universität wien

MASTERARBEIT/MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Time, Interaction, Markets and Prices: What shapes our social norms?

verfasst von / submitted by

Jan Christian Kaiser, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Volkswirtschaftslehre
Betreut von / Supervisor:	Univ.-Prof. Dr. Wieland Müller
Mitbetreut von / Co-Supervisor:	Dipl.-Math. oec. Dr. Christian Koch

Abstract

English

This thesis introduces an experimental framework to examine the effects of prizes as an external stimulus on the outcome of repeated stage games with an externality without changing the underlying monetary incentives.

The validity of this framework is demonstrated by the results of two online economic experiments conducted with paid participants that exhibited statistically significant differences in patterns of behavior by type within each trial but also between treatments.

Additionally, the effects of each type and treatment combination on the formation of social norms is measured by a Krupka-Weber coordination game and we find a significant difference in the level of perceived social appropriateness of the externality causing action for those subjects that are directly affected.

German

Diese Arbeit präsentiert eine Struktur für ökonomische Experimente, um die Auswirkungen von Preisen als externen Reiz auf die Resultate wiederholter Spiele mit einer Externalität zu untersuchen, ohne dabei die grundliegenden monetären Anreize zu ändern.

Deren Validität wird durch die Ergebnisse von zwei Online-Experimenten mit bezahlten Teilnehmern nachgewiesen. Dabei finden wir statistisch signifikante Unterschiede in den Verhaltensmustern nach zugewiesenem Typ innerhalb, aber auch zwischen den Versuchsaufbauten. Zusätzlich werden die Auswirkungen jeder Typ- und Behandlungskombination auf die Bildung sozialer Normen durch ein Koordinationsspiel nach Krupka-Weber gemessen, und wir stellen einen signifikanten Unterschied im Grad der wahrgenommenen sozialen Angemessenheit durch die von der Externalität Betroffenen fest.

Acknowledgements

I’d like to express my gratitude to Christian Koch for his insights, persistence and patience that made this thesis possible. I’d also like to thank my coworkers and superiors for their support that enabled me to study for a master’s degree on top of an often more than full-time work week. Lastly, I would also like to acknowledge the impact of open-source software and standards on this study and countless others.

Contents

1 Introduction

Traditionally economists have argued that interactions and more specifically markets reveal but do not shape preferences. A growing body of literature however does find evidence of erosion of moral behavior depending on the type of interaction especially compared to incentivized choice environments. While a multitude of underlying causal mechanisms have been proposed including social learning, replacements effects and moral wiggle room, this thesis attempts to identify and measure the impact of prizes on behavior independently of their effects on the underlying incentives.

Based on an idea by Christian Koch, the goal of this thesis is to develop an experimental design that enables us to distinguish and jointly analyze the effects of *market interaction*, that have played a large role in recent economic literature, and *"putting a price"* on a good, as discussed in the philosophical literature, on moral behavior.

In this study, this is implemented through a repeated stage game featuring two "active" participants with different sets of monetary payoffs whose actions directly impact a third individual without agency via an externality.

Additionally, this framework is further varied by introducing an external fixed price in a second treatment. This stimulus is calibrated specifically to not affect the monetary incentives of the active participants, enabling us to study its effects in separation.

Lastly, we attempt to measure the effects of both assigned role and treatment on participants' perception of prevailing social norms via a Krupka, Weber (2013) coordination game which incentivizes the correct identification of the modal value of all subjects in their treatment.

Vetted participants were recruited through the Prolific online platform and interacted live with their matched counterparts in a separate customized Lioness environment.

Our hypothesis was to find differences in behavior and norm perception not only between assigned roles and therefore monetary incentives but also between the two treatments, i.e., depending on the presence of an externally set price.

In our results, we find strong statistical evidence for different behavior between assigned types and for the subset with comparatively smaller monetary incentives for the externality causing action also statistical support for differences between treatments.

Counterintuitively, we find no evidence that the introduction of an external price stimulus shapes norm perception for active players. However, those without agency exhibit a significant shift in norm perception to a markedly more positive view of the "negative" action.

While our initial expectation was to find stronger evidence of replacement effects or crowding out of norms for active stage game participants this could be due to the specific calibration of the monetary incentives, the value of the stimulus and the chosen framing device of a tax.

Controlled variation of these 3 factors in future research could enable further insight into the mechanisms affecting behavior and the formation of norms and individual preferences.

2 Literature

2.1 Primary

Falk and Szech (2013) compared two main treatments, one featuring a singular choice: money versus the life of a mouse and the other enabling participants to repeatedly bargain with randomly selected peers with the externality that if an agreement was reached, a mouse would be killed. The authors found that the implicit value of the life of a mouse decreased over time and attributed this effect mainly to the existence of the market they created.

In contrast, Sandel (2013) cites multiple incidents in which introducing a price already shapes behavior by crowding out non-market norms. In some cases, because a fine is regarded as a fee, counterintuitively increasing the demand for action it tries to deter. In others by simply putting a price on behavior that was previously regarded a social norm or civic duty, thereby devaluing the action that was supposed to be incentivized. Thus, our basic idea is to look at a setting in which both aspects, that is market interaction and "putting a price on an action", are varied.

In their seminal paper, Krupka and Weber (2013) introduced a mechanism to measure norms or more specifically social appropriateness by incentivizing participants to identify these for a set of actions. Subjects were specially tasked with and paid for correctly identifying the prevailing social norms for specific economic games.

2.2 Secondary

Bartling, Fehr and Özdemir (2021) up the scale by changing the affected third party in the bargaining to a human being who will depending on the outcome or abstention of bargaining receive necessary medical care. By extending the non-market treatment to also encompass multiple periods, they find a similar decrease of actions associated with moral norms over time in both treatments. Bland and Nikiforakis (2015) however find that the magnitude of an externality strongly shapes behavior therefore implying that both studies cannot be directly compared. Nonetheless, the critique of Bartling et al. 2021 that the factor time is not controlled for in Falk and Szech's (2013) experiment still holds true.

Most prior research for instance Bartling et al. (2015) and Rode (2008) show a stabilization of moral behavior within 5 periods. Given the challenges involved with conducting experiments online this enables us to minimize both the burden on participants as well as the necessary monetary investment to remunerate them for their time and actions. Bartling et al. (2019) find a high deviation from the rational outcome if the harm done to another party can be framed as "personal" instead of "statistical". They also speculate that the higher proportion of pro social behavior is partly caused by the employed market mechanism. In the case of the seminal publication by Plott (1983) double auctions and in their own experiment posted offers. Given the split responsibility in double auctions this could also be an example of "moral wiggle room" as described in Dana (2007).

Kirchler (2016) does not compare different market mechanisms directly but analyzes the effects of for example anonymity and punishment on individual choice and double auction treatments separately, giving valuable insight into possible designs for this thesis. Irlenbush

(2019) provides a review on the literature on framing in related studies and in addition finds that referring to buyers and sellers by their respective role decreases socially responsible behavior. Ziegler (2020) finds that allowing participants to trade in multiple units each round leads to massive decrease in moral behavior compared to single unit treatments. The authors attribute this additional change to a replacement effect, i.e., that participants justify their actions by reasoning that the actions causing externalities would instead be exploited by others.

While this presents a promising avenue for further research, the additional complexity required is beyond the scope of a master thesis and likely a better fit for a more controlled laboratory environment.

3 Study Design

3.1 General Design

"Flygskam" (see e.g., Coffey (2019)) or as it is also known "flight shame" is an example of a current trend in which adherents voluntarily abstain from self-beneficial behavior to minimize their impact on the environment. I.e., by avoiding destinations that require flying and using alternative modes of transportation. The question this thesis tries to address in this context is the following: can introducing a price, for instance a fine on flights devalue the existing behavior of self-imposed constraint? Or is putting a price on something just one of the elements required to undermine moral behavior?

The core idea is to design and implement an experiment in which both "market interaction" and "putting a price on" are varied in a 2 x 2 experimental design. This framework is based on a design proposed by Christian Koch.

It features 3 types of participants, all of which are required to perform a real effort task prior to taking part in the game itself. Type Z does not interact but only observes others behavior. Its income other than the show-up fee is determined by the actions of the other players. Type

Y can increase their own income by engaging in an action that negatively affects those of type Z. Type X is similar but features a much higher payoff for engaging in the “bad” behavior.

To summarize, the two active participants have two options. One involves a negative externality on others. The other does not but is less beneficial. From our study of the relevant literature, we observe that many people may avoid actions similar to the former option. To this baseline, one can introduce treatments that feature “putting a price” or “market interaction”: The “price” treatment includes a tax on the behavior with an externality that is lost and not returned to the participants. This is inspired by Sandel (2013). While the third would exhibit both prices and market interaction, by allowing participants to “trade” their right to conduct the “bad” action similar to Falk and Szech (2013).

To keep the scope of the initial design and therefore the extent of the experiments to a manageable size, only the baseline game and a single additional treatment “putting a price“ were implemented and conducted with online participants.

3.2 Conducting Experiments Online

A key element of this master thesis is the implementation of trial sessions of the experimental designs online using Lioness as a platform and Prolific (www.prolific.co) for recruiting participants. This environment is especially appropriate for our project as it involves only limited interaction between individual players and is simple enough to not require extended briefings or additional aid during each treatment.

Giamattei et al. (2020) created an online framework “Lioness” that enables economists and researchers from related fields to conduct experiments both on and offline on servers or cloud services under their own control while permitting them to interface with common click worker platforms, for instance amazon’s MTurk to recruit participants.

Arechar et al. (2017), conducted already well established and researched public good games both in a laboratory environment as well as with online recruits, found similar behavioral

patterns and came to the conclusion that experiments can be replicated online. One of the key advantages of the online environment is the ability to repeatedly recruit new participants while excluding those who might have been biased by prior treatments and reducing the likelihood of information about the experimental framework spreading if the individual treatments are not conducted within a short timeframe.

This could have for instance enabled to use the “baseline” treatment to calibrate the payoffs prior to conducting the whole experiment. While the results of this first round of experiments did not indicate the need to change the experimental parameters, this allowed us to review not only the quantitative data but also written feedback from the first wave participants prior to scaling up to the full sample size.

Arechar et al. (2017) however also point to and analyze one of the key concerns in online experiments: Participant dropout. Especially in multiplayer games, this can massively impact data quality, because a single mid-game exit affects multiple participants and therefore their behavior. Other than the game design itself, drop-out rates according to the authors are mainly affected by waiting times between rounds and the total time spent in the experiment.

To maximize the amount of valid data gained for the money invested, we chose to minimize the time participants must wait for others by keeping the number of rounds low and aiming to keep the total time including a questionnaire well below 15 minutes.

3.3 Game Design

3.3.1 General Incentives

Approved Participants received a participation fee of 1.5 £ for an estimated completion time of 10 minutes which equals an hourly rate of 9 pounds prior to consenting to taking part in the trial. Any additional rewards during the trial were presented in points with an exchange rate of one point being the equivalent of 2 pence. Participants were made aware of this exchange rate at the beginning of the experiment.

3.3.2 Baseline

Action X	Action Y	Payoff X	Payoff Y	Payoff Z
1	1	50	40	5
1	2	50	35	15
2	1	35	40	15
2	2	35	35	25

Table 1: Payoffs in points, baseline trial

Both types of active participants receive a higher payoff for choosing the "immoral" action 1 that causes an externality to their matched counterpart of type Z. In addition, players of type X receive a higher reward for this "bad" behavior: 50 points compared to Y's 40 points.

In contrast, by individually choosing the action 2, they avoid causing their respective negative impact on the payoff of players of type Z.

3.3.3 Treatment

Action X	Action Y	Earnings X	Tax Due X	Earnings Y	Tax Due Y	Earnings Z	Reparation Due Z
1	1	52.5	2.5	42.5	2.5	0	5
1	2	52.5	2.5	35	0	12.5	2.5
2	1	35	0	42.5	2.5	12.5	2.5
2	2	35	0	35	0	25	0

Table 2: Payoffs in points, treatment trial

The treatment trial was designed such that the actual payoffs and therefore the monetary incentives do not change from the baseline trial.

Given a large enough sample size, any differences in behavior by our subjects must then be caused by our 2 changes. Namely the introduction of our stimulus with the framing device "tax" and its value.

3.3.4 Krupka Weber

Participants were tasked with correctly identifying the prevailing view of all involved subjects on the social appropriateness of the externality causing action. Specifically, they were asked

to consider the appropriateness separately for type X and Y and were awarded 10 points for each correct norm identification.

3.4 Hypothesis

Hypothesis 1: The game elicits different patterns of behavior depending on player type.

This serves the purpose of validating the base game and that the differences in payoff and the scale of the externality are able to influence behavior.

Hypothesis 2: Behavior of participants differs between the baseline and treatment designs.

One of the core ideas of this thesis is that the introduction of an externally set price has an impact on action even if there is no change in the actual monetary incentives.

A decrease in the rate of immoral behavior would be in line with the established literature on taxation. Results similar to those reported by Sandel (2013)on the other hand, i.e., a significant increase, would confirm the ability of this framework and specifically the chosen parameters to demonstrate crowding out of internal values by "putting a prize" on the connected behavior.

Hypothesis 3: Price framing does not only change the observed behavior but also the participants' estimate of the prevailing social norms of the appropriateness of the action with a negative externality.

Finally, we propose that external stimuli of this type affect not just actions but can shift social values and norms.

3.5 Recruitment Platform: Prolific

Prolific is an online participant recruitment and payment processing platform that offers its services to scientific researchers but also to commercial, non-commercial, and governmental clients.

It’s three key differentiators to traditional click worker platforms for example Amazon’s MTurk enable reproducible experiments while avoiding the pitfalls normally associated with online recruitment.

Firstly, the company offers a vetted pool of candidates which minimizes the number of bots participating in studies compared to alternative providers. It also reduces misrepresentation of for instance location, citizenship, age etc. by candidates.

While it is of course still possible to transfer accounts to other individuals and even operators of bot farms, this can only happen on a one-to-one basis further mitigating any impact on data quality and sample integrity.

Secondly Prolific enforces a living hourly wage for participants of 6 GBP as of September 2023. This has the benefit of incentivizing candidates to submit valid responses to avoid nonpayment due to failing attention checks. Additionally , if the average completion time of an experiment is larger than advertised to the participants prior to recruitment, Prolific demands a proportional increase in the guaranteed participation fee for the affected trial. Notably this also means that any in experiment monetary economic incentives have to be designed with consideration to the comparably high participation fee that can also vary depending on actual the completion time of each trial run.

To mitigate this issue for this thesis, multiple unreported treatment runs were conducted with friends and family from diverse backgrounds to calibrate the advertised time frame and payment prior to the randomly controlled trials with recruited and incentivized participants.

Finally, Prolific features extensive filters for candidate pools, not only for personal, socio-economic and demographic attributes but also prior exposure to other experiments by the same team of researchers. The recruitment tool can furthermore enforce balancing of samples during live recruitment and for large samples even representativeness.

In Addition, the paper introducing Prolific and its features and benefits to the scientific community Palan, Schitter 2018 is generally well received with 2241 citations as of 07.09.2023.

3.6 Implementation and Participants' Experience

3.6.1 Experimental Software Stack and Server

After being recruited via Prolific participants entered the server hosting the experiment via individual links that connected their actions during the experiment to their anonymous prolific account to enable individual and outcome-based payment.

This server was deployed on a Google cloud instance configured to host a LAMP Stack and a Lioness Session as the frontend for both user interaction and supervision.

3.6.2 Challenges with Real Time Matching on Prolific

On their help page, Prolific itself advises a cautionary approach to real time interaction between participants.

This is due to two main factors, namely, the size of the recruitment pool and the method used to attract subjects to specific experiments and more importantly for Prolific, retain them as active members on the platform.

Problems stemming from the size of the recruitment pool where mitigated, by minimizing the number of restrictions to only first language and country of residence and balancing by a single characteristic, namely gender. In addition, the two treatments were conducted during UK business hours in the time frame Prolific reported their candidates to be the most active in accepting and starting submissions.

The second issue arises from the method Prolific employs to attract participants to individual experiments. To avoid irritation and therefore a high turnover rate by vetted Prolific participants, the platform solicits experiments in waves and only to a subset of their available pool of candidates. The size of each wave also notably depends on the number of unclaimed spots in an experiment causing a decrease in successful matches as a trial draws to an end.

While the first problem was apparent during the design of the experiments and precautions were taken to minimize its impact, the second issue only came to attention during the supervision of the two trials.

While we can offer no insights into mitigating this concern, to avoid sounding overly negative we have to state that between 75 and 90 percent of all participants were successfully matched to other players and at a cost per person of less than 3.5 British pounds which is well below on-premise experiments with a similar subject pool.

3.6.3 General Instructions

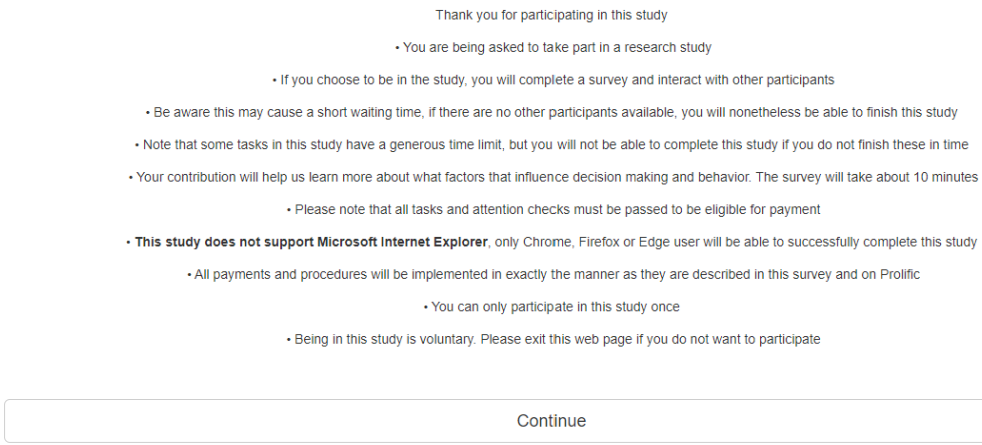


Figure 1: Landing Page

After arriving at the dedicated server hosting the experimental environment, participants were first presented with a message informing them of the expected time frame, the existence of attention checks and control question and that the study required both individual as well as group decisions.

In addition, players were also told that their contribution would aid in understanding what factors influence decision making and behavior. This was included to give a further non-monetary incentive to complete the experiment.

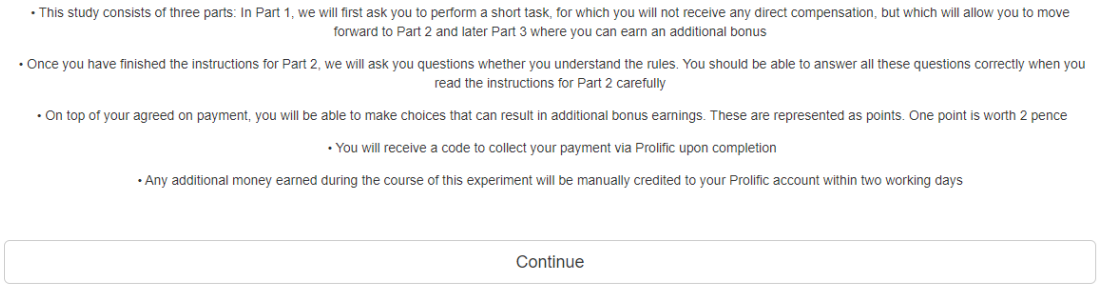


Figure 2: General Instructions

On this page participants were introduced to idea of the real effort task and that while the successful completion of this part would have no inherent value it would allow them to progress to later parts of the experiment that would enable them to earn an additional bonus.

This Task however notably served two purposes. Not only to prime participants and create a sense of relatability with their cohort, but also by randomly creating the puzzle for each player, simple automated survey solvers, similar to those often encountered on Amazon's MTurk, were impeded. Together with an average total run time of each treatment of less than 2 hours we are quite certain that all experimental subjects were human and not automated bots.

3.6.4 Real Effort Task

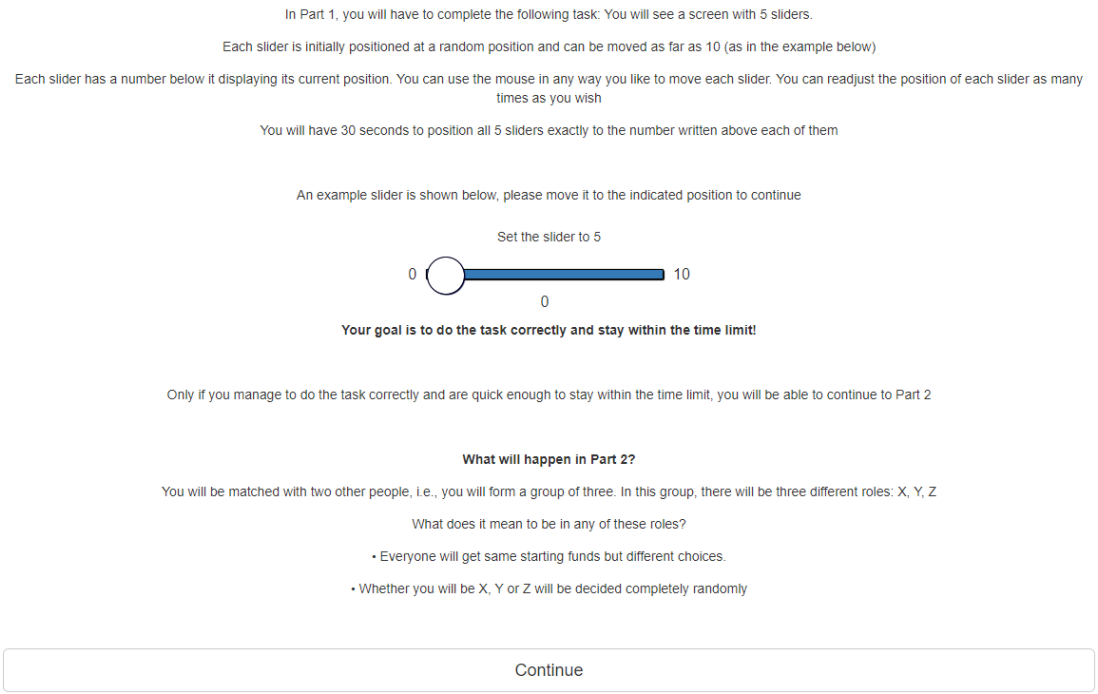


Figure 3: Introduction Real Effort Task

Afterwards the real effort task consisting of a simple slider and a target number was introduced in detail and an example displayed for participants to practice with.

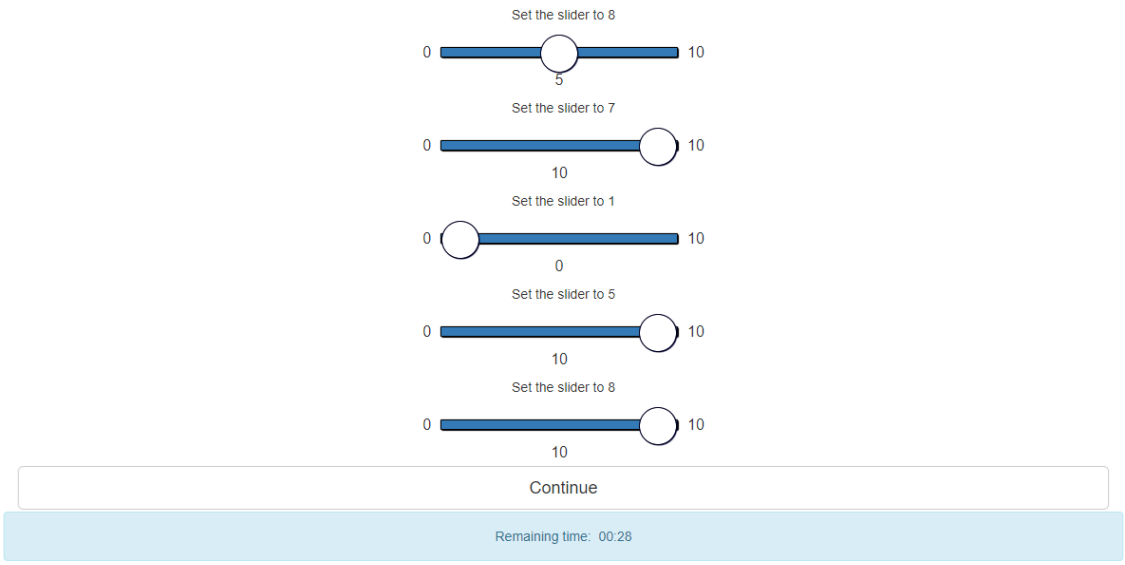


Figure 4: Real effort Task

To reduce the rate of participant dropout and generally the expected time investment for players only five slider tasks had to be solved within a time limit of 30 seconds.

Participants who failed this task were however directly removed from the experiment and shown Figure 22: Exit Screen Time-Out

3.6.5 Instructions: Stage Game

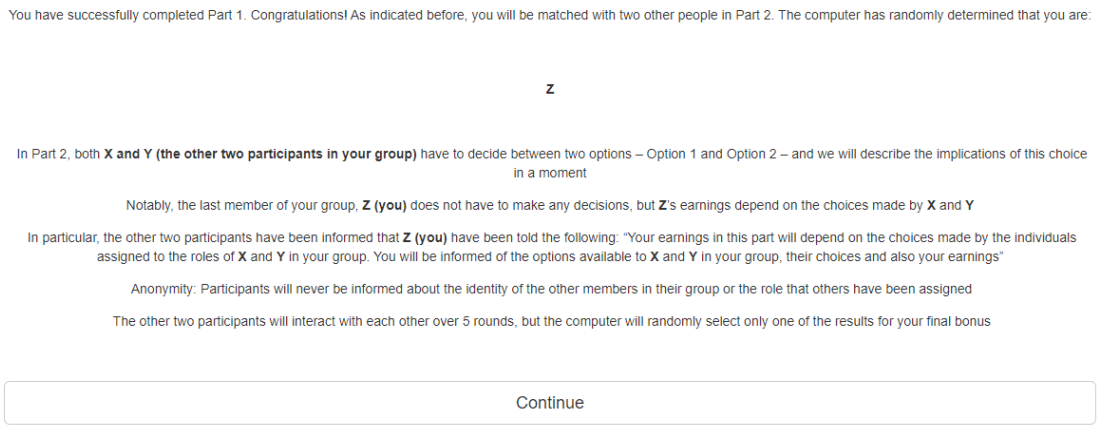


Figure 5: Introduction Part 2 for Type Z

After successfully completing the real effort task players were informed that they had been randomly assigned a role and the implications of said role. Screenshots for participants of Type

X and Y can be found under 6.1 Additional Screens in the appendix. For brevity's sake all subsequent figures are handled in a similar fashion

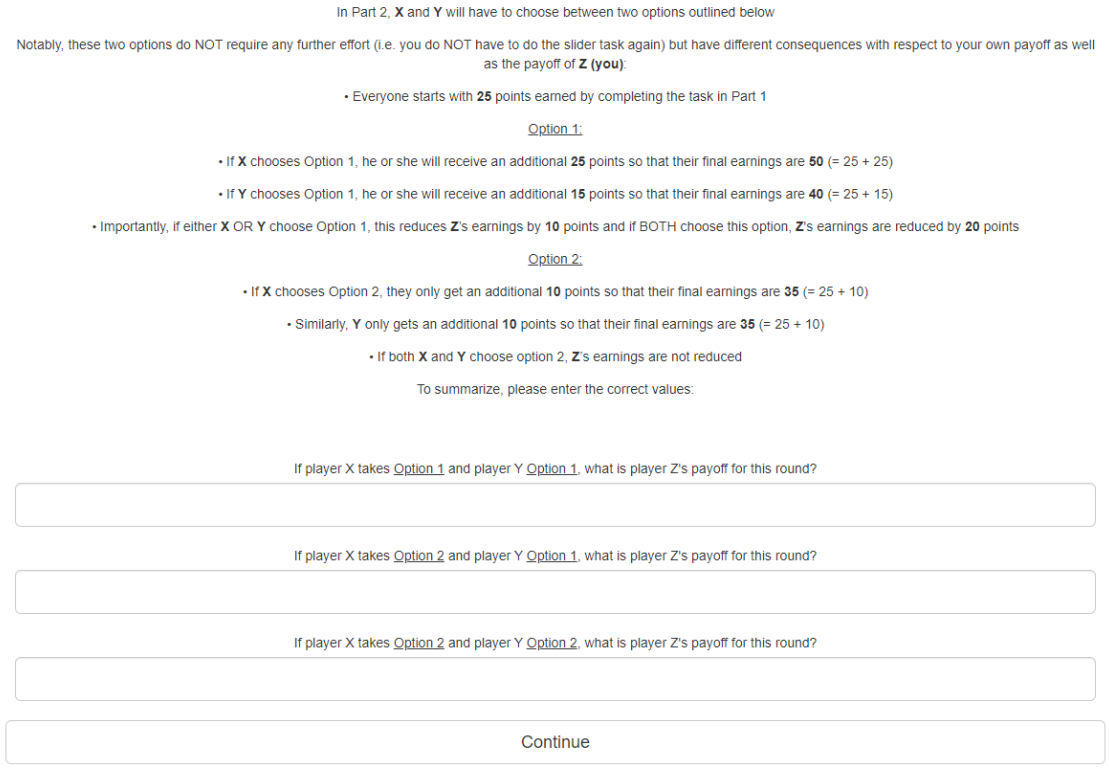


Figure 6: Control Questions for Type Z

To assure that participants fully understood the experiment they were tasked to calculate the payoff for players of type Z under 3 different scenarios. Type Z was chosen to reinforce the idea of actively taking away from another participants by choosing the action with an externality.

After this screen players of type Z were directly moved to Figure 12: Krupka-Weber Survey while those of type X and Y were matched on a one-to-one basis and tasked with playing the stage game for 5 rounds.

3.6.6 Stage Game

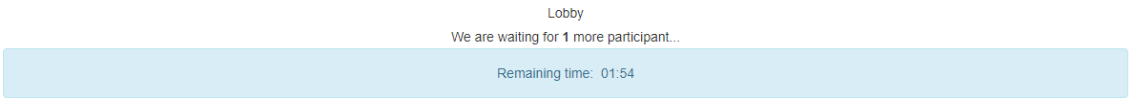


Figure 7: Lobby Stage Game

While in the lobby and waiting to be matched, participants of type X and Y were displayed a life countdown starting at 2 minutes to keep them engaged and avoid unnecessary dropout.

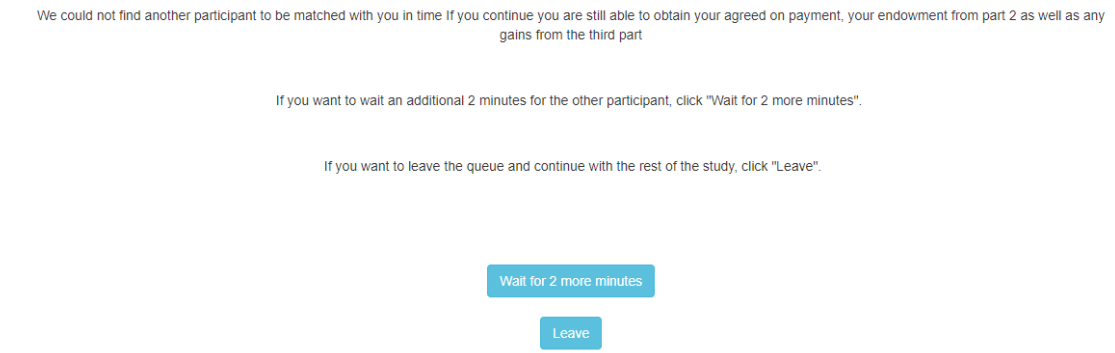


Figure 8: Opt-out Stage Game

If a match could not be assigned during this timeframe, players in the lobby were given the option to directly continue to Figure 12: Krupka-Weber Survey. They were also informed that they would be able to keep their endowment as well as any additional gains from the third part of the experiment.

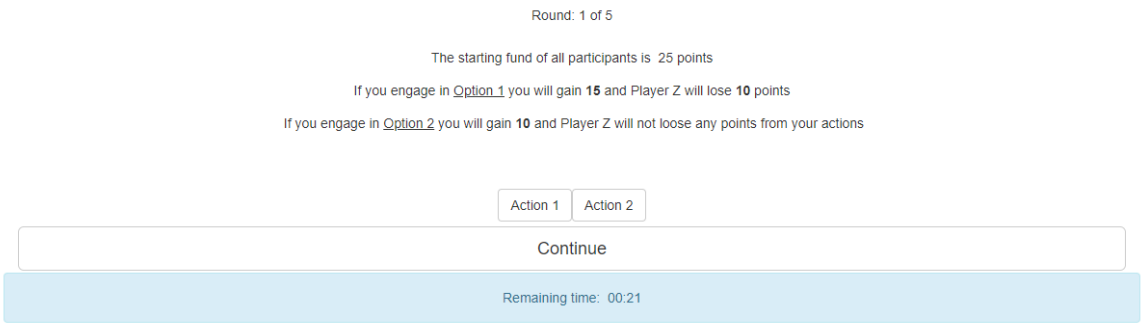


Figure 9: Choice Screen for Type Y

During each of the 5 rounds of the stage game participants were informed of their possible choices and their respective impact on themselves and the third matched player of type Z.

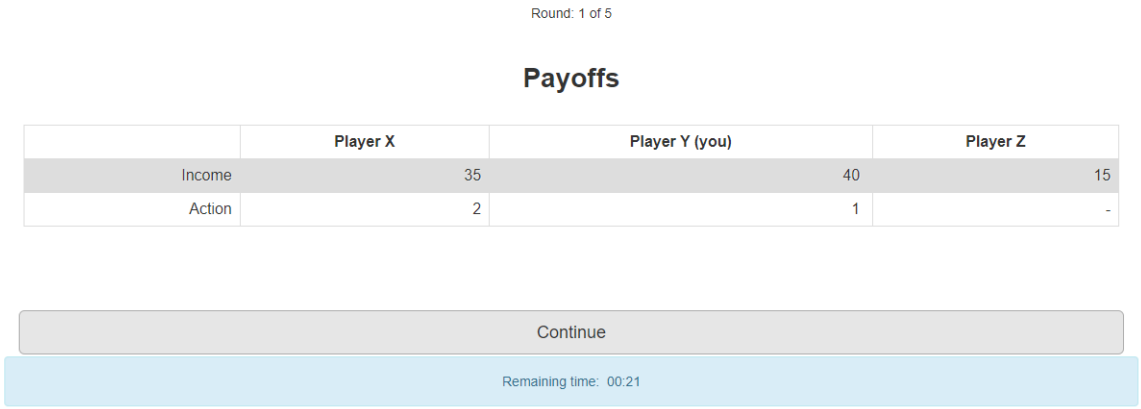


Figure 10: Payoff Screen for Type Y

After both active players (type X and Y) made their choices, the payoffs for this round were calculated and displayed to the active players.

Dear participant,

Thank you for taking part in our study

Your matched participant Y played Option 1 for 5 rounds, while player X chose 1,1,2,1,2 respectively

Round number 1 was randomly chosen for payout and has already been credited to your account

Thanks again and have a nice day!

Figure 11: Payoff Information for Type Z

Players of type Z were informed of their matched players actions and their payoff after the fact via Prolific's messaging service.

3.6.7 Krupka Weber Norm Identification

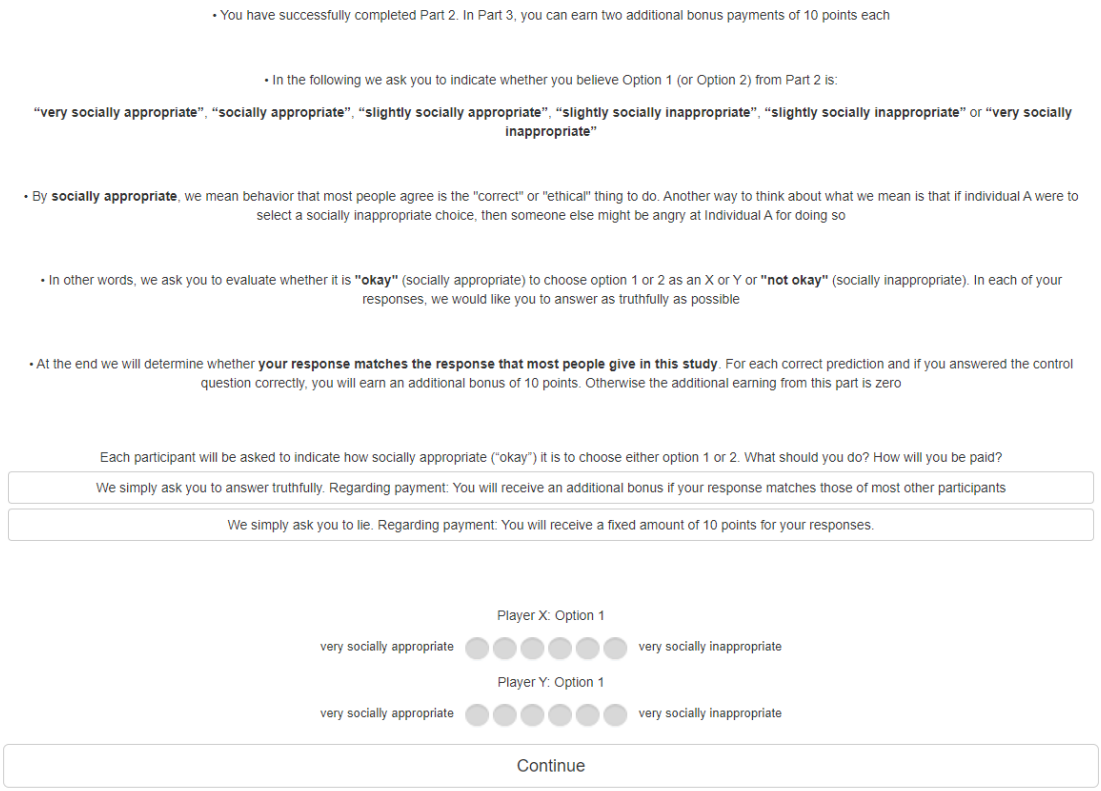


Figure 12: Krupka-Weber Survey

Regardless of their choices, players of all 3 types who successfully completed the first two parts of the experiment were given a short introduction into a Krupka Weber 2013 style coordination game and its aim to elicit their expectation of the social norms of the other participants.

A single control question had to be answered correctly to be eligible for payment for an accurate prediction in the third part of the experiment.

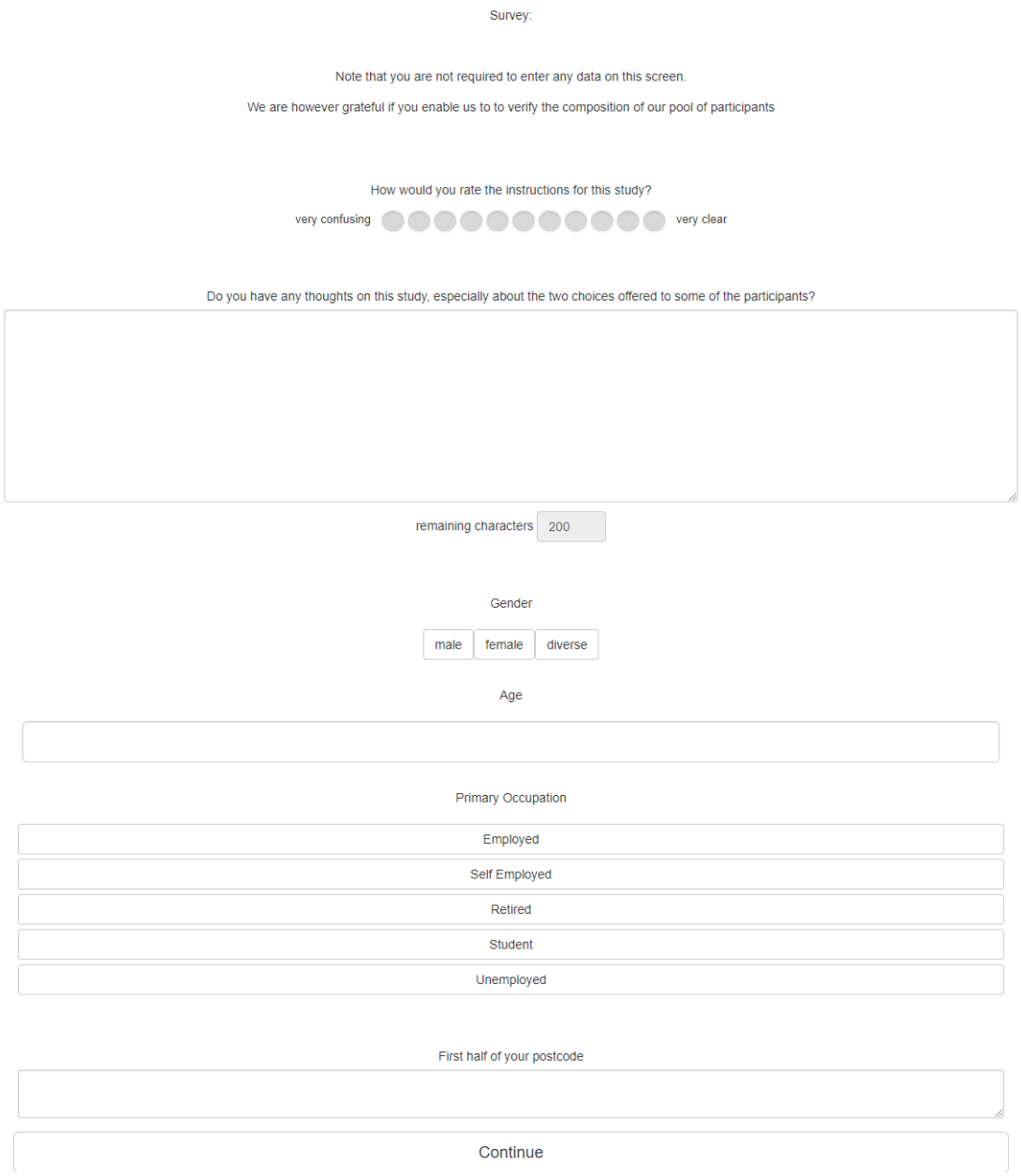


Figure 13: Post Experiment Survey

On the penultimate screen, participants had the option to verify their personal characteristics in accordance with the standards set by Prolific's guidelines for the collection of personal data.

They were also made aware of the fact that this was optional but could be used to verify parts of the study.

This screen also enabled the participant to both give numerical feedback on the perceived quality of the experiments but also to give a short comment on the experiment as a whole.

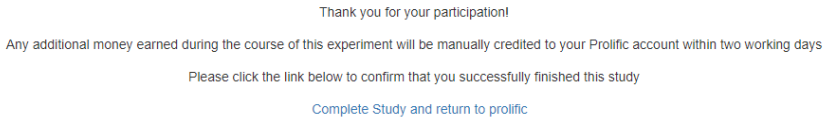


Figure 14: Return Screen

On the final screen the participants were presented with a link that included their personalized ID which allowed us to connect Prolific User IDs and Lioness Experiment IDs thereby enabling individual payment as well as direct messaging.

4 Results

4.1 Participants

	Participants	X matched	Y matched	Z
Baseline	60	18	17	18
Treatment	58	16	15	18

Table 3: Participants by Type

All approved participants successfully completed the real effort task, answered the control questions correctly, if matched played in 5 rounds of the stage game and filled out the Krupka Weber Survey.

In accordance with the terms of service of Prolific only these individuals received full payment. This entailed the participation fee, all additional rewards from both Krupka Weber and the stage game and finally an additional fee if the average completion time for their treatment exceed the advertised one.

Subjects of type X and Y who were matched but whose counterparts timed-out were informed of this fact, that their actions still had impact on the player of type Z and were tasked with selecting their choices for the remaining rounds. Participants of type Z that were affected by time-outs received this information in their post experiment summary and if their randomly chosen round was affected were rewarded as if the other player had chosen the action without an externality.

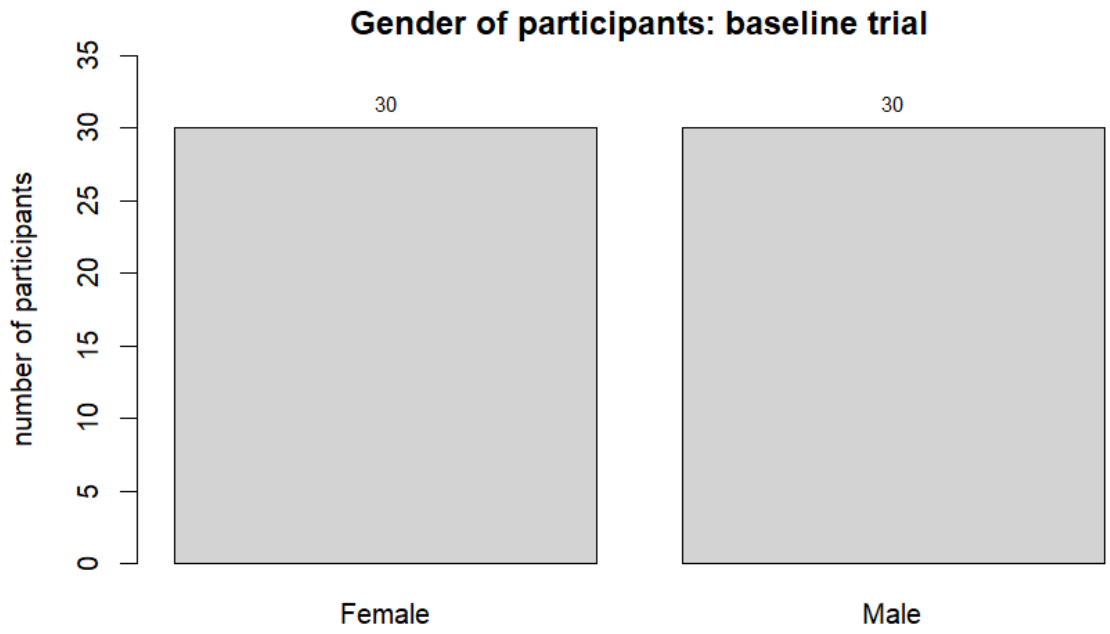
All subjects that Prolific recruited had English as their first language and were verified UK residents. In addition, Prolific gender balanced the sample from which it recruited participants for the study.

This restriction on residence was made to guarantee a fluent level of English in the candidate pool and to minimize additional sources of variance in the ex-ante norms. In addition, this enabled scheduling of the trial runs during business hours for both participants and experimenters.

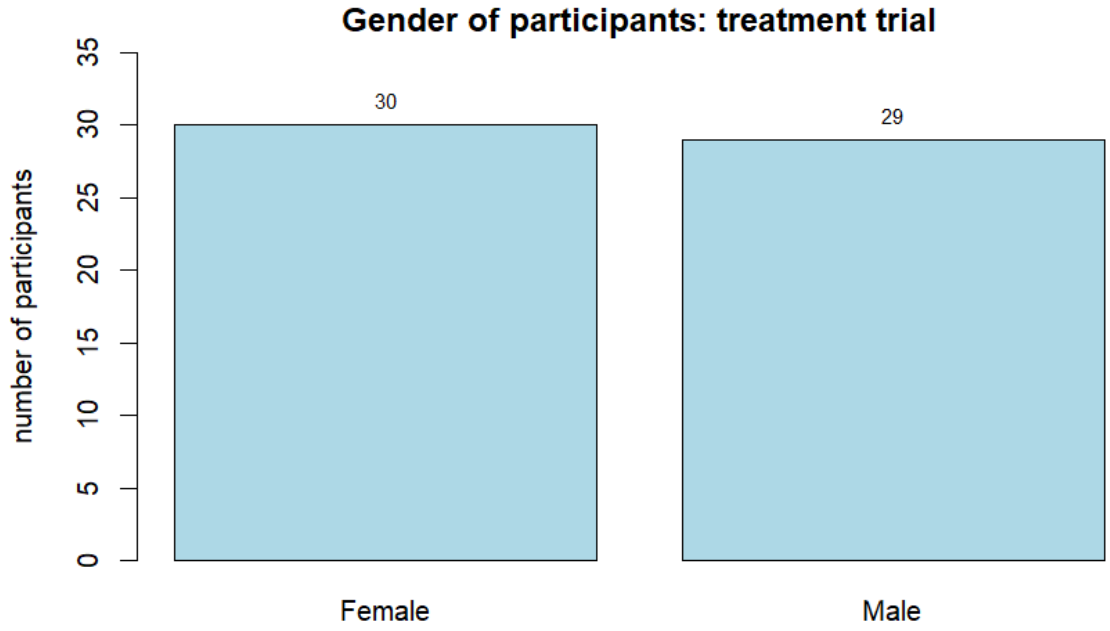
Finally, participants that took part in the baseline study were excluded from the treatment trial via Prolific's screening feature.

4.1.1 Characteristics

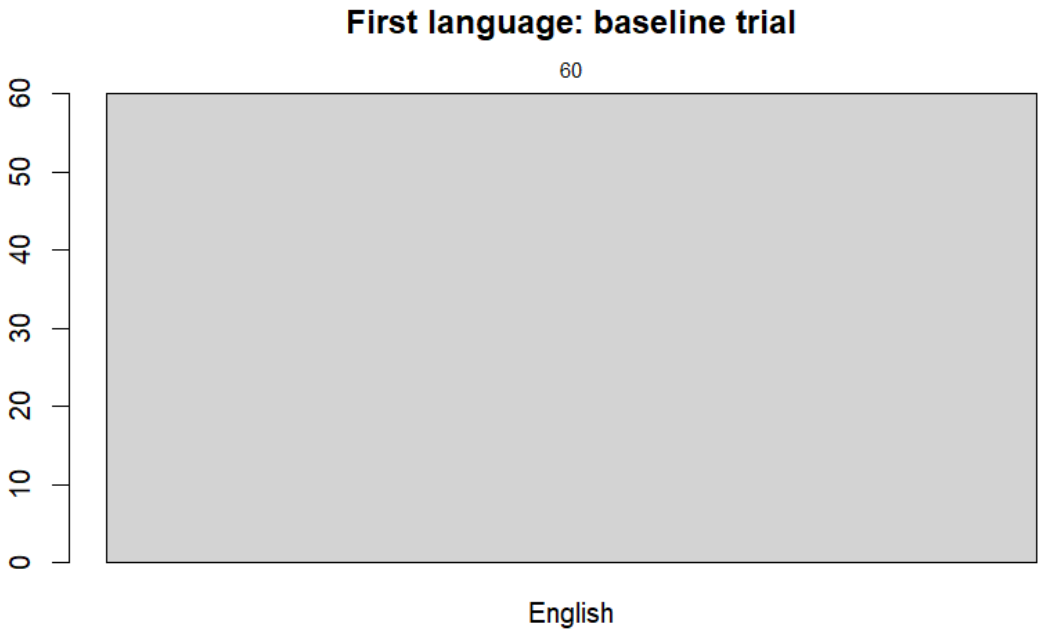
4.1.1.1 Personal



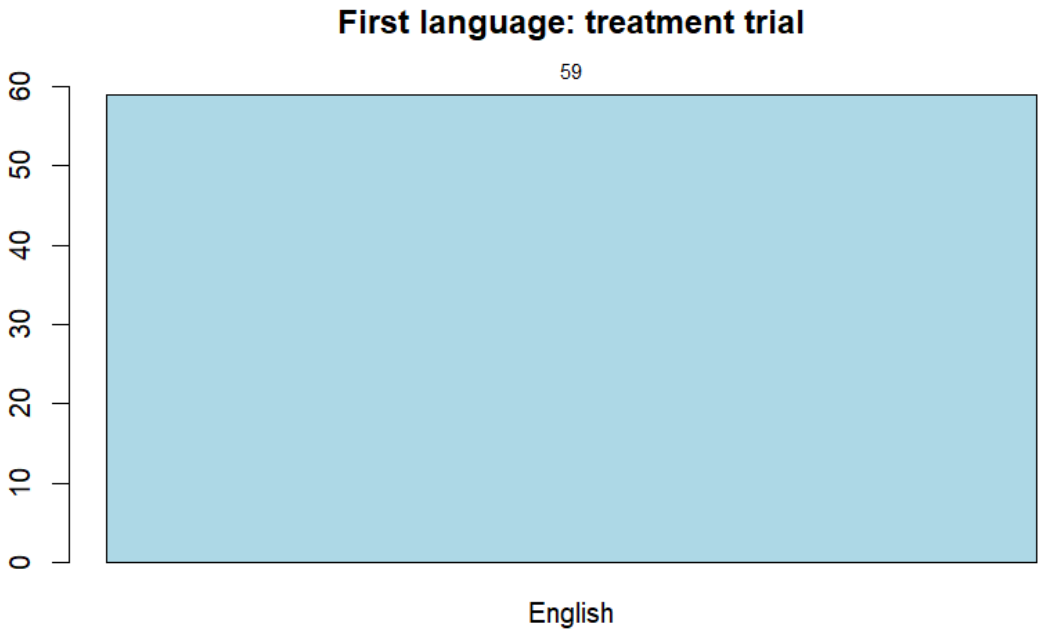
Participants' Characteristics 1: Gender, baseline trial



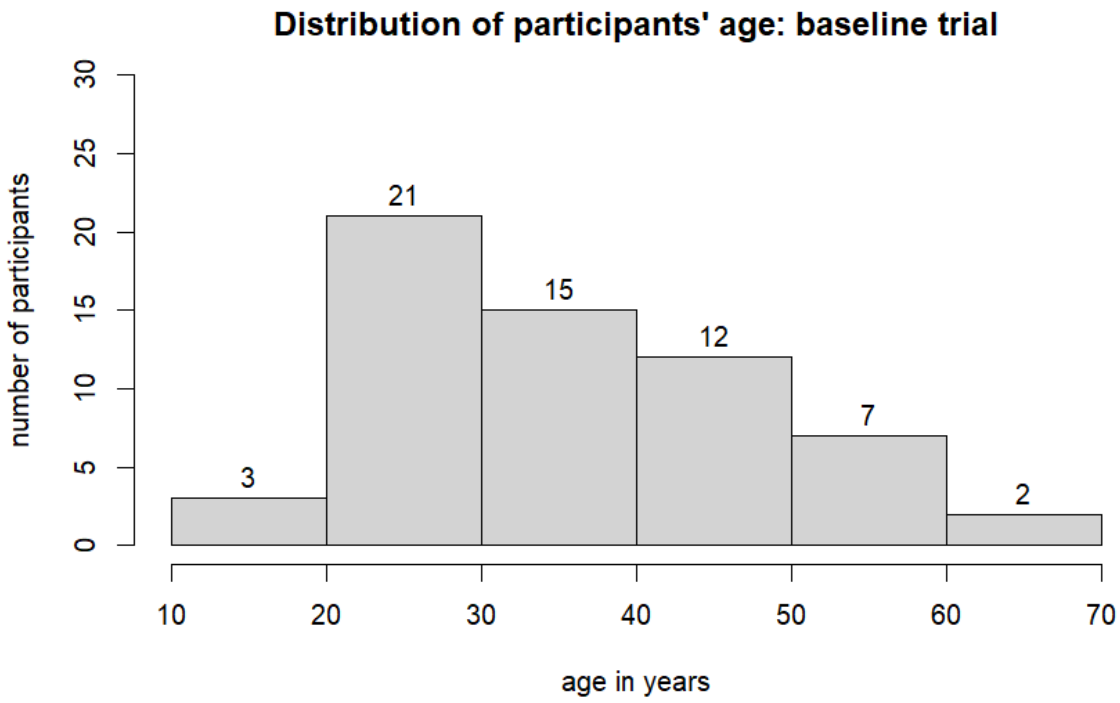
Participants' Characteristics 2: Gender, treatment trial



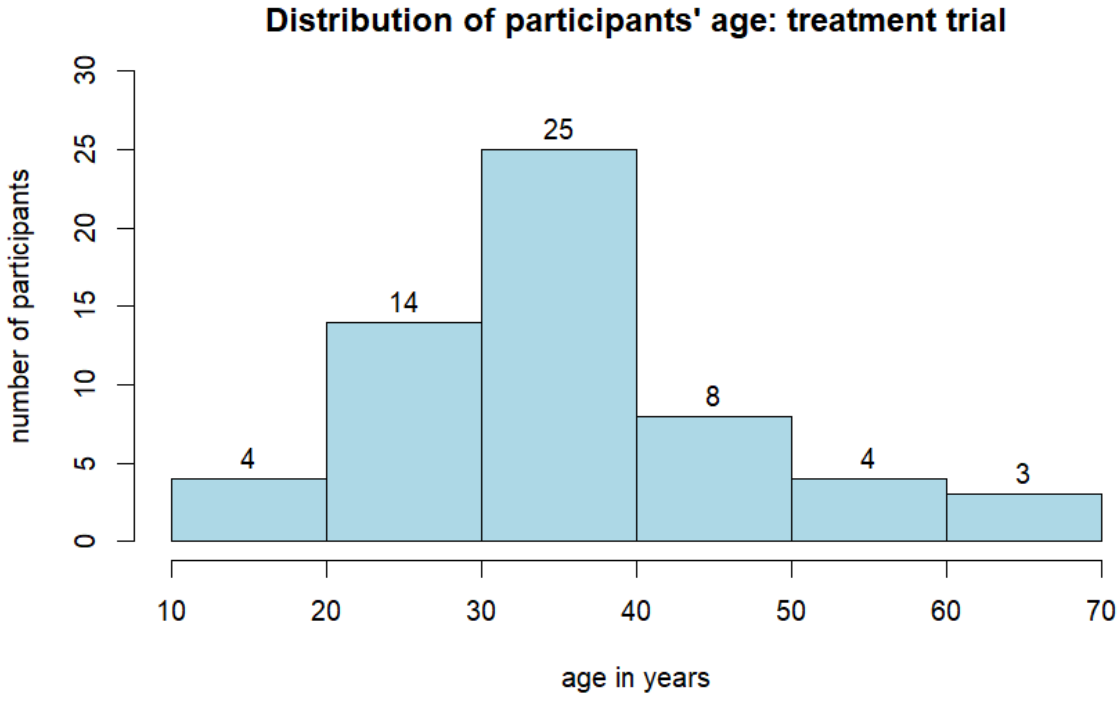
Participants' Characteristics 3: First Language, baseline trial



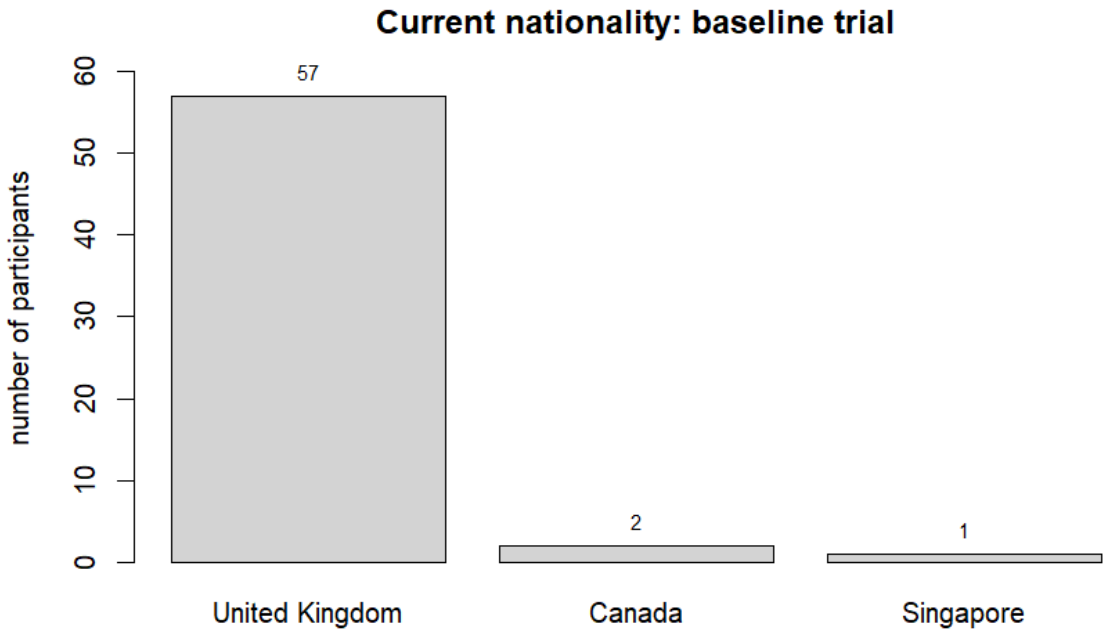
Participants' Characteristics 4: First Language, treatment trial



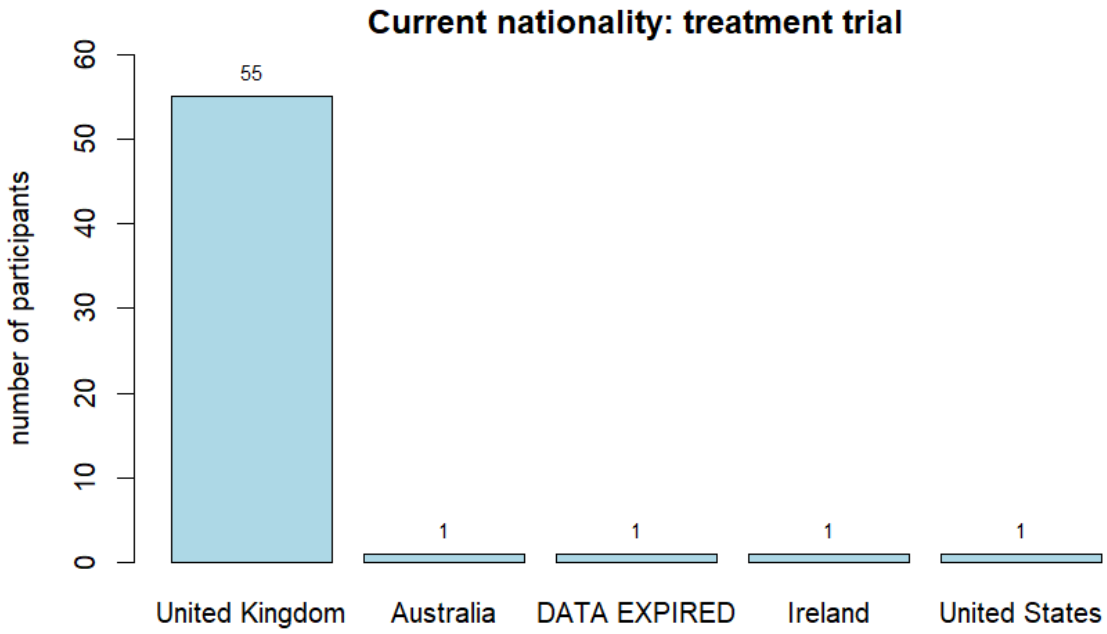
Participants' Characteristics 5: Age, baseline trial



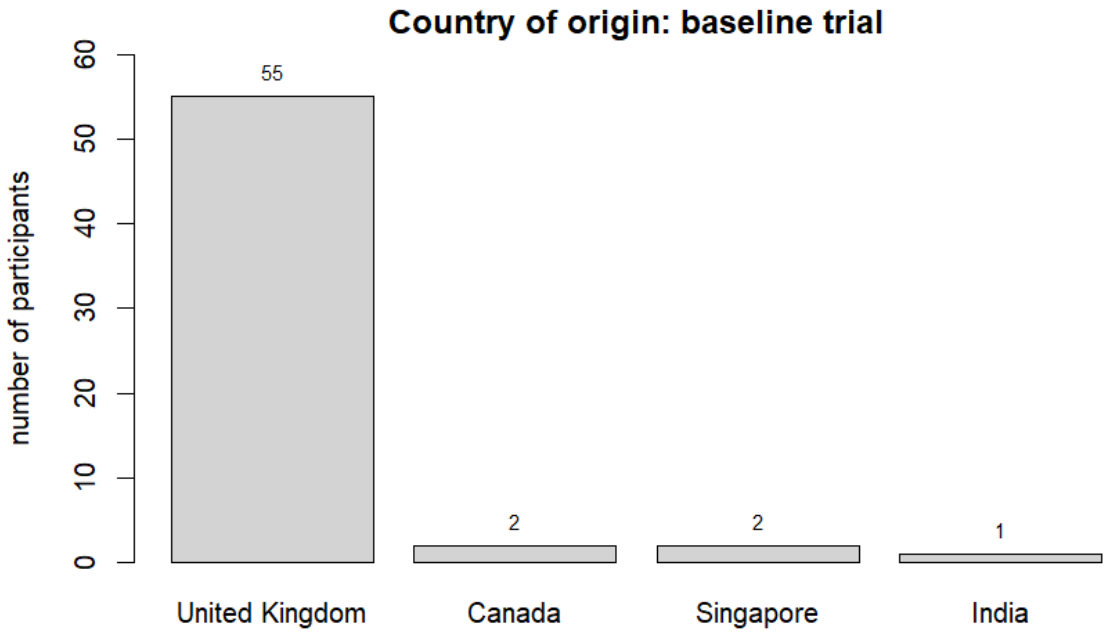
Participants' Characteristics 6: Age, treatment trial



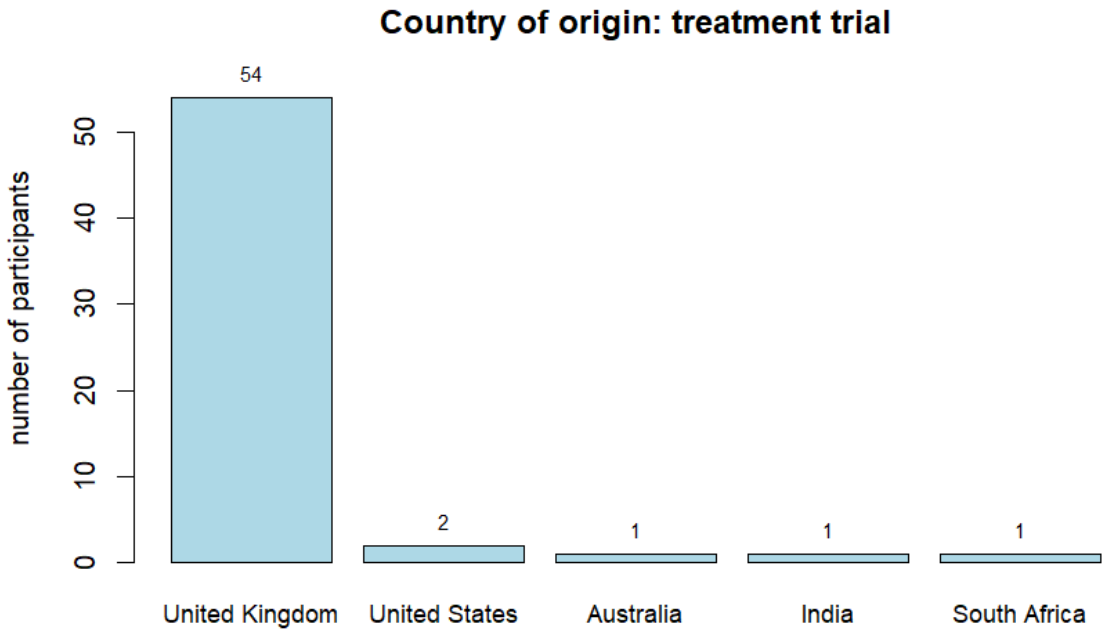
Participants' Characteristics 7: Current Nationality, baseline trial



Participants' Characteristics 8: Current Nationality, treatment trial

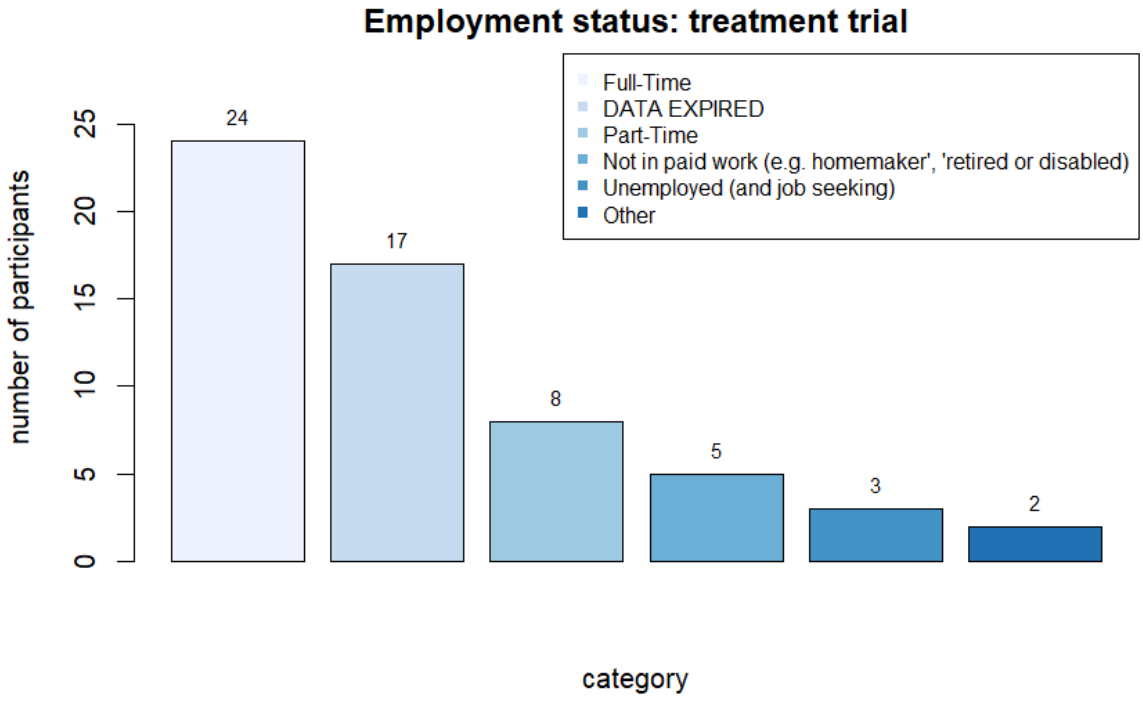


Participants' Characteristics 9: Country of Origin, baseline trial

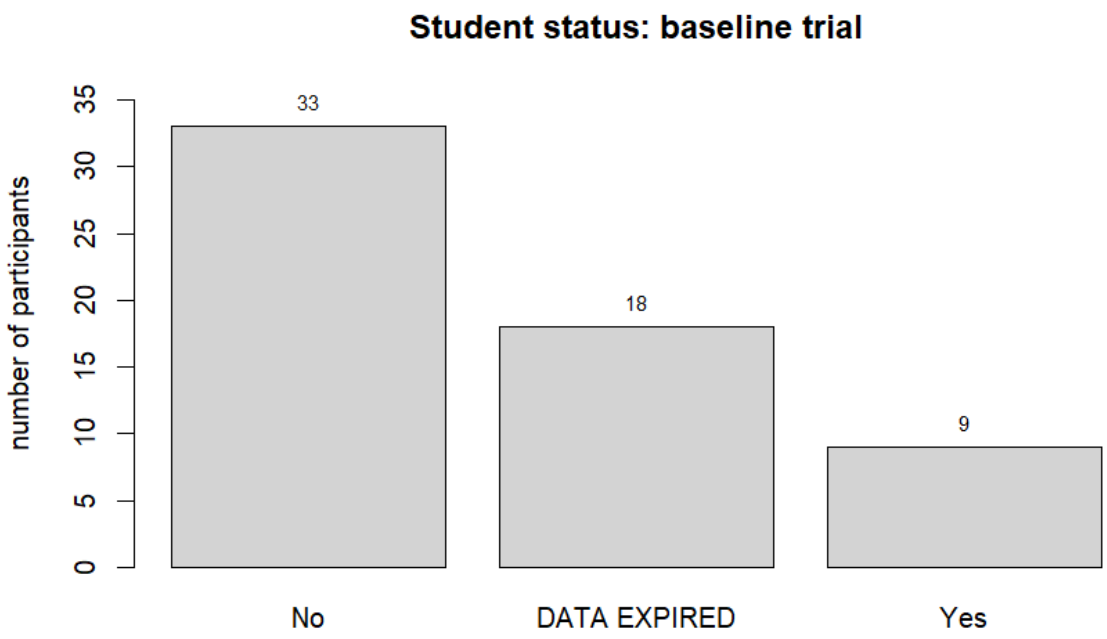


Participants' Characteristics 10: Country of Origin, treatment trial

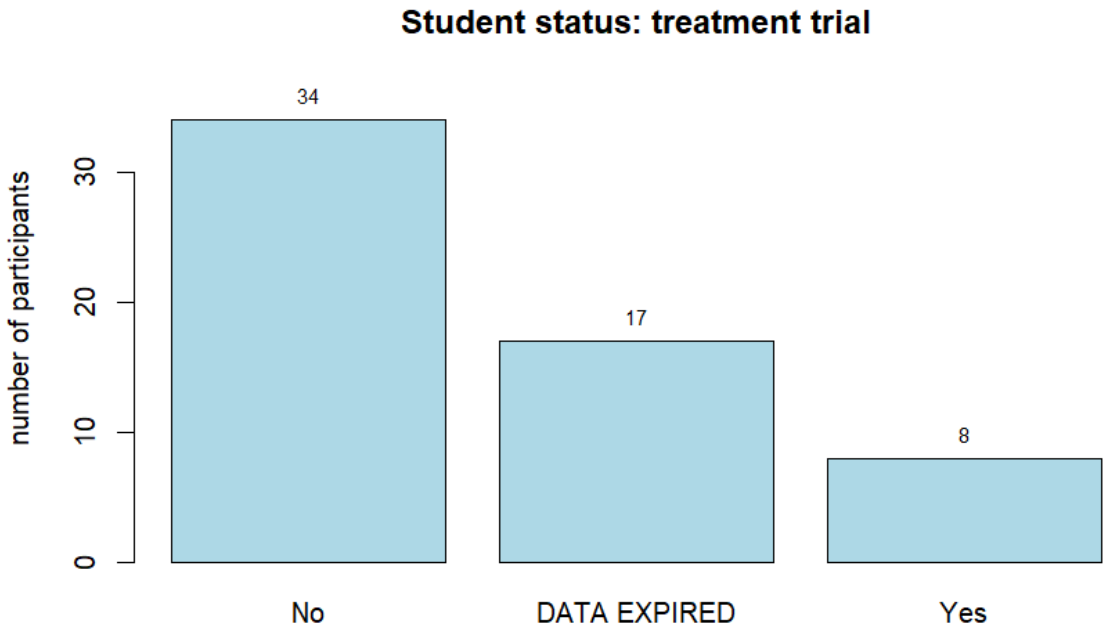
Participants' Characteristics 11: Employment Status, baseline trial



Participants' Characteristics 12:Employment Status, treatment trial



Participants' Characteristics 13: Student Status, baseline trial



Participants' Characteristics 14: Student Status, treatment trial

4.1.1.1.1 Discussion

Please note that the plots based on the characteristics of the treatment dataset supplied by Prolific feature one additional individual that was approved by accident and could not be

identified due to time constraints. This error does not impact the data collected via Lioness and therefore the analysis of the study specific data.

The characteristics reported by Prolific do not show cause for concern on a per trial level as well as in aggregate. Both first language and gender are within in the parameters set during screening and the age distributions for both studies feature an average age of 36 years and an overall similar pattern.

Counterintuitively, for a click worker platform the share of participants with a different country of origin 8,4 % is slightly below the UK average of 14% as reported by Rienzo and Vargas-Silva (2022). This can be explained by the same report: 5 out of the top 10 foreign countries of birth are non-majority English speaking.

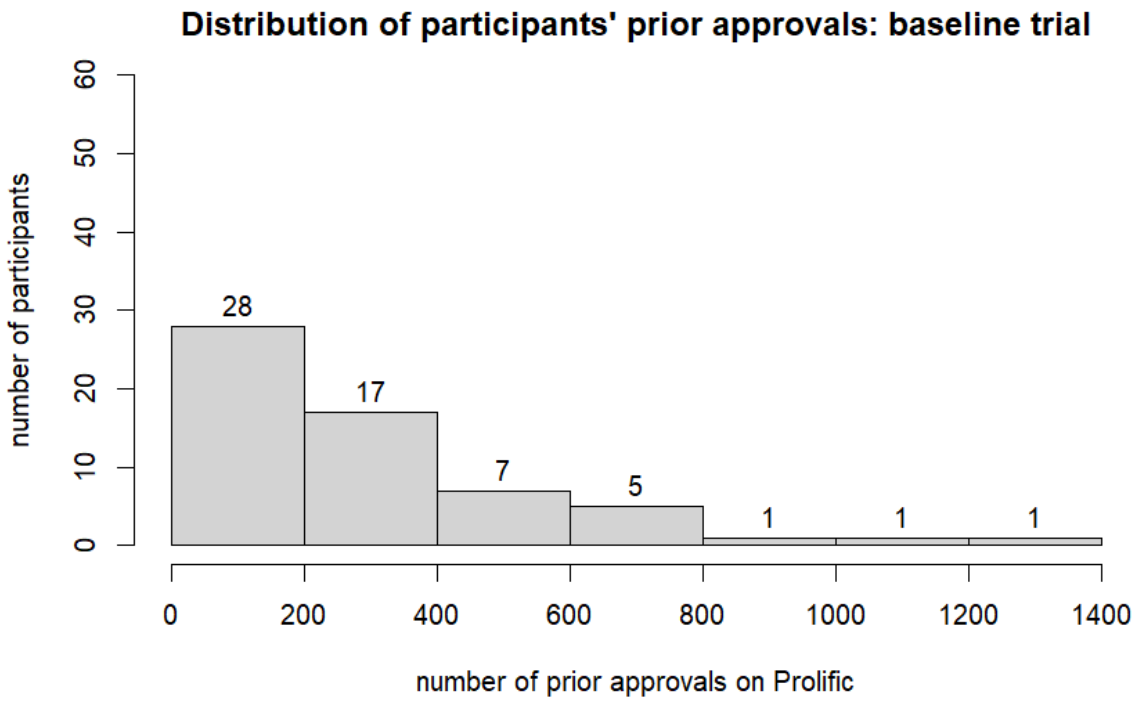
While this could be seen as a benefit for conducting complex studies for instance in regard to communicating difficult tasks to participants this restriction nonetheless biases the recruitment pool and can cause issues with experiments that require representative populations.

Also intriguing, is the share of full-time workers in the dataset, while the relative ease and short time investment necessary for taking part in individual studies and the push nature of recruitment through Prolific could explain the prevalence. The fact that both trials were conducted during business hours casts doubt on the validity of this part of the dataset.

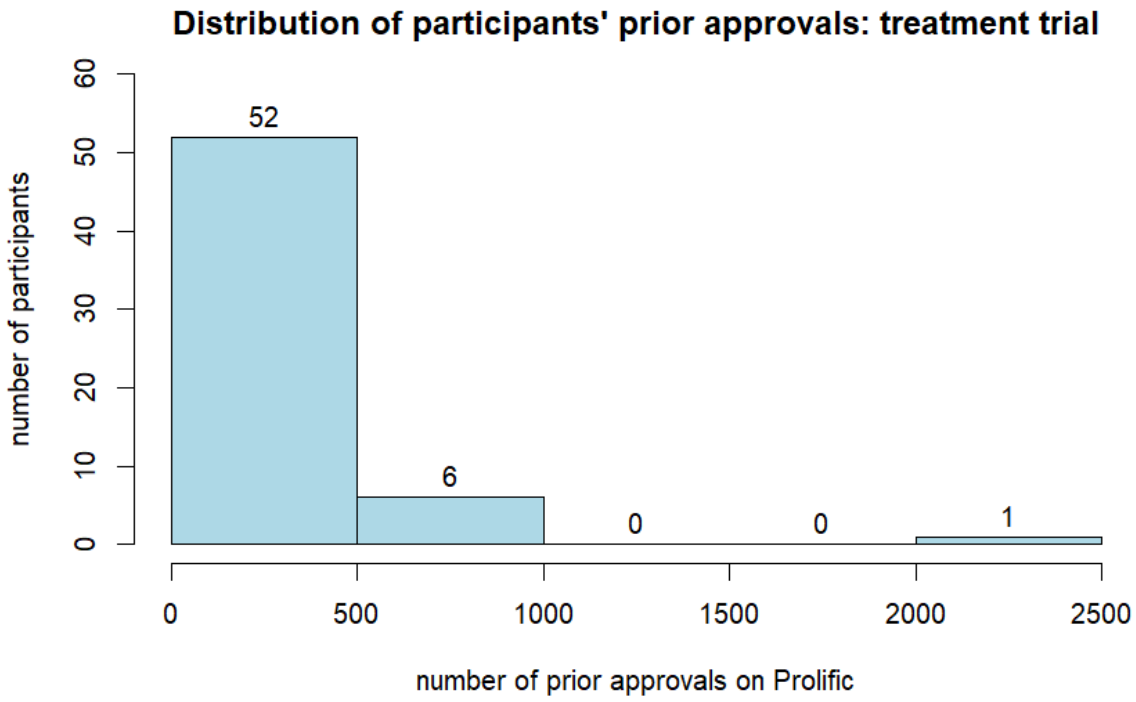
Together with the relatively high number of entries that did not meet Prolific's unreported own standard of quality for currency and were reported as "Data Expired" this implies that participants do not regularly update characteristics on the platform and instead correspond to their status when they originally joined. While this can be counteracted by including an additional survey in studies, the fact that any data gained through it can't be validated, requires additional comparison to Prolific's data if it were to be used in modelling. This approach is also recommended by the company.

Lastly and rather assuringly, the datasets include a relatively small share of students. This opens the door for comparative studies and validation of results of on-premise experiments with a comparatively higher share of participants enrolled in academia.

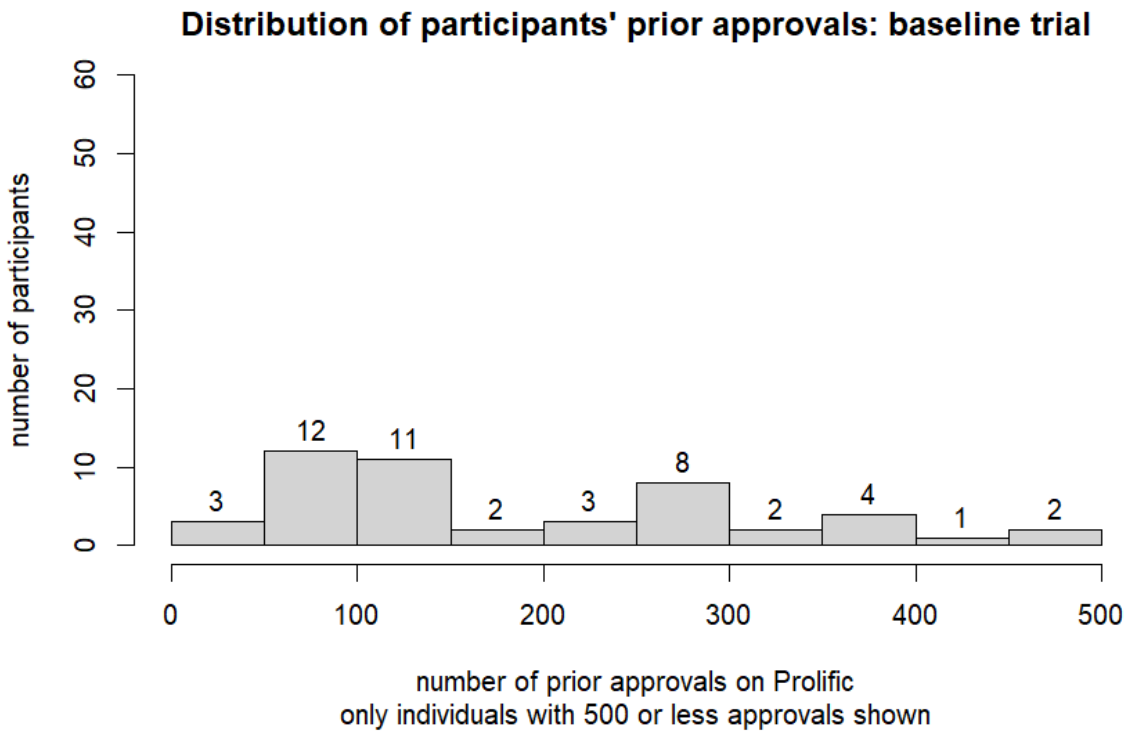
4.1.1.2 Specific to Prolific



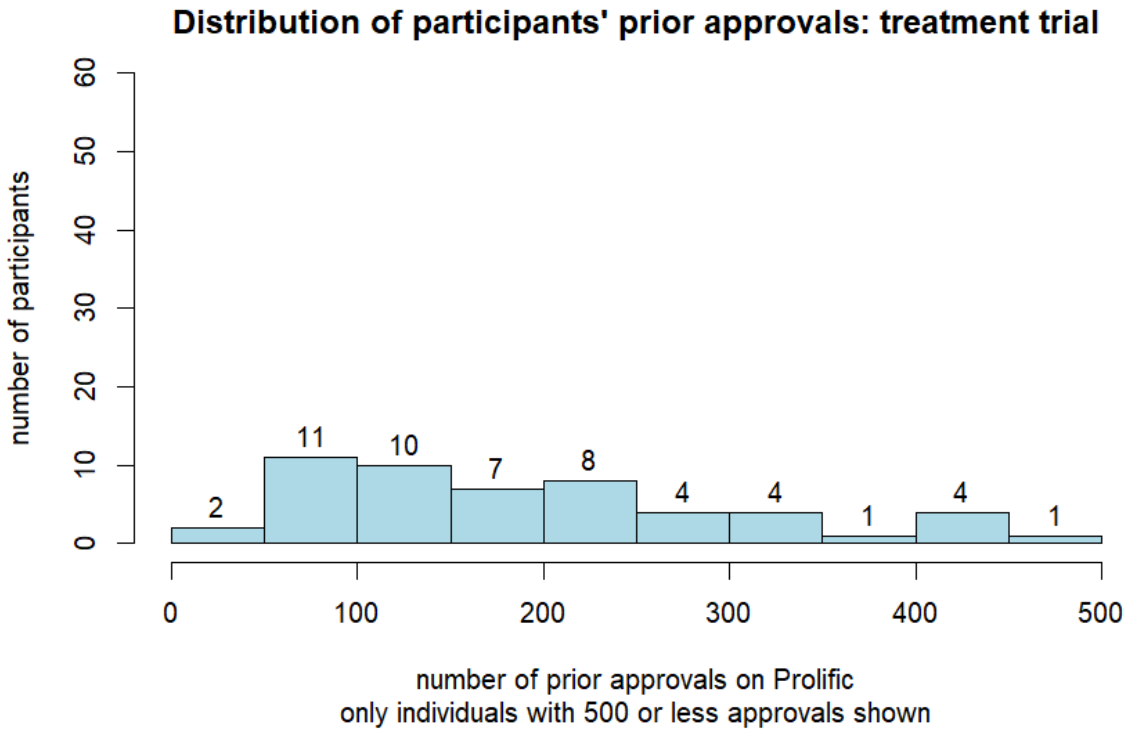
Participants' Characteristics 15: Prior Approvals, baseline trial



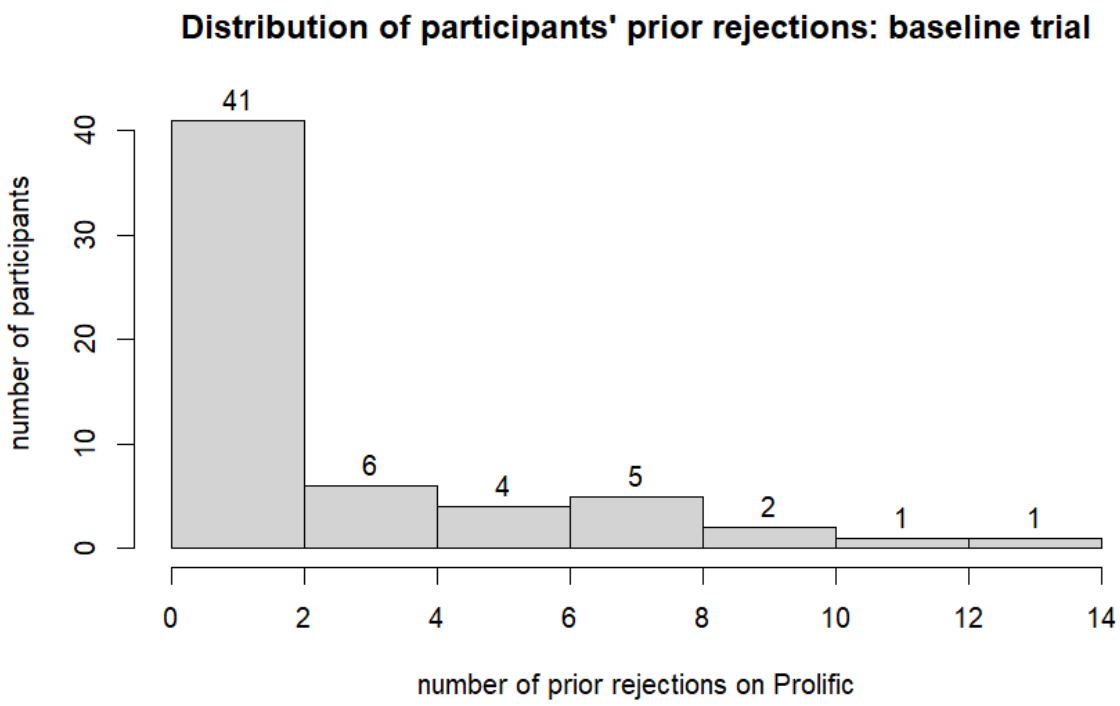
Participants' Characteristics 16: Prior Approvals, treatment trial



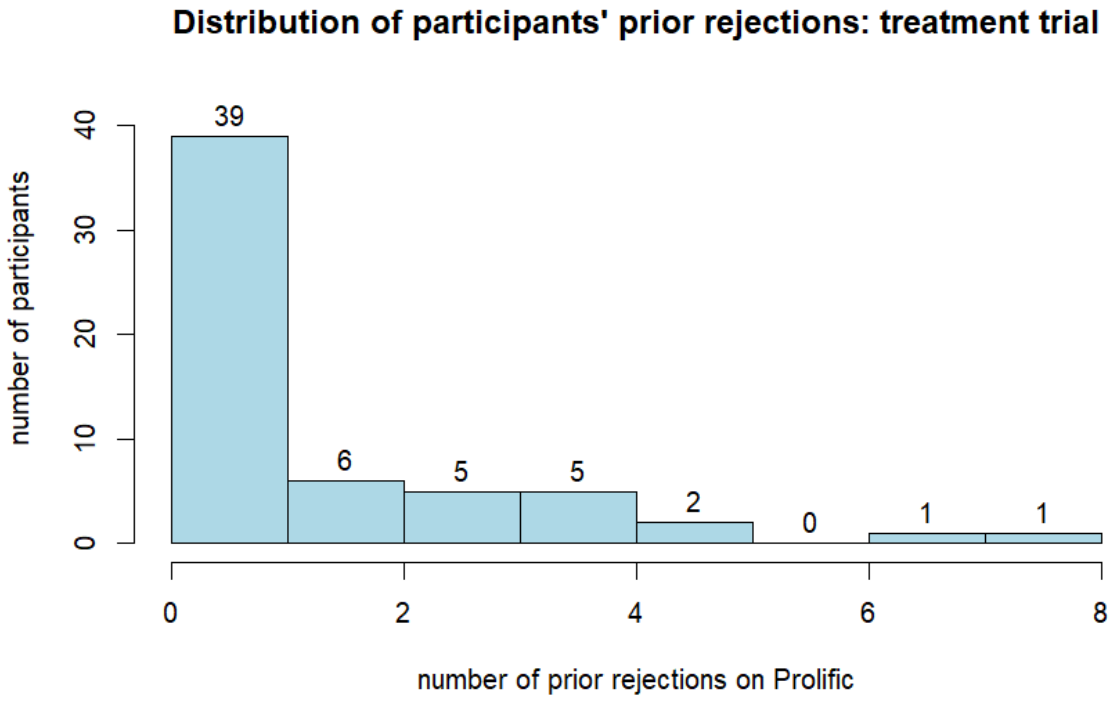
Participants' Characteristics 17: Prior Approvals, less than 500, baseline trial



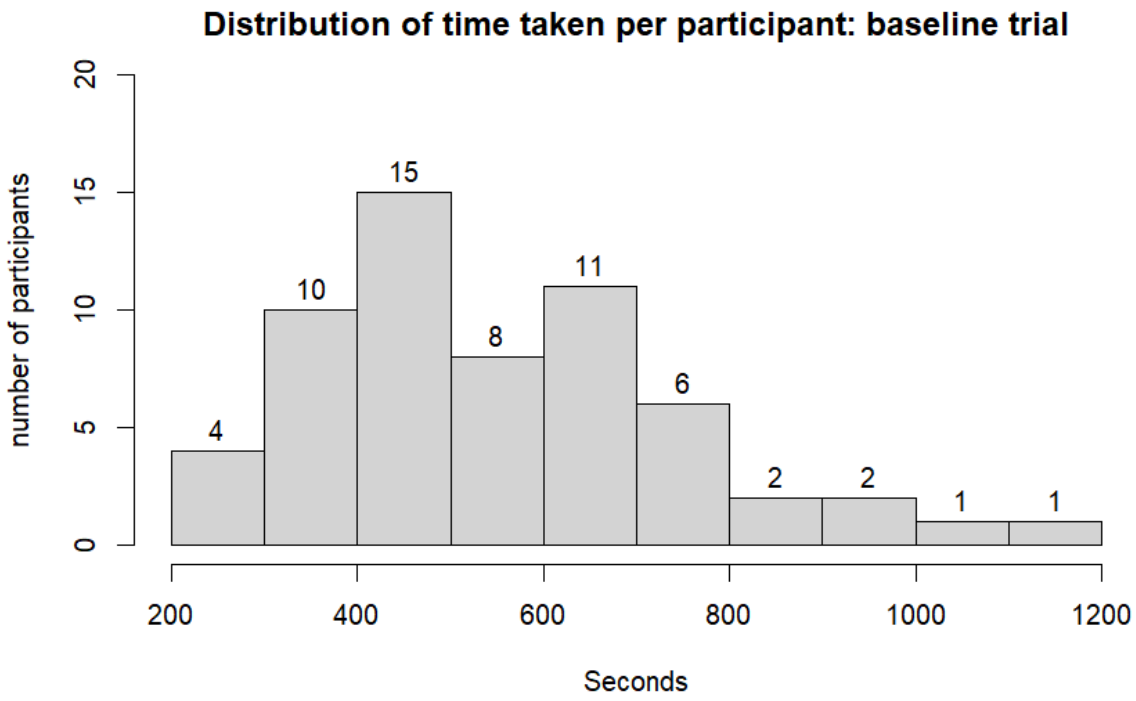
Participants' Characteristics 18: Prior Approvals, less than 500, treatment trial



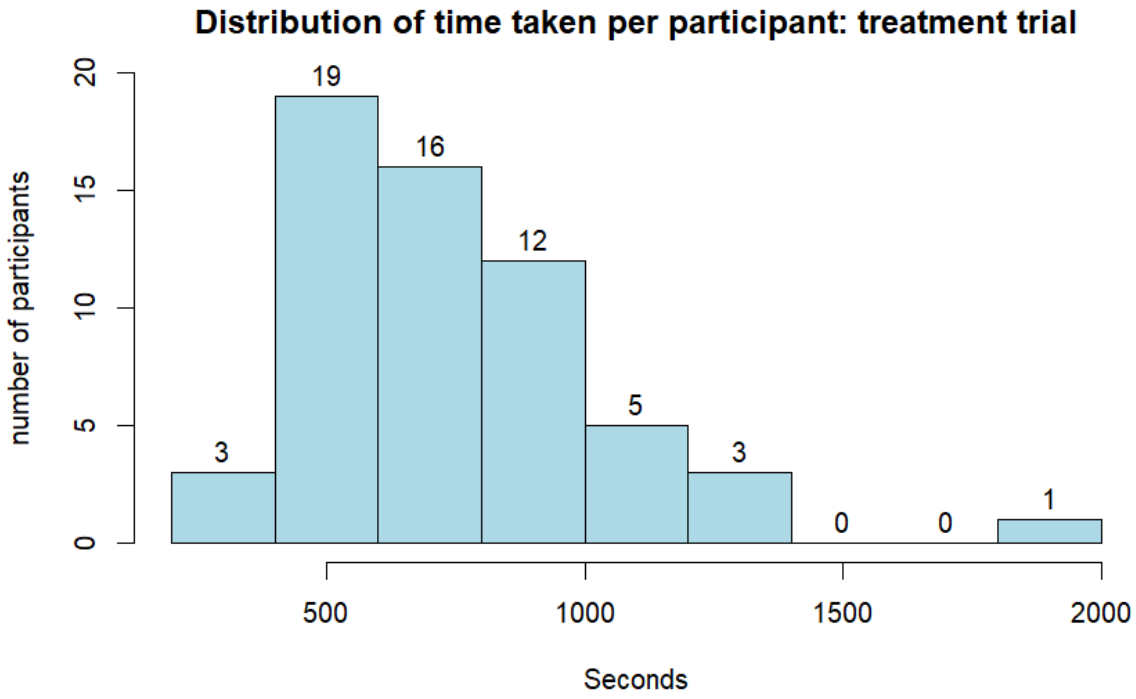
Participants' Characteristics 19: Prior Rejections, baseline trial



Participants' Characteristics 20: Prior Rejections, treatment trial



Participants' Characteristics 21: Time Taken, baseline trial



Participants' Characteristics 22: Time Taken, treatment trial

4.1.1.2.1 Discussion

Compared to their personal characteristics which were well within the expectations of the authors the Prolific specific characteristics of the participants can be described as terrifying:

With an average of 299 prior approved entries per subject in the baseline trial and 273 in the treatment study, referring to the participants as having prior knowledge and acquaintance with economic experiments would be quite an understatement.

While control questions are able to filter out some of the more problematic candidates in general, their experience as subjects dwarfs that of most academic researchers and is a major cause for concern. However, this total contains not only economic, psychological and sociological experimental trials but also a non-negligible number of commercial surveys and tasks more in line with traditional "click work". Nevertheless, we do not have any insights or data into the relative share of each type of assignment.

On a more positive note, most of the participants had a low number of prior rejected submissions, both in relative as well as in absolute numbers.

The last two plots, 21 and 22, are specific to the study itself and illustrate the time taken from accepting the experiment to finishing all tasks for subjects with approved submissions.

The introduction of the "putting a price" stimulus increased the average completion time by 199 seconds. While not much more than conjecture, we have three possible explanations for this observation. Firstly, that the introduction of the "tax" increased both the perceived as well as the actual complexity of the economic game and secondly that it caused them to invest more time into their decision making due to concerns about third party societal norms and judgement. Lastly, we also have to be open about the possibility that the changes to the instructions in the price trial made them simply harder to follow.

4.1.2 Payment

Baseline study:

payment total	total per person	num. of participants	show-up fee total	fee per person	bonus total	avg. bonus per person
159.020	2.683	60	90	1.500	69.020	1.183

Table 4: Payments to Participants in GBP: Baseline

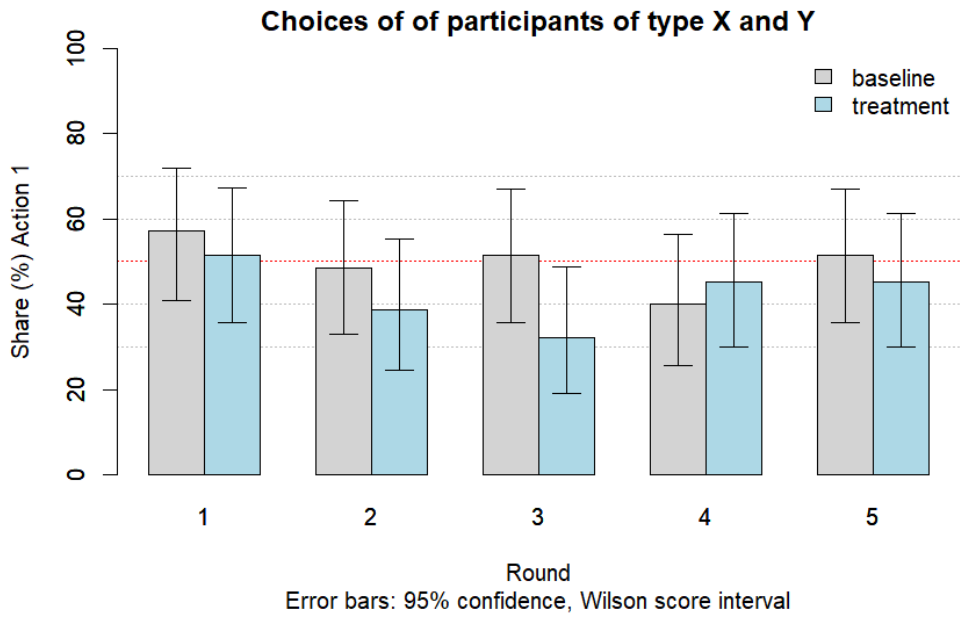
Treatment study:

payment total	total per person	num. of participants	show-up fee total	fee per person	bonus total	avg. bonus per person
163.920	2.826	58	104.980	1.810	58.940	1.016

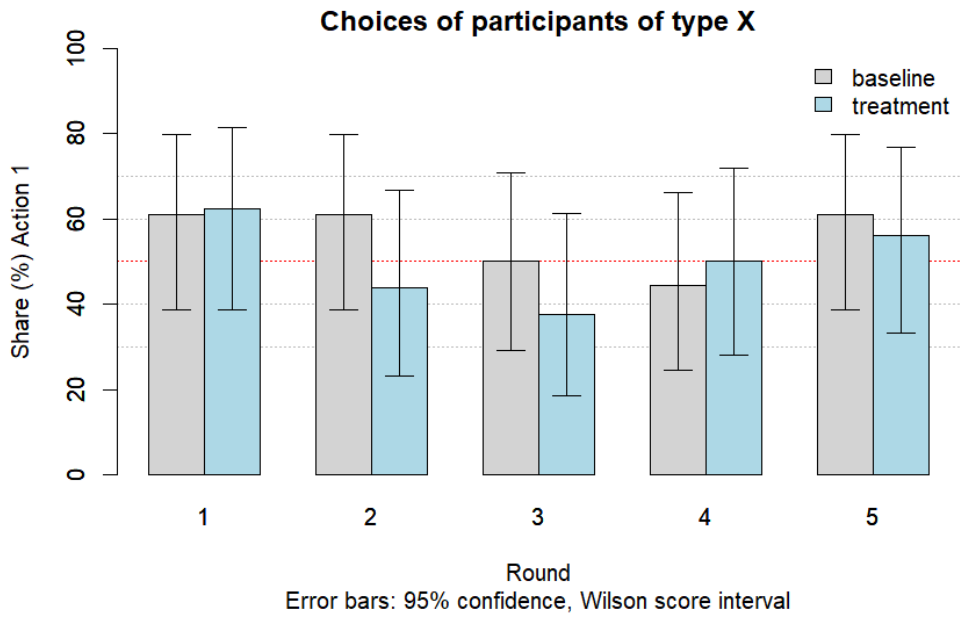
Table 5: Payments to Participants in GBP: Baseline

All values are excluding Prolific's commission of a third of the show-up fee. Participants in the treatment trial received a larger lump sum payment than those in the baseline to achieve the advertised hourly equivalent as required by Prolific's terms of service. This compensation was paid out after the experiment and therefore had no impact on the choices of subjects during the study.

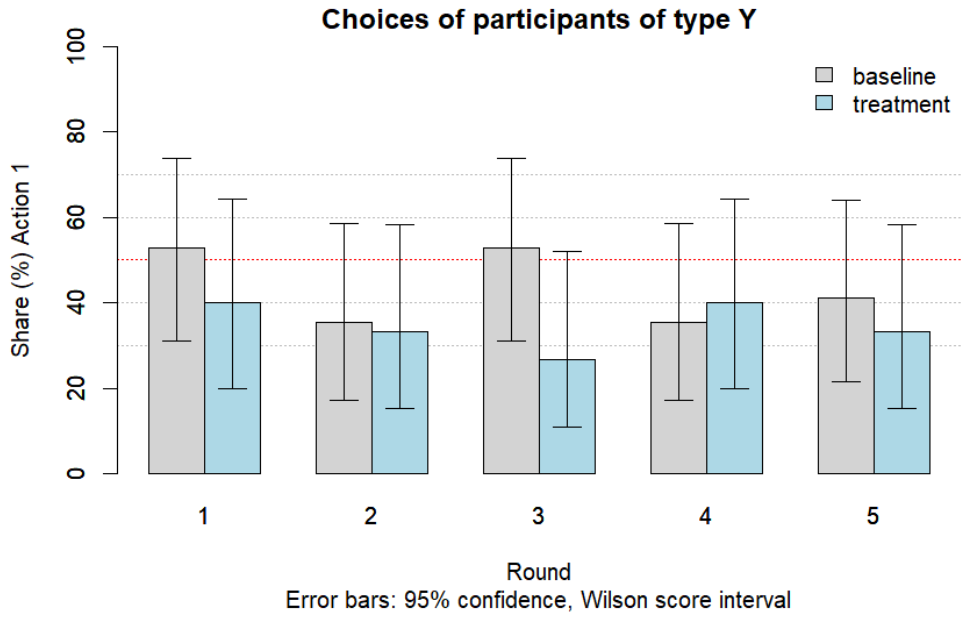
4.2 Stage Game: Actions



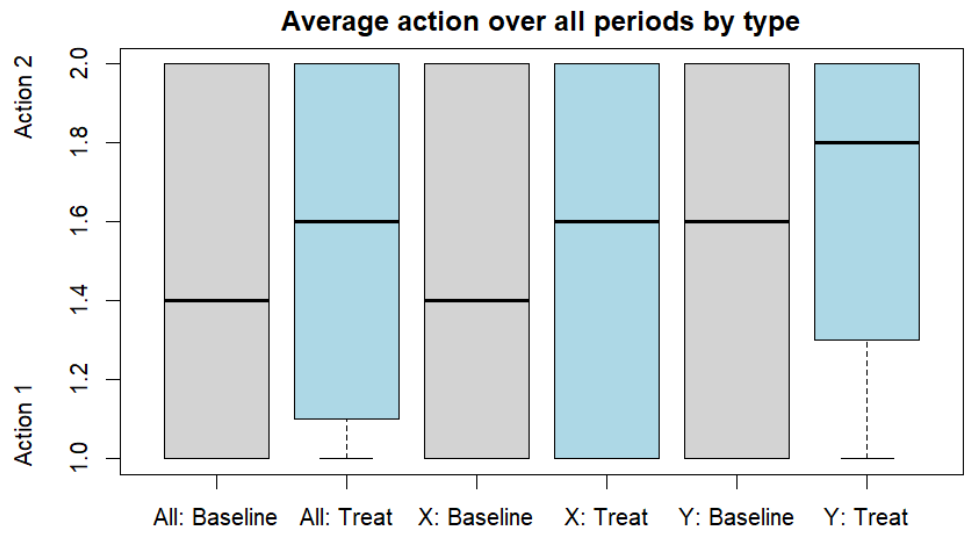
Plot 1: Share of Action 1 for X and Y



Plot 2: Share of Action 1 for X



Plot 3: Share of Action 1 for Y



Plot 4: Average Actions over all Periods by Type

Plot 4 illustrates the choices taken by each type and treatment if their per round action is aggregated to an average for each individual. Together with Table 6 this allows us to analyze changes in behavior by type and by treatment.

In line with our hypothesis H2, we observe a shift in the prevalence of action 1 in the treatment trial compared to the baseline setting. This can be mainly attributed to the choices made by players of type Y. Note however, that even for decisions of the players of type Y, individually only the 3rd round difference is significantly larger than zero at a 0.05 confidence level.

Due to our restricted sample we chose to analyze and test the aggregated data over all periods. While this does not violate the assumptions of the exact binomial test it is possible that this approach affected its statistical power.

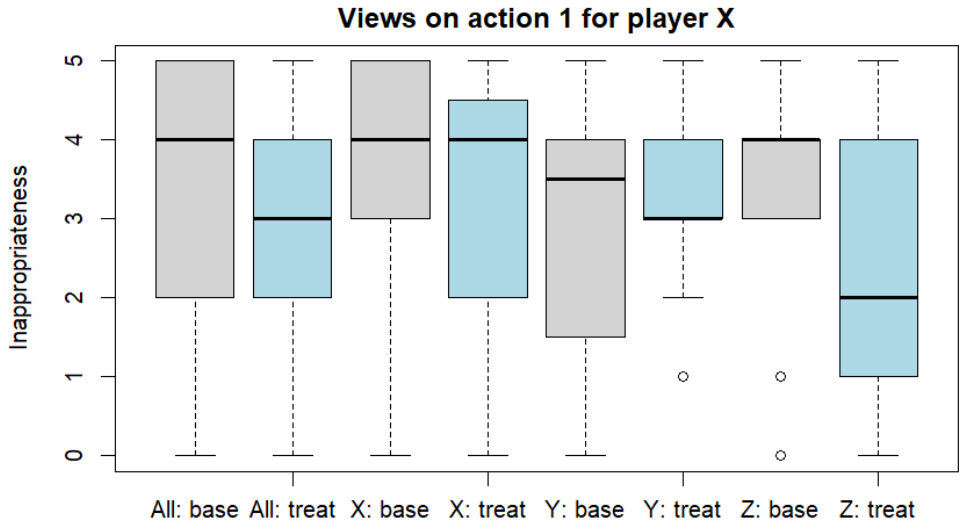
All: base vs treat	X: base vs treat	Y: base vs treat	base: X vs Y	treat: X vs Y
0	0.171	0.056	0.014	0.003

Table 6: Actions by active players: binomial exact one-sided p-values

We find significant differences in behavior of active participants by type in the baseline as well as the "price" treatment as proposed by hypothesis H1, but notably only significant shifts in

behavior by participants of type Y between the two treatments as put forward by hypothesis H2.

4.3 Social Norms

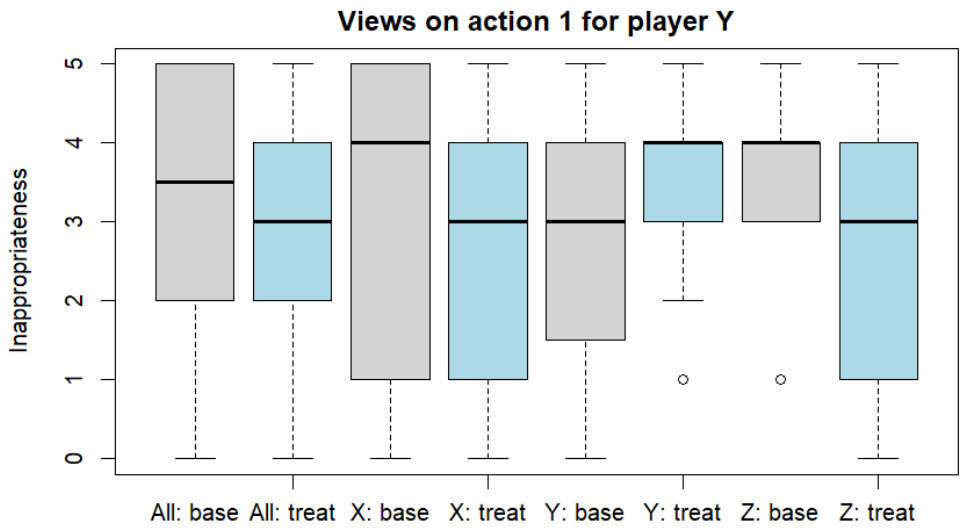


Plot 5: Views on action 1 for player X, by type and treatment

base vs treat	X: base vs treat	Y: base vs treat	Z: base vs treat
0.164	0.297	0.770	0.053

Table 7: Mann-Whitney U one sided p-values for player X: action 1

Using a statistical approach based on de Winter, Dodou (2010) and Derrick, White (2017) to analyze the distribution of the Likert data submitted by participants we only find a significant changes at the 0.1 level in the perception of players of type Z on the social appropriateness of the externality causing action by players of type X. Changes in the norm estimate of players of the other two types are not significantly different from zero.



Plot 6: Views on action 1 for player Y, by type and treatment

base vs treat	X: base vs treat	Y: base vs treat	Z: base vs treat
0.246	0.312	0.820	0.039

Table 8: Mann-Whitney U one sided p-values for player Y

This also holds true for the estimated appropriateness of action 1 by player Y. However, the relatively small sample size of less than 20 individuals per type and treatment could conceal a larger impact on norm perception.

5 Conclusion

Hypothesis 1: The game elicits different patterns of behavior depending on player type.

This can be answered with a resounding yes. Both in the baseline as well as the treatment environment the observed difference in behavior is significantly different from zero. This implies that this framework in combination with the chosen payoff parameters is able to elicit different behavior from different types of players.

Hypothesis 2: Behavior of participants differs between the baseline and treatment designs.

Here however, the answer is less straightforward. While the difference is highly significant at the 0.01 level if all active players are included in the sample, the same cannot be said about the individual types. While this could be caused by the small sample size overall and not necessarily by the underlying pattern of behavior, the fact that the p-value for type Y is close to the 0.05 level while the same value for X is anything but indicates that the introduction of price framing i.e., "putting a price" mainly affected players of type Y. While these players were aware of their excess negative impact on players of type Z in the base game the introduction of the "external tax" in the treatment without any change in monetary incentives had an impact on their actions. Either through signaling "external" social norms or acting as an additional reminder for the externality associated with action 1. Note that this finding is also in line with Bland and Nikiforakis (2015) results that the scale of an externality strongly affects behavior.

Hypothesis 3: Price framing does not only change the observed behavior but also the participants' estimate of the prevailing social norms of the appropriateness of the action with a negative externality.

Yes, but in a counterintuitive way. While the aggregate p-values as well as those for players of type X and Y are well above any reasonable threshold of significance, we observe a drastic change in participants of type Z.

Introducing in a sense not a moral wiggle room for the active players as in Dana et al. (2007) but contrary pacifying those affected by the externality. While this framework is not able to directly elicit utility levels of players of type Z, the change in their estimates of social norms indicates that taxes can have a placating effect and therefore can increase the utility of those affected without changing the underlying actions and effects.

To conclude, this thesis finds evidence that external price framing shapes not only decision making but also the expectations of social norms of participants affected by the externalities caused by these decisions. More importantly it also shows that this effect persists even if there is no change in the underlying monetary incentives.

Note however that the stimulus due to being framed as a tax also encompasses an additional societal component. Implying that the exogenous tax is well known, generally accepted, enforced or at least unavoidable.

A possible avenue for future research could encompass multiple treatments featuring the same price stimulus but varying the label presented to participants. For instance, transaction fee, fine or offset.

Returning to the original research question, namely how markets, interaction and prices shape our norms, this thesis demonstrates that the framework proposed by Christian Koch elicits significant changes in both actions and their social perception of appropriateness and could be extended to include further treatments focused on markets and interaction as the primary stimulus.

6 Appendix

6.1 Additional Screens

6.1.1 Baseline Treatment

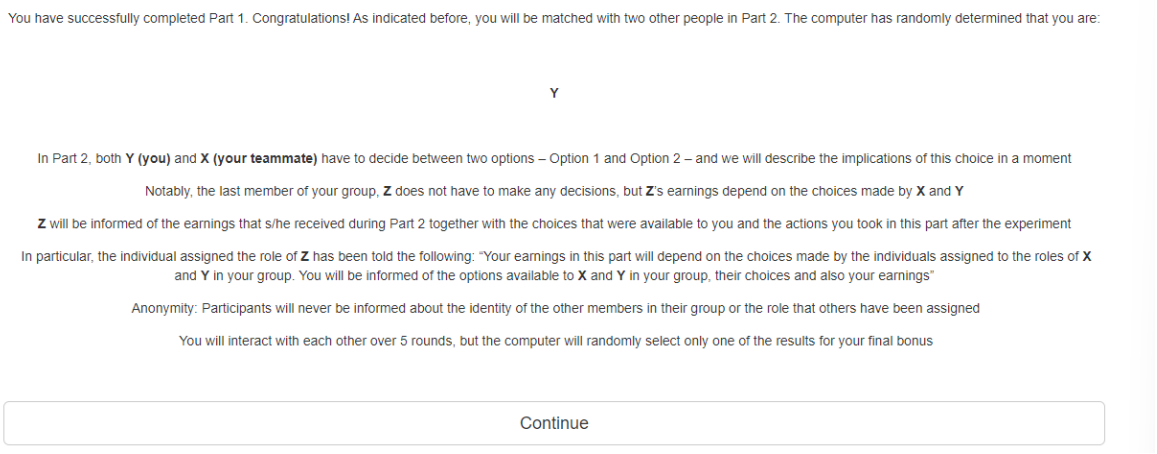


Figure 15: Introduction Part 2 for Type Y

Figure 16: Control Questions for Type Y

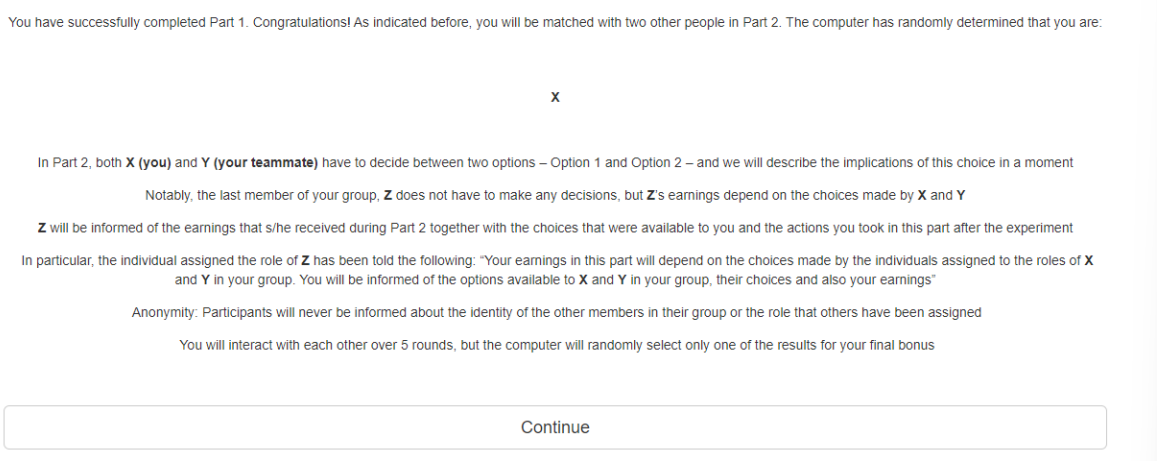


Figure 17: Introduction Part 2 for Type X

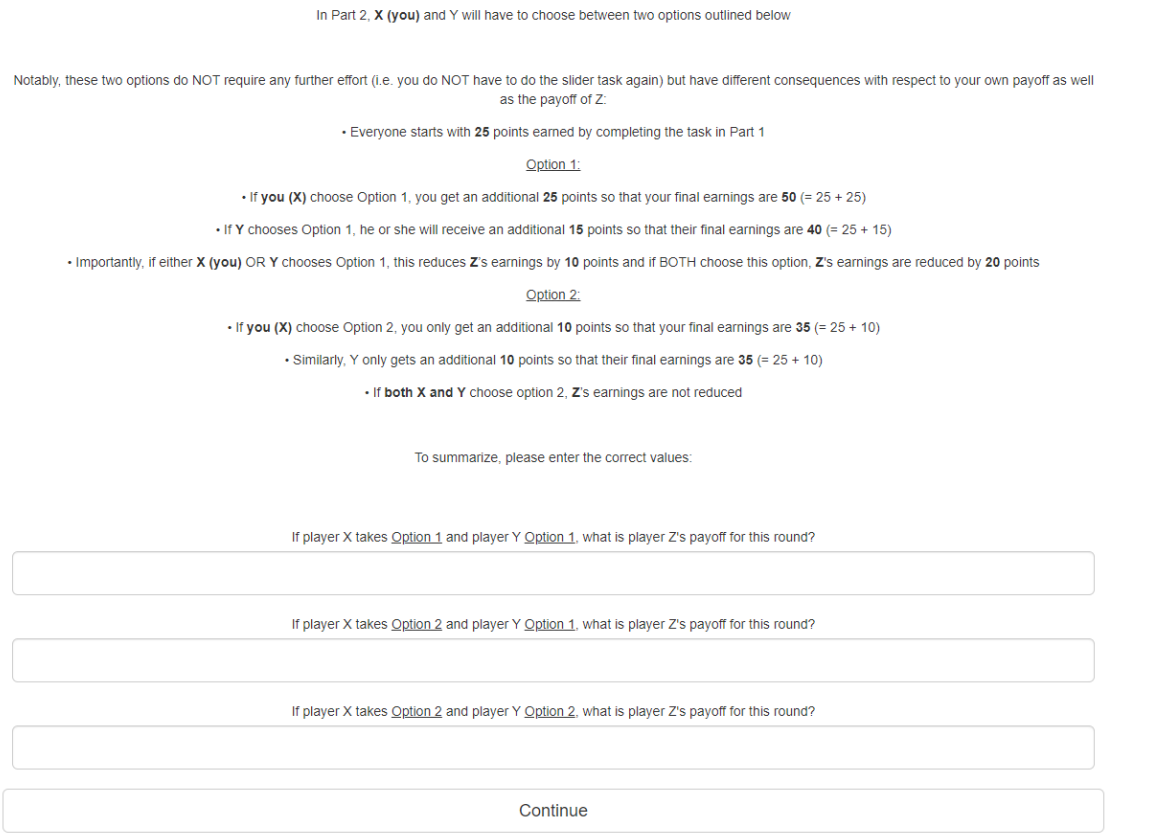


Figure 18: Control Questions for Type X

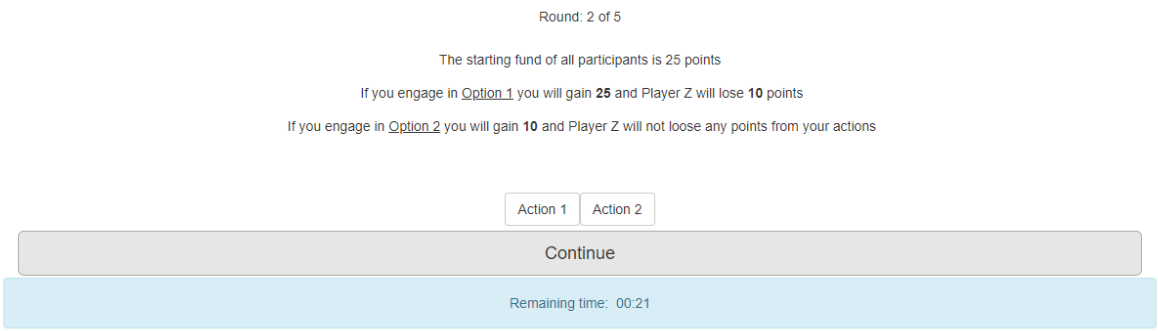


Figure 19: Choice Screen for Type X

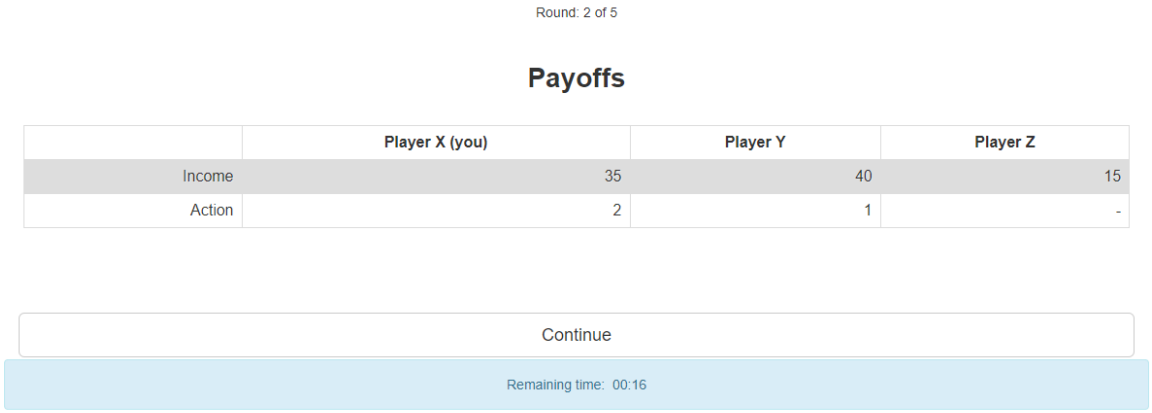


Figure 20: Payoff Screen for Type X

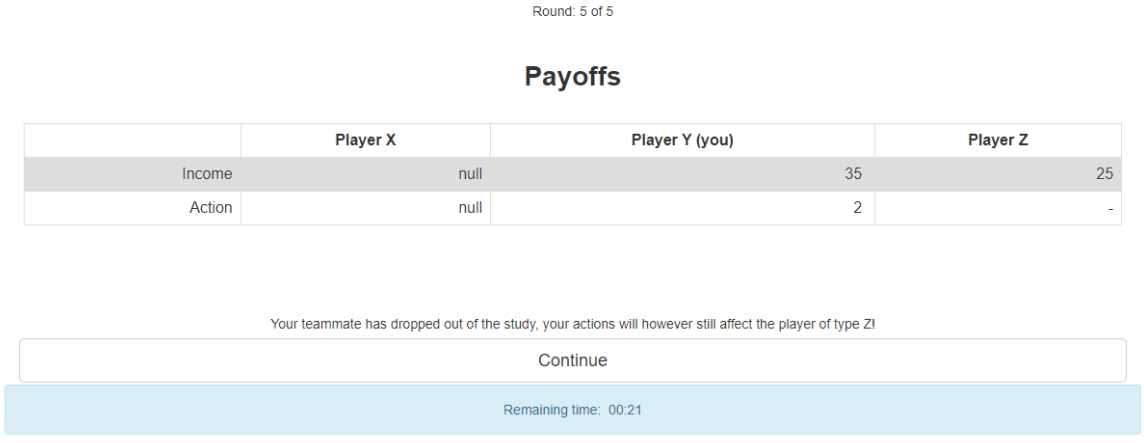


Figure 21: Payoff Screen if matched participant drops out

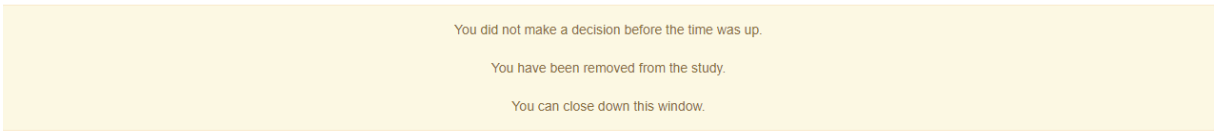


Figure 22: Exit Screen Time-Out

6.1.2 Price Treatment

Note that only screens that differ from the baseline treatment are presented on the following pages.

Figure 23: Control Questions for Type Z - Price Treatment

Figure 24: Control Questions for Type Y - Price Treatment

Figure 25: Control Questions for Type X - Price Treatment

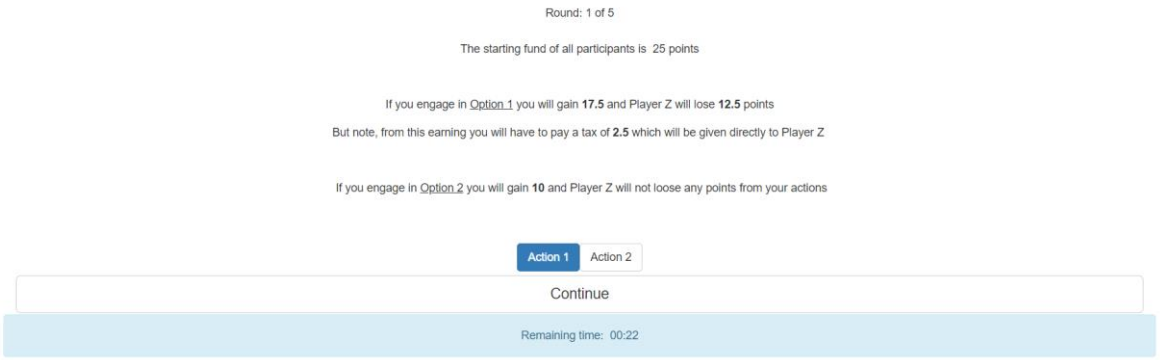


Figure 26: Choice Screen for Type Y - Price Treatment

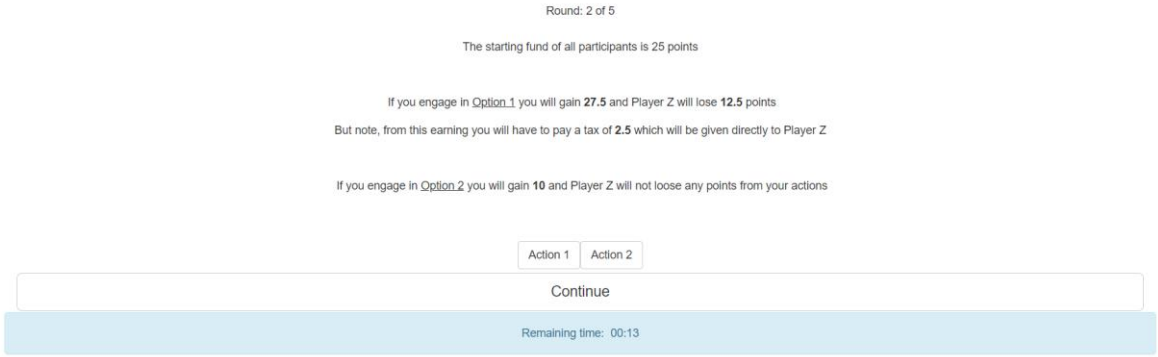


Figure 27: Choice Screen for Type X - Price Treatment

6.2 Additional Tables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
time_taken	60	555.1	200.7	276.3	410.1	667.8	1,199.6
age	60	36.6	12.5	19	27	43.2	69
num_approvals	60	299.4	265.5	11	104.2	401	1,289
num_rejections	60	2.4	3.2	0	0	3.2	13
prolific_score	60	99.4	1.1	96	99	100	100

Table 9: Participants' Characteristics: Baseline

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
time_taken	59	749.9	287.7	282.7	518.9	886.5	1,819.5
age	58	36.4	12.1	18.0	28.2	40.8	69.0
num_approvals	59	272.6	295.4	25	103	328	2,011
num_rejections	59	1.4	1.9	0	0	2	8
prolific_score	59	99.7	0.7	96	100	100	100

Table 10: Participants' Characteristics: Treatment

Round	Baseline	Treatment
1	0.571	0.516
2	0.486	0.387
3	0.514	0.323
4	0.400	0.452
5	0.514	0.452

Table 11: Share of Action 1 by all types

Round	Baseline	Treatment
1	0.611	0.625
2	0.611	0.438
3	0.500	0.375
4	0.444	0.500
5	0.611	0.562

Table 12: Share of Action 1 for type X

Round	Baseline	Treatment
1	0.529	0.400
2	0.353	0.333
3	0.529	0.267
4	0.353	0.400
5	0.412	0.333

Table 13: Share of Action 1 for type Y

Level	base all	treat all	base X	treat X	base Y	treat Y	base Z	treat Z
0	5	4	1	3	3		1	1
1	7	8	2	1	2	2	3	5
2					4	3		4
3	8	10	3	1	1	7	4	2
4	17	18	5	6	7	8	5	4
5	15	8	8	4	3	3	4	1

Table 14:Frequency of Perception of Inappropriateness of Action 1 by type X by type and treatment

Level	base all	treat all	base X	treat X	base Y	treat Y	base Z	treat Z
0	6	3	3	2	3			1
1	8	8	3	3	2	1	3	4
2	6	7	2	1	4	3		3
3	9	13	2	2	2	7	5	4
4	14	18	4	4	5	10	5	4
5	15	6	7	3	4	2	4	1

Table 15: Frequency of Perception of Inappropriateness of Action 1 by type Y by type and treatment

6.3 List of Figures

6.4 List of Tables

6.5 List of Plots

6.6 List Plots: Participants' Characteristics

6.7 Example Messages to Participants of Type Z

6.7.1 Two Matched Participants

Dear participant,

Thank you for taking part in our study

Your matched participant Y played Option 1 for 5 rounds, while player X chose 1,1,2,1,1 respectively

Round number 5 was randomly chosen for payout and has already been credited to your account

Thanks again and have a nice day!

6.7.2 One Matched Participant, One Dropout

Dear participant,

Thank you for taking part in our study

Your matched participant Y played Option 1 for 5 rounds, while player X dropped out without making any choice

Round number 5 was randomly chosen for payout and has already been credited to your account

Thanks again and have a nice day!

6.8 References

Arechar, A.A., Gächter, S. & Molleman, L. “Conducting interactive experiments online” (2018), Exp Econ 21, 99–131

Bartling, Björn, Weber, Roberto A. and Yao, Lan “Do Markets Erode Social Responsibility?” (2015), The Quarterly Journal of Economics, February 2015, Vol. 130, No. 1 (February 2015), pp. 219-266

Bartling, Björn and Fehr, Ernst and Özdemir, Yagiz: “Does Market Interaction Erode Moral Values?” (2021), Review of Economics and Statistics

Bartling, Björn, Vaelero, Vanessa, Weber, Roberto: “On the scope of externalities in experimental markets” (2019), Exp Econ (2019) 22:610–624

Bland, James and Nikiforakis, Nikos: “Coordination with third-party externalities”

(2015), European Economic Review, vol. 80, issue C, 1-15

Coffey, Helen "What is 'flygskam'? Everything you need to know about the environmental movement that's sweeping Europe" (2019), The Independent, 2019-06-05

Dana, Jason, Weber, Roberto A. and Xi Kuang, Jason: “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness” (2007), Economic Theory Vol. 33, No. 1, Symposium on Behavioral Game Theory (October 2007), pp. 67-80 (14 pages)

de Winter, J. F., & Dodou, D. (2010). Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). Practical Assessment, Research, and Evaluation, 15(1), 11.

Derrick, B., & White, P. (2017). Comparing two samples from an individual Likert question. International Journal of Mathematics and Statistics, 18(3).

Falk, Armin, Szech, Nora: “Morals and Markets” (2013), Science Vol 340 10 May: 707-711

Giamattei, M., Yahosseini, K.S., Gächter, S. et al. „LIONESS Lab: a free web-based platform for conducting interactive experiments online” (2020) J Econ Sci Assoc 6, 95–111

Irlenbusch, Bernd, Saxler, David J.: „The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior” (2019), Journal of Behavioral and Experimental Economics 80 (2019) 141–161

Kirchler, Michael, Huber, Jürgen, Stefan, Matthias, Sutter, Matthias: “Market Design and Moral Behavior” (2016), Management Science 62(9):2615-2625.

Krupka, Erin L., Weber, Roberto A. 2013 “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?” (2013) Journal of the European Economic Association, Volume 11, Issue 3, 1 June 2013, Pages 495–524

Palan, Stefan, Schitter Christian: “Prolific.ac—A subject pool for online experiments” (2018) Journal of Behavioral and Experimental Finance ,Volume 17, (2018),Pages 22-27

Plott, C. R. “Externalities and corrective policies in experimental markets” (1983), Economic Journal, 93, 106–127.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rienzo, C. & Vargas-Silva, C. (2022) Migrants in the UK: An Overview. Migration Observatory briefing, COMPAS, University of Oxford

Rode, Julian, Hogarth, Robin M., Le Menestrel, Marc, “Ethical differentiation and market behavior: An experimental approach” (2008), Journal of Economic Behavior & Organization Vol. 66 265–280

Sandel, Michael J.: "Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy." (2013) Journal of Economic Perspectives, 27 (4): 121-40.

Ziegler, Andreas, Romagnoli, Giorgia, Offerman Theo: “Morals in multi-unit markets” (2020), Tinbergen Institute Discussion Paper TI 2020-072/I

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

The participants for this paper were recruited using Prolific (www.prolific.co)