labels: A case study', *Journal of Cleaner Production* **161**, 1237–1250.

Liebe, U., Gewinner, J. and Diekmann, A. (2021), 'Large and persistent effects of green energy defaults in the household and business sectors', *Nature Human Behaviour* **5**(5), 576–585.

Lind, J. T., Nyborg, K. and Pauls, A. (2019), 'Save the planet or close your eyes? Testing strategic ignorance in a charity context', *Ecological Economics* **161**, 9–19.

List, J. A. and Momeni, F. (2021), 'When corporate social responsibility backfires: Evidence from a natural field experiment', *Management Science* **67**(1), 8–21.

Liu, E. M. and Zuo, S. X. (2019), 'Measuring the impact of interaction between children of a matrilineal and a patriarchal culture on gender differences in risk aversion', *Proceedings of the National Academy of Sciences* **116**(14), 6713–6719.

Longoni, C., Gollwitzer, P. M. and Oettingen, G. (2014), 'A green paradox: Validating green choices has ironic effects on behavior, cognition, and perception', *Journal of Experimental Social Psychology* **50**, 158–165.

Löschel, A., Sturm, B. and Vogt, C. (2013), 'The demand for climate protection—Empirical evidence from Germany', *Economics Letters* **118**(3), 415 – 418.

Maas, K., Schaltegger, S. and Crutzen, N. (2016), 'Integrating corporate sustainability assessment, management accounting, control, and reporting', *Journal of Cleaner Production* **136**, 237 – 248.

Maki, A., Burns, R. J., Ha, L. and Rothman, A. J. (2016), 'Paying people to protect the environment: A meta-analysis of financial incentive interventions to promote proenvironmental behaviors', *Journal of Environmental Psychology* **47**, 242 – 255.

Maki, A., Carrico, A. R., Raimi, K. T., Truelove, H. B., Araujo, B. and Yeung, K. L. (2019), 'Meta-analysis of pro-environmental behaviour spillover', *Nature Sustainability* **2**(4), 307–315.

Margetts, E. A. and Kashima, Y. (2017), 'Spillover between pro-environmental behaviours: The role of resources and perceived similarity', *Journal of Environmental Psychology* **49**, 30–42.

Marianne, B. (2011), New perspectives on gender, *in* 'Handbook of labor economics', Vol. 4, Elsevier, pp. 1543–1590.

Martin, P. R. and Moser, D. V. (2016), 'Managers' green investment disclosures and investors' reaction', *Journal of Accounting and Economics* **61**(1), 239 – 254.

Martin, R., Muûls, M. and Wagner, U. J. (2016), 'The impact of the European Union Emissions Trading Scheme on regulated firms: What is the evidence after ten years?', *Review of Environmental Economics and Policy* **10**(1), 129–148.

Mazar, N. and Zhong, C.-B. (2010), 'Do green products make us better people?', *Psychological Science* **21**(4), 494–498.

McDermott, R. C., Brasil, K. M., Borgogna, N. C., Barinas, J. and Levant, R. F. (2022), 'Traditional masculinity ideology and feminist attitudes: The role of identity foreclosure', *Sex Roles* **87**(3), 211–222.

McKay, R., Efferson, C., Whitehouse, H. and Fehr, E. (2011), 'Wrath of god: Religious primes and punishment', *Proceedings of the Royal Society B: Biological Sciences* **278**(1713), 1858–1863.

Meier-Pesti, K. and Penz, E. (2008), 'Sex or gender? Expanding the sex-based view by introducing masculinity and femininity as predictors of financial risk taking', *Journal of Economic Psychology* **29**(2), 180–196.

Meijers, M. H., Noordewier, M. K., Verlegh, P. W., Willems, W. and Smit, E. G. (2019), 'Paradoxical side effects of green advertising: How purchasing green products may instigate licensing effects for consumers with a weak environmental identity', *International Journal of Advertising* **38**(8), 1202–1223.

Meijers, M. H., Verlegh, P. W., Noordewier, M. K. and Smit, E. G. (2015), 'The dark side of donating: How donating may license environmentally unfriendly behavior', *Social Influence* **10**(4), 250–263.

Menyhart, O., Weltz, B. and Győrffy, B. (2021), 'Multipletesting. com: A tool for life science researchers for multiple hypothesis testing correction', *PLOS ONE* **16**(6), e0245824.

Merritt, A. C., Effron, D. A. and Monin, B. (2010), 'Moral self-licensing: When being good frees us to be bad', *Social and Personality Psychology Compass* **4**(5), 344–357.

Mi, L., Gan, X., Sun, Y., Lv, T., Qiao, L. and Xu, T. (2021), 'Effects of monetary and non-monetary interventions on energy conservation: A meta-analysis of experimental studies', *Renewable and Sustainable Energy Reviews* **149**, 111342.

Michaud, C., Llerena, D. and Joly, I. (2013), 'Willingness to pay for environmental attributes of non-food agricultural products: A real choice experiment', *European Review of Agricultural Economics* **40**(2), 313–329.

Miller, D. T. and Effron, D. A. (2010), Psychological license: When it is needed and how it functions, *in* 'Advances in experimental social psychology', Vol. 43, Elsevier, pp. 115–155.

Milne, M. J. and Gray, R. (2013), 'W(h)ither ecology? The triple bottom line, the global reporting initiative, and corporate sustainability reporting', *Journal of Business Ethics* **118**(1), 13–29.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R. and Grafman, J. (2006), 'Human fronto–mesolimbic networks guide decisions about charitable donation', *Proceedings of the National Academy of Sciences* **103**(42), 15623–15628.

Monin, B. and Miller, D. T. (2001), 'Moral credentials and the expression of prejudice', *Journal of Personality and Social Psychology* **81**(1), 33.

Mullainathan, S. and Thaler, R. H. (2000), Behavioral economics, Working paper, National Bureau of Economic Research.

Mullen, E. and Monin, B. (2016), 'Consistency versus licensing effects of past moral behavior', *Annual Review of Psychology* **67**, 363–385.

Mummolo, J. and Peterson, E. (2019), 'Demand effects in survey experiments: An empirical assessment', *American Political Science Review* **113**(2), 517–529.

Murphy, J. J., Allen, P. G., Stevens, T. H. and Weatherhead, D. (2005), 'A meta-analysis of hypothetical bias in stated preference valuation', *Environmental and Resource Economics* **30**(3), 313–325.

Mussweiler, T. and Förster, J. (2000), 'The sex→ aggression link: A perception-behavior dissociation.', *Journal of Personality and Social Psychology* **79**(4), 507.

Myers, R. M. and Griffin, A. L. (2019), 'The geography of gender inequality in international higher education', *Journal of Studies in International Education* **23**(4), 429–450.

Nelson, J. A. (2014), 'The power of stereotyping and confirmation bias to overwhelm accurate assessment: The case of economics, gender, and risk aversion', *Journal of Economic Methodology* **21**(3), 211–231.

Nelson, J. A. (2015), 'Are women really more risk-averse than men? A re-analysis of the literature using expanded methods', *Journal of economic surveys* **29**(3), 566–585.

Nelson, J. A. (2016), 'Not-so-strong evidence for gender differences in risk taking', *Feminist Economics* **22**(2), 114–142.

Nelson, J. A. (2017), *Gender and risk-taking: Economics, evidence, and why the answer matters*, Routledge.

Ness, B., Urbel-Piirsalu, E., Anderberg, S. and Olsson, L. (2007), 'Categorising tools for sustainability assessment', *Ecological Economics* **60**(3), 498 – 508.

Newell, B. R. and Shaw, B. (2017), 'Priming risky choice: Do risk preferences need inferences?', *Journal of Behavioral Decision Making* **30**(2), 332–346.

Niederle, M. (2017), 'A gender agenda: A progress report on competitiveness', *American Economic Review* **107**(5), 115–119.

Niederle, M. and Vesterlund, L. (2007), 'Do women shy away from competition? Do men compete too much?', *The Quarterly Journal of Economics* **122**(3), 1067–1101.

Nilsson, A., Bergquist, M. and Schultz, W. P. (2017), 'Spillover effects in environmental behaviors, across time and context: a review and research agenda', *Environmental Education Research* **23**(4), 573–589.

Nisan, M. and Horenczyk, G. (1990), 'Moral balance: The effect of prior behaviour on decision in moral conflict', *British journal of social psychology* **29**(1), 29–42.

Noblet, C. L. and McCoy, S. K. (2018), 'Does one good turn deserve another? Evidence of domain-specific licensing in energy behavior', *Environment and Behavior* **50**(8), 839–863.

Nordhaus, W. (2019), 'Climate change: The ultimate challenge for economics', *American Economic Review* **109**(6), 1991–2014.

Observatory of Public Sector Innovation (n.d.), 'The OECD network of behavioural insights experts in government', Available Online: https://oecd-opsi.org/bi-network/. Accessed: July 2023.

Office for National Statistics (2021), 'Average household income, UK: Financial year 2020', Available Online: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyear2020. Accessed: August 2021.

Office for National Statistics (2022), 'Statistical bulletin, Population and household estimates, England and Wales: Census 2021, unrounded data', Available Online: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationandhouseholdestimatesenglandandwales/census2021unroundeddata. Accessed: November 2022.

Organisation for Economic Co-operation and Development (2006), 'Cost-benefit analysis and the environment', Available Online: "https://www.oecd-ilibrary.org/content/publication/9789264010055-en".

Organisation for Economic Co-operation and Development (2015), 'What lies behind gender inequality in education?', Available Online: https://www.oecd-ilibrary.org/content/paper/5js4xffhhc30-en.

Oxford English Dictionary (n.d.), 'sex', Available Online: https://www.oed.com/dictionary/sex_n1. Accessed: October 2021.

o'Dea, R. E., Lagisz, M., Jennions, M. D. and Nakagawa, S. (2018), 'Gender differences in individual variation in academic grades fail to fit expected patterns for STEM', *Nature Communications* **9**(1), 1–8.

Pace, D. and van der Weele, J. J. (2020), Curbing carbon: An experiment on uncertainty and information about CO2 emissions, Working paper, Tinbergen Institute Discussion Paper 2020-059/I.

Panzone, L. A., Ulph, A., Zizzo, D. J., Hilton, D. and Clear, A. (2021), 'The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping', *Journal of Environmental Economics and Management* **109**, 102137.

Pathak, M., Slade, R., Shukla, P., Skea, J., Pichs-Madruga, R. and Ürge Vorsatz, D. (2022), Technical summary, *in* P. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz and J. Malley, eds, 'Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press.

Peer, E., Brandimarte, L., Samat, S. and Acquisti, A. (2017), 'Beyond the turk: Alternative platforms for crowdsourcing behavioral research', *Journal of Experimental Social Psychology* **70**, 153–163.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z. and Damer, E. (2021), 'Data quality of platforms and panels for online behavioral research', *Behavior Research Methods* pp. 1–20.

Pichert, D. and Katsikopoulos, K. V. (2008), 'Green defaults: Information presentation and pro-environmental behaviour', *Journal of Environmental Psychology* **28**(1), 63–73.

Pichon, I., Boccato, G. and Saroglou, V. (2007), 'Nonconscious influences of religion on prosociality: A priming study', *European Journal of Social Psychology* **37**(5), 1032–1045.

Pizzol, M., Weidema, B., Brandão, M. and Osset, P. (2015), 'Monetary valuation in Life Cycle Assessment: A review', *Journal of Cleaner Production* **86**, 170 – 179.

Ploner, M. and Regner, T. (2013), 'Self-image and moral balancing: An experimental analysis', *Journal of Economic Behavior & Organization* **93**, 374–383.

Prothero, A., Dobscha, S., Freund, J., Kilbourne, W. E., Luchs, M. G., Ozanne, L. K. and Thøgersen, J. (2011), 'Sustainable consumption: Opportunities for consumer research and public policy', *Journal of Public Policy & Marketing* **30**(1), 31–38.

Ranehill, E., Zethraeus, N., Blomberg, L., von Schoultz, B., Hirschberg, A. L., Johannesson, M. and Dreber, A. (2018), 'Hormonal contraceptives do not impact economic preferences: Evidence from a randomized trial', *Management Science* **64**(10), 4515–4532.

Rennert, K., Errickson, F., Prest, B. C., Rennels, L., Newell, R. G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B. et al. (2022), 'Comprehensive evidence implies a higher social cost of CO2', *Nature* **610**(7933), 687–692.

Reuben, E., Sapienza, P. and Zingales, L. (2015), Taste for competition and the gender gap among young business professionals, Working Paper 21695, National Bureau of Economic Research.

Reuben, E., Wiswall, M. and Zafar, B. (2017), 'Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender', *The Economic Journal* **127**(604), 2153–2186.

Reuter, M., Frenzel, C., Walter, N. T., Markett, S. and Montag, C. (2011), 'Investigating the genetic basis of altruism: the role of the COMT Val158Met polymorphism', *Social Cognitive and Affective Neuroscience* **6**(5), 662–668.

Rigdon, M., Ishii, K., Watabe, M. and Kitayama, S. (2009), 'Minimal social cues in the dictator game', *Journal of Economic Psychology* **30**(3), 358–367.

Ryan, M. and San Miguel, F. (2000), 'Testing for consistency in willingness to pay experiments', *Journal of Economic Psychology* **21**(3), 305–317.

Saccardo, S., Pietrasz, A. and Gneezy, U. (2018), 'On the size of the gender difference in competitiveness', *Management Science* **64**(4), 1541–1554.

Sachdeva, S., Iliev, R. and Medin, D. L. (2009), 'Sinning saints and saintly sinners: The paradox of moral self-regulation', *Psychological Science* **20**(4), 523–528.

Samson, A. (2014), 'The behavioral economics guide 2014', *London, UK: Behavioral Economics Group* .

Sapienza, P., Zingales, L. and Maestripieri, D. (2009), 'Gender differences in financial risk aversion and career choices are affected by testosterone', *Proceedings of the National Academy of Sciences* **106**(36), 15268–15273.

Schmitz, J. (2019), 'Temporal dynamics of pro-social behavior: An experimental analysis', *Experimental Economics* **22**, 1–23.

Schöller, V. and Schlereth, L. (2023), 'Be green or feel green? An experiment on moral balancing in pro-environmental decision making'.

Schöller, V. and Ulmer, C. (2023), 'Can monetized carbon information increase pro-environmental behavior? Experimental evidence', *Ecological Economics* **206**, 107759.

Schultz, P. W., Estrada, M., Schmitt, J., Sokoloski, R. and Silva-Send, N. (2015), 'Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms', *Energy* **90**, 351 – 358.

Schwartz, D., Bruine de Bruin, W., Fischhoff, B. and Lave, L. (2015), 'Advertising energy saving programs: The potential environmental cost of emphasizing monetary savings', *Journal of Experimental Psychology: Applied* **21**(2), 158–166.

Sen, S. and Vollebergh, H. (2018), 'The effectiveness of taxing the carbon content of energy consumption', *Journal of Environmental Economics and Management* **92**, 74–99.

Serafeim, G., Zochowski, T. R. and Downing, J. (2020), 'Impact-weighted financial accounts: The missing piece for an impact economy', Harvard Business School. Available Online: https://www.hbs.edu/impact-weighted-accounts/Pages/default.aspx. Accessed: October 2020.

Shariff, A. F. and Norenzayan, A. (2007), 'God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game', *Psychological Science* **18**(9), 803–809.

Shurchkov, O. and Eckel, C. C. (2018), *Gender differences in behavioral traits and labor market outcomes*, Oxford, UK: Oxford University Press.

Simbrunner, P. and Schlegelmilch, B. B. (2017), 'Moral licensing: A culture-moderated meta-analysis', *Management Review Quarterly* **67**(4), 201–225.

Singh, R. K., Murty, H., Gupta, S. and Dikshit, A. (2009), 'An overview of sustainability assessment methodologies', *Ecological Indicators* **9**(2), 189 – 212.

Sintov, N., Geislar, S. and White, L. V. (2019), 'Cognitive accessibility as a new factor in proenvironmental spillover: results from a field study of household food waste management', *Environment and Behavior* **51**(1), 50–80.

Sisco, M. R. and Weber, E. U. (2019), 'Examining charitable giving in real-world online donations', *Nature Communications* **10**(1), 1–8.

Spence, A., Leygue, C., Bedwell, B. and O'Malley, C. (2014), 'Engaging with energy reduction: Does a climate change frame have the potential for achieving broader sustainable behaviour?', *Journal of Environmental Psychology* **38**, 17 – 28.

Steele, J. R. and Ambady, N. (2006), '"Math is hard!" The effect of gender priming on women's attitudes', *Journal of Experimental Social Psychology* **42**(4), 428–436.

Steg, L. and Vlek, C. (2009), 'Encouraging pro-environmental behaviour: An integrative review and research agenda', *Journal of Environmental Psychology* **29**(3), 309–317.

Steinhorst, J. and Klöckner, C. A. (2018), 'Effects of monetary versus environmental information framing: Implications for long-term pro-environmental behavior and intrinsic motivation', *Environment and Behavior* **50**(9), 997–1031.

Stern, N. (2007), 'Climate change, ethics and the economics of the global deal', Annual Public Lecture, 2007-11-29 – 2007-11-30.

Sunstein, C. R. and Reisch, L. A. (2014), 'Automatically green: Behavioral economics and environmental protection', *Harvard Environmental Law Review* **38**, 127.

Sunstein, C. R. and Thaler, R. H. (2008), 'Privatizing marriage', *The Monist* **91**(3/4), 377–387.

Sustainable Consumption Roundtable (2006), 'I will if you will. Towards sustainable consumption', Available Online: https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/2312/sdc-2006-i-will-if-you-will.pdf?sequence=1&isAllowed=y. Accessed: March 2023.

Sutter, M. and Glätzle-Rützler, D. (2015), 'Gender differences in the willingness to compete emerge early in life and persist', *Management Science* **61**(10), 2339–2354.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D. and Trautmann, S. T. (2013), 'Impatience and uncertainty: Experimental decisions predict adolescents' field behavior', *American Economic Review* **103**(1), 510–31.

Tan, H. B. and Forgas, J. P. (2010), 'When happiness makes us selfish, but sadness makes us fair: Affective influences on interpersonal strategies in the dictator game', *Journal of Experimental Social Psychology* **46**(3), 571–576.

Taschini, L., Dietz, S. and Hicks, N. (2013), 'Carbon tax v cap-and-trade: Which is better', *The Guardian* . Available Online: https://www.theguardian.com/environment/2013/jan/31/carbon-tax-cap-and-trade, Accessed: July 2023.

Thaler, R. (1985), 'Mental accounting and consumer choice', *Marketing Science* **4**(3), 199–214.

Thaler, R. H. and Sunstein, C. R. (2021), *Nudge: The Final Edition*, Penguin.

Thøgersen, J. and Crompton, T. (2009), 'Simple and painless? The limitations of spillover in environmental campaigning', *Journal of Consumer Policy* **32**(2), 141–163.

Thøgersen, J. and Ölander, F. (2003), 'Spillover of environment-friendly consumer behaviour', *Journal of Environmental Psychology* **23**(3), 225–236.

Thöni, C. and Volk, S. (2021), 'Converging evidence for greater male variability in time, risk, and social preferences', *Proceedings of the National Academy of Sciences* **118**(23).

Tiefenbeck, V., Staake, T., Roth, K. and Sachs, O. (2013), 'For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign', *Energy Policy* **57**, 160–171.

Tomar, S. (2023), 'Greenhouse gas disclosure and emissions benchmarking', *Journal of Accounting Research* **61**(2), 451–492.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T. and Vandenbergh, M. P. (2014), 'Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework', *Global Environmental Change* **29**, 127–138.

Truelove, H. B., Yeung, K. L., Carrico, A. R., Gillis, A. J. and Raimi, K. T. (2016), 'From plastic bottle recycling to policy support: An experimental test of pro-environmental spillover', *Journal of Environmental Psychology* **46**, 55–66.

Tversky, A. and Kahneman, D. (1974), 'Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty', *Science* **185**(4157), 1124–1131.

United Nations (2020), 'The sustainable development goals report 2020'. Available Online: https://unstats.un.org/sdgs/report/2020/The-Sustainable-Development-Goals-Report-2020.pdf. Accessed: January 2021.

United Nations Entity for Gender Equality and the Empowerment of Women & Department of Economic and Social Affairs (2022), 'Progress on the sustainable development goals: The gender snapshot 2022', Available Online: https://dspace.ceid.org.tr/xmlui/bitstream/handle/1/2114/Progress-on-the-sustainable-development-goals-the-gender-snapshot-2022-en_0.pdf?sequence=1&isAllowed=y. Accessed: July 2023.

United Nations Environment Program (2019), 'Emissions Gap Report 2019', Available Online: https://wedocs.unep.org/bitstream/handle/20.500.11822/30797/EGR2019.pdf?sequence=1&isAllowed=y. Accessed: October 2020.

United Nations Framework Convention on Climate Change (2015), 'Paris Agreement', Available Online: https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf. Accessed: October 2020.

Urban, J., Bahník, Š. and Kohlová, M. B. (2019), 'Green consumption does not make people cheat: Three attempts to replicate moral licensing effect due to pro-environmental behavior', *Journal of Environmental Psychology* **63**, 139–147.

van Anders, S. M., Steiger, J. and Goldey, K. L. (2015), 'Effects of gendered behavior on testosterone in women and men', *Proceedings of the National Academy of Sciences* **112**(45), 13805–13810.

van der Werff, E., Steg, L. and Keizer, K. (2014), 'Follow the signal: When past pro-environmental actions signal who you are', *Journal of Environmental Psychology* **40**, 273–282.

VandenBos, G. R. (2007), *APA dictionary of psychology.*, American Psychological Association.

Venmans, F. (2012), 'A literature-based multi-criteria evaluation of the EU ETS', *Renewable and Sustainable Energy Reviews* **16**(8), 5493 – 5510.

Villeval, M. C. (2012), 'Ready, steady, compete', *Science* **335**(6068), 544–545.

Vlek, C. and Steg, L. (2007), 'Human behavior and environmental sustainability: Problems, driving forces, and research topics'.

von Gaudecker, H.-M., van Soest, A. and Wengstrom, E. (2011), 'Heterogeneity in risky choice behavior in a broad population', *American Economic Review* **101**(2), 664–94.

Wee, S.-C., Choong, W.-W. and Low, S.-T. (2021), 'Can "nudging" play a role to promote pro-environmental behaviour?', *Environmental Challenges* **5**, 100364.

Weitzman, M. L. (1974), 'Prices vs. quantities', *The Review of Economic Studies* **41**(4), 477–491.

Wenzlaff, F., Briken, P. and Dekker, A. (2018), 'If there's a penis, it's most likely a man: Investigating the social construction of gender using eye tracking', *PloS one* **13**(3), e0193616.

West, C. and Zhong, C.-B. (2015), 'Moral cleansing', *Current Opinion in Psychology* **6**, 221–225.

Whitmarsh, L. and O'Neill, S. (2010), 'Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours', *Journal of Environmental Psychology* **30**(3), 305–314.

Wilcox, K., Vallen, B., Block, L. and Fitzsimons, G. J. (2009), 'Vicarious goal fulfillment: When the mere presence of a healthy option leads to an ironically indulgent decision', *Journal of Consumer Research* **36**(3), 380–393.

Willis Towers Watson (2022), '2022 global gender wealth equity report: The role of gender in wealth equity', Available Online: https://www.wtwco.com/en-nz/insights/2022/11/2022-global-gender-wealth-equity-report. Accessed: July 2023.

World Bank (2023), 'State and trends of carbon pricing 2023', Available Online: https://openknowledge.worldbank.org/bitstreams/bdd449bb-c298-4eb7-a794-c80bfe209f4a/download. Accessed: July 2023.

World Commission on Environment and Development (1987), 'Our Common Future', Available Online: https://digitallibrary.un.org/record/139811. Accessed: October 2020.

World Health Organization (2021), 'Climate change and health', Available Online: https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health. Accessed: June 2023.

Wozniak, D., Harbaugh, W. T. and Mayr, U. (2014), 'The menstrual cycle and performance feedback alter gender differences in competitive choices', *Journal of Labor Economics* **32**(1), 161–198.

Wozniak, D., Harbaugh, W. T. and Mayr, U. (2016), 'The effect of feedback on gender differences in competitive choices.', *Available at SSRN 1976073* .

Wu, J.-J., Ji, T., He, Q.-Q., Du, J. and Mace, R. (2015), 'Cooperation is related to dispersal patterns in sino-tibetan populations', *Nature Communications* **6**(1), 1–6.

Wynes, S. and Nicholas, K. A. (2017), 'The climate mitigation gap: Education and government recommendations miss the most effective individual actions', *Environmental Research Letters* **12**(7), 074024.

Wynes, S., Zhao, J. and Donner, S. D. (2020), 'How well do people understand the climate impact of individual actions?', *Climatic Change* **162**(3), 1521–1534.

XE.com Inc. (2021), 'Historical rate tables', Available Online: https://www.xe.com/currencytables/?from=USD&date=2021-06-16#table-section. Accessed: February 2023.

Yokessa, M. and Marette, S. S. (2019), 'A review of eco-labels and their economic impact', *International Review of Environmental and Resource Economics* **13 (1–2)**(1-2), 119–163.

Zak, P. J., Kurzban, R., Ahmadi, S., Swerdloff, R. S., Park, J., Efremidze, L., Redwine, K., Morgan, K. and Matzner, W. (2009), 'Testosterone administration decreases generosity in the ultimatum game', *PLOS ONE* **4**(12), e8330.

Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., von Schoultz, B., Hirschberg, A. L. and Johannesson, M. (2009), 'A randomized trial of the effect of estrogen and testosterone on economic behavior', *Proceedings of the National Academy of Sciences* **106**(16), 6535–6538.

Zhang, Y. J. (2013), 'Can experimental economics explain competitive behavior outside the lab?', *Available at SSRN 2292929* .

Zhang, Y., Jiang, J., Sun, Y., Gu, D. and Jiang, W. (2021), 'Engagement in cause-related marketing reduces pro-environmental behaviors', *Environment and Behavior* **53**(10), 1047–1069.