Empathy, Motivated Reasoning, And Redistribution

Tingyan Jia*

Stanford University

This paper is evolving. Please click here for the latest version.

October 31, 2022

Abstract

I investigate both theoretically and experimentally the economics of empathy and its implications for redistribution. I first define empathy as an accurate simulation of how one would feel if they were in another's position, distinguishing it from altruism. I propose a novel mechanism by which personal experience affects distributional motives through empathy: wealthy individuals have selfish motivation not to be empathetic towards the poor in order to justify less redistribution; in addition, more varied personal experience of consumption constrains such motivated reasoning, therefore, increasing empathy and redistribution. I provide a test of the mechanism in a laboratory setting. I create exogenous variation in experiences and manipulate the timing of information to identify the role of motivated reasoning for subjects with different experiences. I find strong support for the validity of the mechanism: subjects with uniform experience are more susceptible to self-serving motivated reasoning in both their empathy beliefs and redistribution choices.

*I thank my committee members, B. Doug Bernheim, Muriel Niederle, and Matthew Jackson for their advice and support. I am especially grateful for the countless conversations I had with Professor Bernheim in developing this paper. I also thank Ran Abramitzky, Sandro Ambuehl, Arun Chandrasekhar, Pascline Dupas, Christine Exley, Marcel Fafcamps, Caroline Hoxby, Collin Raymond, Alvin Roth, David Zuckerman, participants in Stanford Behavioral and Experimental Lunch, Stanford Applied Micro Lunch for their useful comments. All errors are my own.

1 Introduction

Economic inequality has been increasing around the globe, and its significance was arguably exacerbated by the COVID-19 crisis.[1] Redistributing resources from the wealthy to the poor could reduce such inequality, but this would face resistance from wealthy individuals who are usually politically powerful. What determines wealthy individuals' attitudes toward redistribution? The behavioral economic literature has explored several mechanisms in depth: other-regarding preferences (e.g. altruism, fairness), reputation (e.g. utility from being viewed as fair or altruistic), and beliefs about the structural reasons for inequality (e.g. attribution of success to skill vs. to luck).

This paper provides a theory and a test of an additional mechanism, *empathy*, that affects the propensity of the wealthy to redistribute their income to the poor. The word "empathy" has many different meanings.[2] This paper focuses on "empathetic simulation", by which I mean the ability to accurately simulate how one would feel if they were in another's position. My focus of empathetic simulation follows Adam Smith who wrote, in the Theory of Moral Sentiments, "As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation".

By defining empathy in terms of the accuracy of beliefs about "in-their-shoes" experience, I am able to study how it engages with motivated reasoning in the context of redistribution. Motivated reasoning is a term from psychology that informally refers to behaviors such as finding excuses or lying to oneself (e.g. Erdelyi 1974, Kunda 1990). A burgeoning literature in behavioral economics tackles motivated reasoning theoretically (e.g. Benabou and Tirole 2006, Brunnermeier and Parker 2005, Bernheim et al. 2021) and empirically (e.g. Dana et al. 2007, Exley 2016).

It is intuitive that people may resort to motivated reasoning when they are redistributing resources from themselves to others. According to my definition of empathy, wealthy individuals are more empathetic if they are more willing to acknowledge that the marginal utility of consumption when poor is much higher than that when rich. Without motivated reasoning, wealthy individuals would need to either lower their own consumption utility if they choose to redistribute or suffer from utility loss (e.g. feeling guilty) if they do not redistribute. With motivated reasoning, my hypothesis is that wealthy individuals can talk themselves into believing the marginal utility of consumption for the poor is *not* much higher than that of their own consumption, so that they could mitigate both the material loss from consumption redistribution and the utility loss from appearing selfish. Such

[1] For income inequality in the US as well as globally, see: e.g. Piketty and Saez 2003, Piketty 2003, Atkinson et al. 2011, Piketty and Saez 2014, Piketty et al. 2019, Chancel and Piketty 2021. For debates on income inequality since COVID-19, see: e.g. Stiglitz 2020, Deaton 2021.

[2] For empathy in the psychology literature, see e.g. Batson 2009, Elliott et al. 2011, Stocks and Lishner 2012, Zaki 2014.

self-deception is at least partially constrained by glimmers of the truth: lying to oneself leads to cognitive dissonance, which is psychologically costly.[3]

Supposing these excuse-seeking behaviors do exist in the redistribution decisions of wealthy individuals, I wonder what could potentially lower the degree of such self-serving bias and increase redistribution. In this paper, I address the research question: how does personal experience impact redistribution through empathy?

There is mixed empirical evidence on how experience affects redistributive preferences. On the one hand, there is motivating evidence that wealthy individuals who grew up poor are more supportive of left-wing parties and redistributive policies than those who have always been rich (e.g. Cherkaoui 1992), and that personal hardships from systematic economic shocks make people more supportive of redistribution (e.g. Margalit 2013, Giuliano and Spilimbergo 2014). On the other hand, there is also empirical evidence that experience of poverty and economic hardships might not only affect empathy, but also views on social mobility and inequality (e.g. Piketty 1995, Cohn et al. 2019, Roth and Wohlfart 2018). Building on these empirical findings, I will isolate the effect of exogenous experience and focus on its interaction effect with motivated reasoning.

My main hypothesized mechanism is that wealthy individuals have selfish motivations not to be empathetic towards the poor so that they can justify less redistribution. In addition, I argue whether or not someone has experienced being poor themselves in the past can affect their empathy today through limiting motivated reasoning. However, experience can constrain this tendency, because having been poor provides the individual with direct evidence of how it feels to be poor. To provide intuition, let us compare a wealthy individual who has experienced poverty themselves to one who has always been wealthy. Any belief that contradicts their experience would cause the former a larger cognitive dissonance than the latter. Hence, people who have experienced greater income variation reduce the costs of that dissonance by accepting a belief that is closer to the truth.

In this paper, I first formalize the intuition with an economic model. I consider different emotional sources of motivated reasoning, and show that they lead to different conclusions. I then test my proposed empathy mechanism using a laboratory experiment.

In my model, a wealthy individual has a history of consumption experience. There is also a poor individual who is endowed with no income and needs to consume out of the transfer from the wealthy individual. The wealthy individual chooses not only how much consumption to redistribute to the poor, but also what to think about the poor's marginal utility of consumption (relative to that of their own) to justify their redistribution decision. I define altruism as the weight they put on the well being of the poor relative to that of their own, thus separating it from empathy.

Holding altruism level fixed, the wealthy individual could conduct motivated reasoning in order to transfer less or more to the poor. I assume that people experience visceral

[3]For cognitive dissonance theory, see e.g. Festinger 1962, Aronson 1992.

emotions, such as guilt (because they feel they ought to transfer more) and temptation (because they would selfishly like to transfer less) when they make their transfer decisions. However, their preferences are potentially time-inconsistent: they may place less weight on those reactions, or even ignore them altogether, when considering the problem in advance.[4]

Time inconsistency is an essential part of the mechanism I study: I show that, if the wealthy individual has identical sensitivities to those emotions when they think about redistribution ahead of time and when they are redistributing, there would be no bias in their beliefs. Accordingly, their redistribution would be the same as when they could not conduct motivated reasoning.

Different visceral emotions lead to divergent conclusions: I show that if the wealthy individual expects themselves to feel only *guilt* and no temptation in redistribution, then they could choose to be *less* empathetic towards the poor to prevent their future selves from redistributing too generously to the poor. To the contrary, I show that if the wealthy individual expects to feel only *temptation* and no guilt in redistribution, then they could choose to be *more* empathetic towards the poor in order to refrain their future selves from keeping too much resources to themselves.[5]

My model solution confirms that a larger variance of their own consumption experience always shrinks their motivated bias. Varied experience offsets some of the motivated reasoning effect on their chosen redistribution. That is, I show that for a wealthy individual who feels guilt but not temptation, more varied personal experience leads to higher empathy and higher redistribution. For a wealthy individual who feels temptation but not guilt, more varied personal experience leads to lower empathy and less redistribution. Hence, my model provides a simple conceptual framework to explain people's distributional preferences through the endogenous determination of empathy by experience.

These opposing results set up a horse rate between three hypotheses: motivated reasoning to overcome guilt, motivated reasoning to overcome temptation, and no motivated reasoning.

In the second part of the paper, I simulate the model environment in a laboratory experiment to isolate empathy from other factors that might affect redistribution and test my model mechanism. I deduce whether guilt or temptation is the dominant emotion based on the detected empirical results in the data.

The experiment has three stages. In Stages 1 and 2, I assign subjects with various levels of real effort tasks randomly. In Stage 3, I measure their empathy and have them redistribute some tasks from their partners to themselves. The real effort tasks consist of the same basic unit, with the only difference being shorter or longer duration of the

[4]One could assume that emotions are more intense in advance, which would reverse my results. This alternative assumption, however, does not seem realistic.

[5]For a wealthy individual who feel *both* guilt and temptation, the results would lie somewhere *in between* the two extremes.

task. The first two stages constitute my experience variation treatment: subjects have either uniform experience (assigned the short version of the tasks in both stages), or varied experience (assigned the short version for the first stage and the long version in the second stage).

The task I use is a clicking task that requires subjects to click inside a small red box for either 100 seconds or 500 seconds with a minimum speed of 4 times per second. The task is easy enough for everyone to perform immediately without learning, yet it is unpleasant enough for nobody to be willing to do for leisure. Also, due to muscle fatigue and boredom, it is increasingly difficult to keep up to the required speed as the duration of the task increases. Combining these traits, my real effort task with a convex cost schedule can correspond to the consumption experience in the model with a concave utility schedule. One needs to think of more clicking as lower consumption (of leisure) and less clicking as higher consumption.[6]

In the third and final stage of the experiment, all subjects become the "wealthy individual" in my model. That is, all subjects are endowed with the short version of the task in Stage 3 regardless of what kind of tasks they did in the first two stages. I match every wealthy individual with another subject from a separate participant pool as partners. This other subject is endowed with the long version of the task and stands for "the poor" in my model. The subject who plays the role of the wealthy individual in the model can reallocate some of their partner's task to themselves.

I use a 2-by-2 design in this experiment. One dimension of variation is whether the subject has varied experience or uniform experience with the real effort task in Stages 1 to 2. The other dimension of variation is whether I elicit subjective beliefs about the curvature of the effort cost function *before or after* subjects are informed about the opportunity to make a transfer to their partner in Stage 3. The purpose of this second dimension is to mitigate the scope for motivated reasoning–once they have expressed an unbiased belief, they will experience additional cognitive dissonance if they try to talk themselves into something else. Hence, I call the second treatment Cognitive Dissonance Amplification (CDA) treatment.

I measure empathy by asking subjects about what they think others in the study report as the marginal unpleasantness of completing the short versus the long task. Subjects are incentivized for accuracy in their predictions. Based on how I characterize empathy in the model, the wealthy individual in Stage 3 of the experiment are more empathetic towards someone assigned the long task if they acknowledge a larger convexity of the effort cost function.[7]

[6]Without a budget constraint, one could create large variation of consumption-utility experiences between subjects by endowing some of them with large sums of money. Real effort tasks with no learning curve and a convex cost function are an economical substitute for consumption experience.

[7]Note that empathy measured in the experiment could be either upward or downward biased with regard to the true convexity of the cost function. I am not interested in the absolute value of empathy, instead, I

After I measure their empathy, I have subjects redistribute tasks from their partners to themselves. For subjects in the CDA control group, they knew about redistribution before they formed their empathy beliefs, and the process mimics real world decision making where people form their beliefs about the poor knowing that redistribution is a possibility. For subjects in the CDA treatment group, they did not know about the choice to redistribute when formulating beliefs. These stated beliefs should then constrain subsequent motivated reasoning when subjects make a transfer, because any attempt to revise those beliefs would create additional cognitive dissonance. Hence, I anticipate a CDA treatment effect for both the empathy beliefs and redistribution choices.

I analyze the results using a Difference-in-Difference method. Looking at simple differences is not informative for the following reasons: a simple comparison across subjects with different dissonance constraints (arising from the timing of information concerning the transfer) does not necessarily shed light on motivated reasoning, because the treatment may serve as a "nudge" to make generous transfers. Similarly, a simple comparison across subjects with different experiences does not necessarily shed light on motivated reasoning, because their differing levels of "wealth" (arising from the different amount of real effort task completed in Stages 1 to 2) for the entire experiment may affect their levels of generosity. However, looking at Difference-in-Difference is informative because it eliminates both of those confounds and speaks to the central theoretical prediction that experience weakens motivated reasoning.

The experiment was conducted online through the Prolific platform. I find that compared to subjects with varied experience, subjects with uniform experience are more subject to selfish motivated reasoning in their empathy beliefs and redistribute less generously as a result. This detected interaction effect between experience variation and motivated reasoning is driven by subjects with median and below median level of altruism in my sample. As a secondary result, my data also show some gender differences in empathy and redistribution propensity. Women are on average more empathetic than men in acknowledging doing the hard task is more unpleasant than doing the easy one, and they are more likely to help, after controlling for individual-level altruism. [8] These findings favor guilt over temptation in the theoretical horse race that I set up earlier. The results also corroborate the key mechanism of my model: without the constraint of experiencing poverty themselves, wealthy individuals do conduct motivated reasoning to reduce empathy in order to not feel guilty for redistributing too little effort task from their partners to themselves.

This paper has four main contributions. My first contribution is to provide, to the best of my knowledge, the first formal economic model of empathy. I treat empathy as

study how empathy changes with selfish motivations and prior personal experience.

[8] For an overview about gender differences in preferences, see e.g. Croson and Gneezy 2009. In terms of gender differences in social preferences, the findings are mixed, and seem to depend on specific experiment design and implementation.

a belief phenomenon, and thereby distinguishes it from altruism, which is an aspect of preferences. My second contribution is to propose a novel mechanism by which personal experience affects distributional motives through empathy: more varied experience limits the self-serving reduction in empathy when a wealthy individual is asked to redistribute to the poor. [9] My third contribution is to provide a test of the mechanism in a laboratory setting: I create exogenous variation in experiences and identify the role of motivated reasoning in empathy and redistribution for subjects with different prior experiences. Finally, I find strong support for the validity of the mechanism: subjects with uniform experience are more susceptible to self-serving motivated reasoning in both their empathy beliefs and redistribution choices.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 develops a theory of empathy. Section 4 introduces my experimental design. Section 5 describes the findings of the laboratory experiment. Section 6 concludes.

2 Literature Review

The paper relates to several strands of literature: empathy, distributional preferences, motivated reasoning, experience effect, and positive welfare economics.

Compared to altruism, the literature on empathy is sparse.[10] Loewenstein 2005 discusses the intrapersonal and interpersonal hot-cold gap of empathy and discuss its implication for medical decision making. Their projection bias mechanism is about the impact of the *current* state, whereas this paper focuses on the impact of *previous* states. Stark and Falk 2000 study empathy in the context of a transfer and a reverse transfer. Their definition of empathy is how much the receiver of a transfer factors the giver's utility in his own preference, which would be defined instead as altruism in this paper.

The most related work on empathy in psychology is Zaki 2014 and Lönnqvist and Walkowitz 2019. Zaki 2014 provides a thorough description of the psychological process of empathy formation and, in particular, argues conceptually that avoiding material costs may motivate people to be less empathetic, which is consistent with the mechanisms studied in this paper. Lönnqvist and Walkowitz 2019 shows that givers in a dictator game report more empathy toward the receiver if they need to write about feelings of the receiver, but such increase in reported empathy does not lead to more transfers. There are more psychology work on empathy that enumerates various phenomena as empathy (e.g. Batson 2009), defines empathy as both emotional simulation and perspective taking in psychotherapist and client relationships (e.g. Elliott et al. 2011), and measures empathy

[9] Empathy is not the only mechanism by which experience can impact personal preferences. People with varied experience might, for example, have a different view about climbing out of poverty than people who have always been wealthy. This is a different kind of motivated belief from empathy and may have a different impact on distributional preferences.

[10] Economic research on altruism dates back to, for example, Becker 1974 and Andreoni 1989.

using the Interpersonal Reactivity Index questionnaire (e.g. Stocks and Lishner 2012).

The literature suggests several prominent factors other than empathy that may affect wealthy individual's preference for redistribution. The first factor is altruism: Andreoni (1995) finds that differences in altruism explains a lot of heterogeneity in people's pro-social decisions. Parker 2012 surveys 2508 adults in the US and 55% of them believe the wealthy are greedier than others. The second factor is fairness: Roth and Wohlfart 2018 argues that attitudes towards redistribution are determined through fairness views using the US, German and European social survey data sets. Fisman et al. 2015 shows Yale law school students are more efficiency than fairness focused than average Americans when they play modified dictator games. Fisman et al. 2017 shows that equality-focused Americans are more likely to be Democratic. Cohn et al. 2021 shows that wealthy Americans tolerate more unequal distributions between two third parties than other Americans regardless of whether the inequality is caused by luck, skill, or a mixture of both. Fisman et al. 2021 studies distributional preference in large groups of around seven people, and find subjects prefer to lower the top income and increase the bottom income within a group.

The third factor is reputation or self-image: Andreoni and Bernheim 2009 shows both theoretically and experimentally that how likely a dictator divides equally between himself and someone else depends on the observability of his choice. Konow 2000 decomposes a dictator's choice into consumption utility, value of fainess to himself, and cost of self-deception of what is fair. The fourth factor is the attribution of success to luck versus skill: Benabou and Tirole 2006 provides a theory explaining the differences of tax rates in the US versus in Europe with differences in beliefs about a "just" world where efforts pay off. In this paper, I show empathy is an additional factor that affects redistribution.

This paper also identifies a new class of environments in which motivated reasoning may play an important role. Existing motivated reasoning literature usually requires one of the following: asymmetric updating in ego-related domains such as intelligence, beauty, or political affliations (Eil and Rao 2011, Thaler 2020), uncertainty in the mapping from one's selfish decision to another party's payoffs (Dana et al. 2007, Di Tella et al. 2015, Exley 2016, Exley 2020), choice architecture that allows one to attribute his selfish actions to framing or anchoring bias (Exley and Kessler 2019). In this paper, the choice environment is about ego-irrelevant real effort tasks. The main choice is simple redistribution with no uncertainty in payoffs to either party. And the choice interface is a standard multiple price list without any framing or anchoring effect. My findings show the generality of motivated reasoning.

Literature shows various ways attitudes and actions are affected by experience. Lowe 2021 conducted a field experiment randomly assigning Indian men from different castes to different teams in cricket games and found that those who have been teammates with players from a different caste increases cross-caste friendship and decreases own-caste fa-

voritism. Ager et al. 2017 analyzed data on World War II fighter pilots victory and death rates, and found that when a pilot's performance was recognized, those who were former peers with him also fought more heroically and achieved more victories. The public economic literature also provides evidence that personal history and historical experiences affect people's preference for redistribution, see Corneo and Grüner 2002, Lee and Roemer 2006, Alesina and Giuliano 2011. These documented experience effects motivate me to consider the role of experience in empathy beliefs and redistribution choices.

In the macroeconomic literature, recent economic episodes can be extrapolated to form beliefs about the future. In particular, stock market and inflation experiences were documented to be correlated with beliefs about future risk-adjusted returns of stocks and inflation rates (Malmendier and Nagel 2011, Malmendier and Nagel 2016). Roth and Wohlfart 2018 provides empirical evidence that experiencing inequality during one's lifetime may change views about fairness. Bernheim et al. 2021 provides a theory of how these experience effects may play out on future worldviews and preferences in a persistent fashion. In this paper, personal experience affects the accuracy with which one project's oneself into another's situation. More varied personal experience constrains self-benefiting biases by causing larger cognitive dissonance (given any biased beliefs), making beliefs more accurate as a result.

This paper proposes empathy as another mechanism that experience with income mobility might change redistribution preferences. Cherkaoui 1992 documents that voters who are wealthy now but grew up poor are 50% more likely to vote for left-wing parties than those who have always been wealthy. Margalit 2013 shows that the financial crisis has a sizable yet transient effect on voter's preference for welfare spending. Giuliano and Spilimbergo 2014 utilizes data from a longer horizon and multiple countries and shows that the effect of recession on beliefs and redistribution preferences are long-lasting. The empirical evidence these papers provide is consistent with my empathy mechanism and provide additional motivation to study this issue.

Aside from empathy, there are other ways personal experience might matter for preferences toward inequality and redistribution. First, experience might affect the attribution of success to skill versus luck. Piketty 1995 argues that different norms in attributing success to effort versus luck could explain the difference in distributional preferences between the US and EU. Second, experience might affect views on the difficulty of social mobility. Benabou and Tirole 2006 shed theoretical insights that redistribution differences in Europe and the US might be attributed to differently perceived social mobility rates. Third, experience might affect distributional preferences. Cohn et al. 2019 has subjects from top income percentiles redistribute among two other subjects. They found those self-made high-income subjects have higher tolerance for inequality caused purely by luck. These perspectives are complementary to empathy, which is the focus of this paper.

Finally, the economics of empathy studied in this paper relates to the literature of pos-

itive welfare economics, which studies how people make welfare judgments about others. This paper speaks to both how people assess the welfare of individuals (e.g. Ambuehl et al. 2021), and to how they aggregate the welfare of others and of themselves (e.g. Andreoni et al. 2020, and Ambuehl and Bernheim 2021). Ambuehl et al. 2021 shows that when intervening in other people's choices, subjects project their own time inconsistent preference onto that of others and believe removing an impatient option would benefit the other party. Andreoni et al. 2020 find subjects most commonly conduct a time-consistent decision-making rule to achieve fairness in allocations in an environment with uncertainty. Ambuehl and Bernheim 2021 find strong evidence that subjects use cardinal utility to aggregate group members' ordinal preferences. This paper highlights motivated reasoning as an additional obstacle for people to understand others accurately–motivated reasoning driven by self interests leads to biased beliefs and reduced empathy.

3 Theoretical Framework

I construct a multi-self model following the spirit of Benabou and Tirole 2006, and I have the agent choose beliefs similar to the mechanism in Brunnermeier and Parker 2005.

3.1 A Model of Empathy and Redistribution

A wealthy individual with initial resources X must choose a transfer, t to a poor individual whose initial resources I normalize to zero for convenience. Thus, the rich individual consumes $x_1 = X - t$, while the poor individual consumes t. As in Andreoni 1995, altruistic preferences motivate the transfers. Several additional assumptions differentiate this model from the standard setting.

First, the wealthy individual is imperfectly informed about the relationship between utility and resources. The wealthy individual has experienced various consumption levels and their associated utilities with the following data-generating process.

$$u_i = e^{\beta} \cdot \frac{c_i^{1-\gamma}}{1-\gamma} \cdot e^{\varepsilon_i}$$

where $i \in \{1, 2, ..., n\}$ enumerates his consumption experience, β is the level parameter, γ is the curvature parameter, and ε is the independent and identically distributed mean-zero noise. Let $\theta \equiv (\beta, \gamma)$ be a vector that summarizes both the level and the curvature parameters of the utility generating process. Then the expected utility function is:

$$u(c, \underbrace{\theta}_{\equiv(\beta,\gamma)}) = e^{\beta} \cdot \frac{c_i^{1-\gamma}}{1-\gamma} \tag{1}$$

The wealthy individual has an altruism level $a \in [0, 1]$. Their intrinsic utility function

is a weighted sum of their own expected utility from consuming $X-t$ and the other party's utility from consuming t, weighted by altruism, a:

$$W(t,\theta) = u(X-t,\theta) + a \cdot u(t,\theta)$$

To focus on the process of belief formation, I assume beliefs about the utility function, $\hat{\theta}$ are formed in period 1, and the transfer, t, is selected in period 2.

Second, the individual also experiences different visceral emotions. I consider guilt and temptation and these visceral emotions lowers the utility function. On the one hand, the wealthy individual may feel guilty about how selfish his redistribution decision is in the eye of a more altruistic person. The guilt function is:

$$G(t,\theta) = \max_{\tilde{t}}[u(X-\tilde{t},\theta) + \tilde{a} \cdot u(\tilde{t},\theta)] - [u(X-t,\theta) + a \cdot u(t,\theta)]$$

where $0 \leq a < \tilde{a} \leq 1$.

On the other hand, the wealthy individual may also feel tempted to redistribute little to the other party and to consume more themselves. Following Gul and Pesendorfer 2001, this feasible but not chosen outcome is tempting and reduces the individual's utility by the temptation function:

$$\begin{aligned} T(t,\theta) &= \max_{\tilde{t}} u(X-\tilde{t},\theta) - u(X-t,\theta) \\ &= u(X,\theta) - u(X-t,\theta) \end{aligned}$$

Third, the weight the individual places on guilt or temptation in the moment of redistribution may be excessive from the perspective taken away from the moment. [11] The objective functions in the two periods are:

Period 1:

$$\max_{\hat{\theta}} W(t(\hat{\theta}),\theta) - \lambda_1 G(t(\hat{\theta}),\theta) - \nu_1 T(t(\hat{\theta}),\theta) - \phi D(\hat{\theta}|\{(u,c,)\}) \tag{2}$$

Period 2:

$$\max_{t} W(t,\hat{\theta}) - \lambda_2 G(t,\hat{\theta}) - \nu_2 T(t,\hat{\theta})$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, $0 \leq \nu_1 \leq \nu_2 \leq 1$. I am using weak inequalities here to include the scenarios when the individual is equally emotional in both periods. Later in this section, I will show that strict inequalities in emotional sensitivities give rise to time-inconsistency.

Fourth, as in Equation (2) above, the individual chooses her belief, $\hat{\theta}$, in period 1 to

[11] One may assume the other way around–if one assumes the wealthy individual is more emotional when thinking ahead of time than when making the decision, the results of my model will reverse. However, this alternative assumption does not seem realistic.

maximize her utility, given some prior belief. This prior belief is what best represents their own experience and would be their belief if they did not conduct motivated reasoning. The posterior belief they choose is an excuse to justify their anticipated redistribution in the next period. Here I write the prior belief simply as the "truth", but this is not necessary.[12]

Fifth, as in Equation (2), the individual has prior experience that provides observations on utility, $\{(u,c)\}$. The individual's chosen belief $\hat{\theta}$ leads to cognitive dissonance, D, based on inconsistencies with that experience of theirs, $\{(u,c)\}$.

3.2 Redistribution Decisions Given Empathy

Consider a wealthy individual allocating consumption between himself and another party given a set of beliefs on the utility of consumption. The wealthy individual's Period 2 objective is shown in Equation (3) below. His belief about consumption utility is characterized by $\hat{\theta}$ in blue, which is taken as fixed and given for now. His choices of transfer is denoted by t in red, with t being the part of transferred to the other party, and $X-t$ being the part left to be consumed by themselves.

Solve from backwards. In Period 2:

$$\max_{t} \quad \underbrace{W(t,\hat{\theta})}_{[u(X-t,\hat{\theta})+au(t,\hat{\theta})]} - \lambda_2 \cdot \underbrace{G(t,\hat{\theta})}_{\max_{\tilde{t}}[u(X-\tilde{t},\hat{\theta})+\tilde{a}u(\tilde{t},\hat{\theta})]-[u(X-t,\hat{\theta})+\tilde{a}u(t,\hat{\theta})],\ 0<a<\tilde{a}<1} - \nu_2 \cdot \underbrace{T(t,\hat{\theta})}_{\max_{\hat{t}} u(X-\hat{t},\hat{\theta})-u(t,\hat{\theta})} \tag{3}$$

$$t^* = \frac{A^{1/\hat{\gamma}}}{A^{1/\hat{\gamma}}+1} \cdot X \tag{4}$$

where

$$A \equiv \frac{a+\lambda_2\tilde{a}}{1+\lambda_2+\nu_2} \in (0,1) \tag{5}$$

3.3 Optimal Empathy

From the optimal transfer given an empathy level in Equation (4), I can solve for the optimal empathy level that maximizes Period 1 objective. Note that my key assumption is that the wealthy individual is time inconsistent in their emotions, and they are sophisticated about such inconsistency. Without time inconsistency, there would be no motivated reasoning. Formally:

Proposition 3.1. *Assume the cognitive dissonance term D increases whenever $|\hat{\gamma}-\gamma|$ increases, when the wealthy individual has the same sensitivities to their visceral feelings*

[12] When the wealthy individual has limited experience, per the data-generation-process Equation (1), the prior that they infer from their own experience may not equal to the truth. I take the prior as the truth for simplification.

before and during redistribution, i.e. $\lambda_1 = \lambda_2$ *and* $\nu_1 = \nu_2$*, the optimal chosen belief of the consumption utility curvature (empathy) is unbiased with* $\hat{\gamma}^* = \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) = t^*(\gamma)$.

I prove Proposition 3.1 in Appendix A.7. The gist of the proof is that, with the same weights on emotions in both periods, the objective functions for the wealthy individual is the same in both periods when we do not consider the cognitive dissonance term, D. Choosing a belief in the first period and choosing an action conditional on this belief in the next period to achieve the same goal can be combined into a single problem. That is, there is no reason for the individual to involve in self-deception. Furthermore, the unbiased belief (and its corresponding action) also minimizes the cognitive dissonance term, D. Proposition 3.1 therefore establishes that my following results on motivated reasoning come from time inconsistency.

Next, I show that the direction of motivated reasoning depends on which emotion is more dominant. In Proposition 3.2, I show that to prevent one's future self from feeling guilty about redistributing less than what is ideal for a more altruistic person, the wealthy individual chooses to lower empathy in order to justify less redistribution. As a result, they redistribute less than what they would redistribute without such motivated reasoning. On the contrary, in Proposition 3.3, I show that to prevent one's future self from feeling tempted to not redistribute, the wealthy individual chooses to increase their empathy level to preempt such temptations. As a result, they redistribute more than when they could not motivated reason.

Proposition 3.2. *Assume the cognitive dissonance term D increases whenever* $|\hat{\gamma} - \gamma|$ *increases, when only temptation is turned off, the optimal chosen belief of the consumption utility curvature (empathy) is downward biased with* $\hat{\gamma}^* < \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) < t^*(\gamma)$.

Proposition 3.3. *Assume the cognitive dissonance term D increases whenever* $|\hat{\gamma} - \gamma|$ *increases, when only guilt is turned off, the optimal chosen belief of the consumption utility curvature (empathy) is upward biased with* $\hat{\gamma}^* > \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) > t^*(\gamma)$.

I prove Propositions 3.2 and 3.3 in Appendices A.5 and A.6.

3.4 Personal Experience As A Cognitive Constraint

One feature of my model that's a little peculiar is that the period 1 self knows the truth independently of the observations from experience. In other words, their observations aren't informing the prior – they're just driving the cognitive dissonance.

I assume that the dissonance penalty depends on the fit of the wealthy individual's chosen belief to experience, as summarized by mean squared error (MSE) in a regression. This assumption is reasonable because people will find it harder to maintain a belief that is at odds with the experience they recall.

I reduce the two slope and curvature parameters, $\hat{\beta}$ and $\hat{\gamma}$, into just one curvature parameter, $\hat{\gamma}$. The reason is that given any consumption-utility data set, $\{u,c\}$, and any chosen curvature parameter, $\hat{\gamma}$, the slope parameter $\hat{\beta}$ can auto-adjust to achieve a best fit for the data. In this way, the cognitive dissonance term is a function of the chosen empathy level, $\hat{\gamma}$, alone.

I formalize this intuition in Propositions 3.4 and 3.5 below.

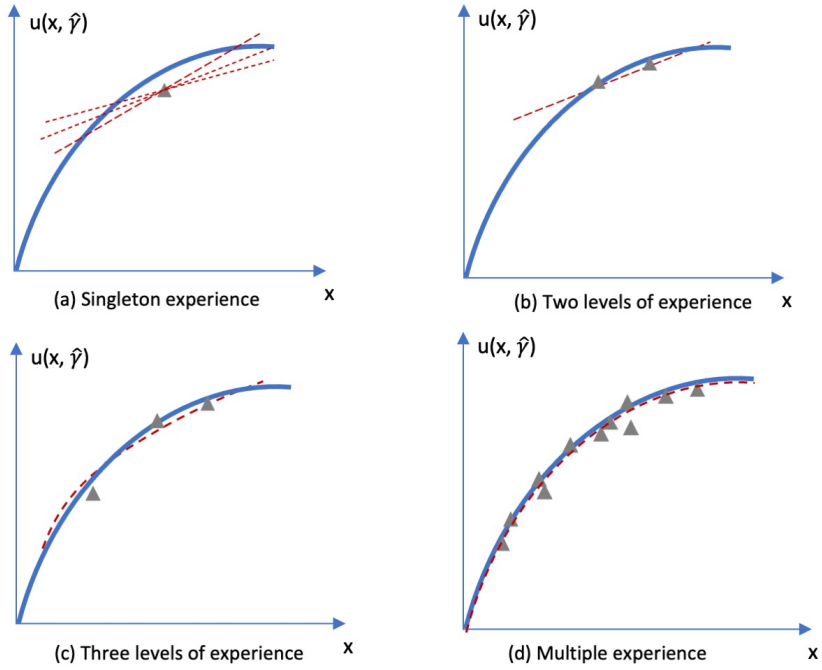
Proposition 3.4. *When $\nu_1 = \nu_2 = 0$, $\hat{\gamma}^*$ is increasing in $var(c)$, $t(\hat{\gamma}^*)$ is increasing in $var(c)$.*

Proposition 3.5. *When $\lambda_1 = \lambda_2 = 0$, $\hat{\gamma}^*$ is decreasing in $var(c)$, $t(\hat{\gamma}^*)$ is decreasing in $var(c)$.*

I prove Proposition 3.4 and 3.5 in Appendix A.8.

The intuition is simple and I illustrate it in Figure 1 below. When the wealthy individual has only one level of consumption experience, he can believe in any arbitrary marginal utility without penalty (see Figure 1, Panel (a)). When his experience is fixed at two levels, he can still draw a line between those two data points and believe that marginal utility is the same everywhere, so that he wouldn't need to redistribute to the poor (see Figure 1, Panel (b)). When the person has three or more levels of consumption experiences, choosing to believe in a linear utility function costs cognitive dissonance (see Figure 1, Panel (c) and (d)). The more varied his experience is, the larger this cognitive dissonance term would be for a fixed bias in beliefs.

Figure 1: Model Intuition: More varied consumption-utility experience constrains motivated reasoning



The y-axis is consumption utility, whereas the x-axis is consumption level. The solid blue curve is the true utility function. The dashed red curves are beliefs about the utility function that may be manipulated to deviate from the true one. The gray triangles are data points in the person's consumption-utility database that is generated from the utility function but with some noise. Panel (a) shows when the person has only a single consumption experience; Panel (b) shows when the person has two levels of consumption experience; Panel (c) shows when the person has three levels of consumption experience; Panel (d) shows when the person has varied consumption experience.

4 Experiment Design

4.1 The Clicking Task

In the theory, a wealthy individual with certain consumption history is now endowed with a large amount of consumption, and he's redistributing some to another poor party for efficiency. To simulate the theory, exogenous experiences in the laboratory experiment should have the following characteristics:

First, I should be able to vary the extent of experience with this task in large scale, which rules out money experiences–it's unlikely that I am able to dramatically change subject's wealth through participation fee or bonuses alone. On the other hand, with real effort tasks, I can create "wealthy" individuals assigned easy tasks and "poor" individuals assigned hard tasks with larger cross-subject differences in experience. Second, analogous to decreasing marginal utility of consumption in the model, the real effort task needs to have increasing marginal cost of effort (due to muscle fatigue and/or boredom). Under this setup, reallocating one unit of task from the poor who are assigned much work to the wealthy who are assigned little work improves efficiency. Third, there needs to be

no learning curve, that is, the task should be easy enough that everyone can immediately take it up, yet it should also be unpleasant enough that nobody enjoys doing it at all. The point here is to make this task pure effort and a negative consumption experience to everybody, without introducing any confounds in learning abilities or heterogeneous tastes for activities.

Therefore, I use a clicking task. In the clicking task, subjects are required to click in a red square with a minimum speed of 4 times per second. The computer monitors their speed and if they fail to reach the minimum speed, they would need to do it again until they meet the speed requirement. I have everyone complete a trial task with 50 seconds of clicking at the beginning to familiarize them with the task interface. For an individual to be "wealthy" in my world of clicking, they are lucky will be assigned only 100 seconds. For them to be "poor", they are unlucky and will be assigned 500 seconds. The tedious repetitive clicking should be boring and tiring enough that generates the desired convex cost of effort, yet its mild nature would not physically harm subjects.

4.2 3 Stages

There are 3 stages in the longitudinal experiment, the first two stages are experience stages that lays the foundation for testing empathy and redistribution by creating either uniform experience with the clicking task or varied experience. The final stage is the redistribution stage and generates the main beliefs and allocations data that I am interested in. The mandatory time break in between two consecutive stages is between 1 hour and 24 hours, so that subjects get enough rest from previous clicking experiences, yet they still remember how it feels vividly in their mind.

4.2.1 Stages 1 & 2: Experience Stages

In Stage 1, all subjects first do 50 seconds of clicking to familiarize themselves with the task, and then all of them are assigned 100 seconds. Hence, all subjects are "wealthy" and get the easy assignment in Stage 1. In Stage 2, half of the subjects are still assigned the easy task, 100 seconds, whereas the other half are assigned the hard task, 500 seconds. For the first half, their experience with clicking in the two experience stages are uniformly "wealthy-wealthy". For the second half, their experience with clicking are varied "wealthy-poor".

4.2.2 Stage 3: Redistribution Stage

In Stage 3, all subjects are "wealthy" and assigned the easy 100 seconds task. Nevertheless, half of them do not know that they are wealthy in Stage 3 so that I can elicit their honest opinions about relative marginal costs of clicking 500 seconds versus 100

seconds before redistribution, see Section 4.3.1 for details about the motivated reasoning treatment. All subjects in Stage 3 are first asked about their beliefs about how unpleasant it is for a poor subject assigned 500 seconds to click for the final 50 seconds, as well as how unpleasant it is for a lucky subject assigned 100 seconds to click an additional 50 seconds. See Section 4.5.1 for details about how I elicitealtd these beliefs. Finally, all subjects in Stage 3 are asked to redistribute some clicking time from their poor partner assigned 500 seconds to themselves. I use a Titration-BDM mechanism to find the reserve value when subjects help their partner, see Section 4.5.2 for details.

4.3 A 2-by-2 Design

4.3.1 The Cognitive Dissonance Amplification (CDA) Treatment

I vary the sequence in between telling subjects that they are "wealthy" in the redistribution stage and asking their beliefs about relative marginal costs. When subjects do not know that they are "wealthy" or that this stage requires redistribution from themselves to their partners, subjects should have little motivation to manipulate their beliefs about how unpleasant doing the hard task versus the easy task is. And after they have stated their beliefs, any additional revisions to their beliefs would increase their cognitive dissonance. Hence, in the CDA treatment group, these stated beliefs should be correlated with their subsequent actions. For the CDA control group, subjects know that they are assigned the short task in the redistribution stage and that they will need to redistribute to their poor partner later. This process mimics real world decision making where individuals have motivations to bias their beliefs in order to justify less redistribution to their partner later.

For subjects in the CDA control group, I conduct the redistribution stage in the following sequence: (1) I inform subjects that they will do the easy task and their partner will do the hard task, and that they will have a chance to redistribute some of their partner's task to themselves later. (2) I elicit subject's beliefs about marginal costs doing the hard versus easy task. (3) I have subjects redistribute tasks among themselves. They can make their partner's task shorter by making their own task longer.

For subjects in the CDA treatment group, I only change the order of the three steps for the redistribution stage while keeping contents of each step identical to those in the CDA control group. That is, subjects were first elicited beliefs about marginal unpleasantness of the long versus short tasks without knowing that they are assigned the short task in this redistribution stage, or knowing that they will need to redistribute later. Without the context of redistribution and the position of them being wealthy, subjects are less likely to motivate their beliefs. After their beliefs are collected, I then inform them about their task assignment in this stage and prelude the upcoming redistribution step. Finally, they redistribute tasks exactly as those in the motivated reasoning treatment do.

4.3.2 The Experience Variation Treatment

I also manipulate the history of relevant experience of the real effort task. In the uniform experience treatment, subjects are assigned both easy tasks in the two experience stages. In the varied experience treatment, subjects are also assigned the easy task in the first experience stage, but then the hard task in the second experience stage. In this way, I study how the variation of experience, holding the number of experience the same, impact empathy and redistribution.

4.4 Difference-in-Difference Method

A simple comparison across subjects with different experiences does not necessarily shed light on motivated reasoning, because their differing levels of "wealth" for the entire experiment may affect their levels of generosity. Similarly, a simple comparison across subjects with different dissonance constraints (arising from the timing of information concerning the transfer) does not necessarily shed light on motivated reasoning, because the treatment may serve as a "nudge" to make generous transfers. However, diff-in-diffs eliminates both of those confounds, and it speaks to the central theoretical prediction that experience weakens motivated reasoning.

I generate the following hypothesis to be tested under the experiment design: Compared to subjects with more varied clicking experience prior to the redistribution, subjects with uniform clicking experience are more subject to motivated reasoning in their empathy beliefs and redistribution decisions.

4.5 Measures

4.5.1 Measure of Beliefs – Relative Marginal Unpleasantness

Regardless of whether subjects are in the motivated reasoning treatment, I use the same set of measures to elicit their beliefs about marginal costs of clicking 500 seconds versus 150 seconds. They are told that there are two other participants in this study, A and B, who are assigned 150 seconds and 500 seconds of clicking, respectively. I ask them how unpleasant it is for A and B to complete their last 50 seconds of clicking using a scale from 0 to 30. That is, for person A assigned 150 seconds, how unpleasant person A think it is to click from 100 second to 150 second. For Person B assigned 500 seconds, how unpleasant person B think it is to click from 450 seconds to 500 seconds. I also explain to them that unpleasantness includes boredom and muscle fatigue from of clicking repetitively.

Subjects are incentivized for accuracy in their guesses. Since unpleasantness of the task is not observable, I am not incentivizing them for the accuracy with which they report unpleasantness. Rather, I am assessing subjective unpleasantness from others, and then

incentivizing them to accurately report their belief about what others will say. I am then assuming that the belief about what others will say is a good stand-in for the belief about unpleasantness. Those whose guess is within 3 of person A or B's actual answers will receive a small bonus.

I arrange the two slider questions on the same page, so subjects would need to think about the marginal costs of clicking 150 seconds and 500 seconds together. This type of thinking is directly related to the tradeoff subjects engage in once they know that they are the wealthy individual assigned 100 seconds, and that they can help their poor partner by clicking 50 seconds more. Therefore, the wealthy individual's redistribution decision should be based on the beliefs I measured and reflect whether or not they are in the motivated reasoning treatment when I elicit their beliefs.

4.5.2 Measure of the Willingness to Redistribute – Titration BDM

I follow Mazar et al. 2014 and use a Titration BDM mechanism to find the minimum reduction in the poor partner's clicking time at which the wealthy individual's is willing to help by volunteering to click 50 seconds more. The process starts by asking subjects to write down their reserve value: How many seconds at least should their partner's clicking time be reduced so that they are willing to help? Suppose subjects enter a number, X. I then ask them two confirmation questions. Are they willing to help when their partner's time reduction is X+10 seconds? And are they not willing to help when their partner's time reduction is X-10 seconds? Suppose subjects answer incorrect to at least one of the two confirmation questions, they will receive advice to revise their reservation value. And then the process repeats itself until subjects answer correct to both the confirmation questions. After I find the reservation value, I provide subjects with a multiple price list pre-filled with whether or not to help the partner at different amounts of time reduction. This Titration-BDM process has the following advantages: (i) it helps subjects understand the contingent reasoning implied by a price list, and (ii) it eliminates boundary effects resulting from the minimum and the maximum on a price list.

I then construct a measure of helpfulness from the redistribution decision this wealthy individual makes.

$$\text{Helpfulness} := \frac{\text{the additional 50 seconds the wealthy individual needs to click to help}}{\text{the minimum reduction in partner's clicking time at which the wealthy individual helps}}$$

The numerator is fixed at 50 seconds, whereas the denominator is the reserve value in the reduction of partner's clicking time and ranges from 25 seconds to 350 seconds. Therefore, the helpfulness measure ranges from 0.14 to 2. The larger it is, the more generous the wealthy individual is in redistribution.

4.5.3 Measure of Altruism as a Control

I also measure altruism and contrasts empathy with it in the experiment. After subjects are done with everything else in the redistribution stage, they play a dictator game with another stranger in the experiment. The wealthy individual is told that there is 10% chance that they will be selected to receive an additional bonus from the dictator game. I then ask them out of $5, how much they would share with another participant in this study. I ask this question four times, using various dollar amounts and various types of lottery tickets to split in between the wealthy individual and another participant. I then take the average percentage the wealthy individual shared with the other participant across these questions as a measure of altruism. My altruism measure has a value between 0 and 1, and the larger it is, the more altruistic the wealthy individual is.

5 Experiment Results

5.1 Sample Collection

5.1.1 Demographic Representativeness

I recruited 609 subjects for Stage 1 on Prolific in September 2021. They need to reside in the US, hold an undergraduate, master, or doctorate degree, and in between 25 and 56 years old to qualify for the experiment. These educational restriction helps me to zoom in American elite's attitudes toward redistribution. The age redistribution makes sure that subjects have enough physical capacity to click quickly for an extended period of time, and that they have formed a world view in terms of inequality and redistribution. My final sample has a median age of 40 and a Male to Female ratio of 6:4.

5.1.2 Attrition Rate

Given the longitudinal nature of the 3-stage experiment, there is some attrition of the sample both within stage and across stages. 5% of my sample who started Stage 1 did not finish it. 15% of those 609 subjects I recruited for Stage 1 did not come back for Stage 2. While in Stage 2, subjects are assigned to click for 500 seconds ("poor") or 100 seconds ("wealthy"). 5% of the subjects in the poor treatment didn't finish Stage 2, whereas 1% of those in the wealthy treatment dropped out during Stage 2. 8% of subjects in the poor treatment didn't come back for Stage 3, whereas 6% of those in the wealthy treatment did not come back. There are no attrition during Stage 3. Since I arranged the majority of payments to arrive after Stage 2 and after Stage 3 to incentivize continued participation, the attrition rates are overall very low and similar for the two groups, so that there isn't evidence of differential selection.

After attrition, I am left with 438 subjects. Among these, 2% of them have duplicate IP addresses and are thrown out of the sample. 16% have inconsistent redistribution choices, that is, they would choose to help their partner when the time reduction in partner's clicking time is low, but they would not help when the time reduction is high. These inconsistent choice data are low quality and I drop these observations. Finally, 5% of the sample spent less than 10 seconds on the page with the two belief elicitation questions. Given the amount of information they need to process on that page, these submissions turn the page too quickly and are thrown out. The final usable data size is 334. See Table 1 for a summary of the attrition rates.

Table 1: Data Attrition Rates Over Three Stages

	Num of Obs		Attrition Rate	
	Poor (Stage 2)	wealthy (Stage 2)	Poor (Stage 2)	wealthy (Stage 2)
Started Stage 1	609			
minus dropped out in Stage 1	32		5%	
minus dropped out between Stages 1 and 2	89		15%	
Started Stage 2	263	224		
minus dropped out in Stage 2	13	2	5%	1%
minus dropped out between Stages 2 and 3	20	13	8%	6%
minus dropped out during Stage 3	0	0	0%	0%
equals finished Stage 3	229	209		
Total finished Stage 3	438			
minus duplicated IPs	10		2%	
minus inconsistent choices in MPL	72		16%	
minus less than 10 sec for belief questions	22		5%	
Total usable data size	334			

5.2 Different-in-Difference Test

5.2.1 Baseline Effects

Before I run the difference-in-difference test, I first interpret the two baseline effects, experience variance treatment and motivated reasoning treatment, as well as their respective confounding effects.

First, when I vary subject's experience with clicking in the first two stages, I also vary their pay per hour in this experiment. Those with uniform experience are assigned 100 seconds throughout the first two experience stages, whereas those with varied experience are assigned 100 seconds and 500 seconds. Therefore, those with uniform experience get paid the same as those with varied experience in this experiment while doing a lot less work. This pay per hour difference may play a role in their redistribution decisions through wealth effect. That is, other than the empathy and motivated reasoning channel, wealth effect may also play a role in changing redistribution between the two experience treatment. Per the wealth effect, subjects in the "wealthy-wealthy" treatment should be more helpful than those in the "wealthy-poor" treatment, which is against my hypothesized direction of experience variation effect. Therefore, the baseline effect of the varia-

tion of experience on redistribution generosity should be an *under-estimation* due to the offsetting wealth effect.

Furthermore, the experience treatment may also have confounding effect on beliefs aside from the empathy channel. Even though subjects in both uniform and varied experience treatments have the same *information* regarding the length and intensity of clicking 500 seconds, their experience-based knowledge of clicking 500 seconds are different. Therefore, we cannot read too much from the baseline differences in the beliefs of the two groups about marginal costs. The fact that experience effect alone has ambiguous effects on beliefs further stresses the importance of interacting it with the motivated reasoning channel. Only by studying the interaction effect, can we zoom in on the empathy effect and how it affects the beliefs of subjects with different experiences differently.

Second, in the CDA treatment, I manipulate when subjects know they are wealthy individual in Stage 3 faced with a redistribution decision. If they don't know they are wealthy, or that there is redistribution coming up, their guesses about the marginal costs are more likely to be honest and less likely impacted by selfish motivations. The undesired side effect of this treatment is a nudging effect, that is, subjects who are informed of the redistribution purpose of this experiment might behave more generously than subjects who did not know.

Due to this nudging effect, my estimation of the baseline difference between CDA treatment and control in both beliefs and decisions should be an *under-estimation* of the hypothesized motivated reasoning effect. This is because nudging effect makes subjects more generous, whereas motivated reasoning effect allows them to be more selfish. The former effect cancels out some of the motivated reasoning effect for the baseline differences, making it harder to detect baseline effect of the motivated reasoning treatment and calling for attention on its interaction effect with experience variation.

5.2.2 The Interaction Term

Given the confounds of the baseline effect, I take a difference-in-difference strategy and focus on the interaction term between experience variation and motivated reasoning treatments. Difference-in-difference subtracts out the confounding effects because confounds for one treatment doesn't impact the other treatment.

Confounds for experience variation does not affect the CDA treatment. Suppose there are two subjects–subject Y is in the CDA control and knows that he is wealthy in Stage 3 and needs to redistribute later when he tells me his guesses about the marginal costs. Subject N is in the CDA treatment and has no idea about her assignments and positions in Stage 3 when she guesses about the marginal costs of clicking 500 seconds versus 100 seconds. Both subject Y and N have the same kind of wealth effect and experience-based knowledge effect from certain clicking experience in the first two stages. These confounds for experience treatment do not change whether or not subjects have selfish motivations

when they form beliefs about how difficult the task is at various lengths, and therefore, do not confound the motivated reasoning treatment.

Confounds for CDA treatment does not affect the experience variation treatment. The reason is simple: CDA treatment happens in Stage 3, whereas experience variation treatment happens in Stages 1 to 2.

To summarize, difference-in-differences eliminates both confounds for the baseline effects and identifies the channel I am interested in in this paper–how experience variation constrains motivated reasoning and promotes empathy in beliefs and generosity in redistribution.

5.3 Redistribution Results

Redistribution results are summarized in Figure 2. The bars measure how helpful the wealthy individual is in redistributing some of their partner's tasks to themselves, as defined in Section 4.5.2. A taller bar means the wealthy individual is more helpful.

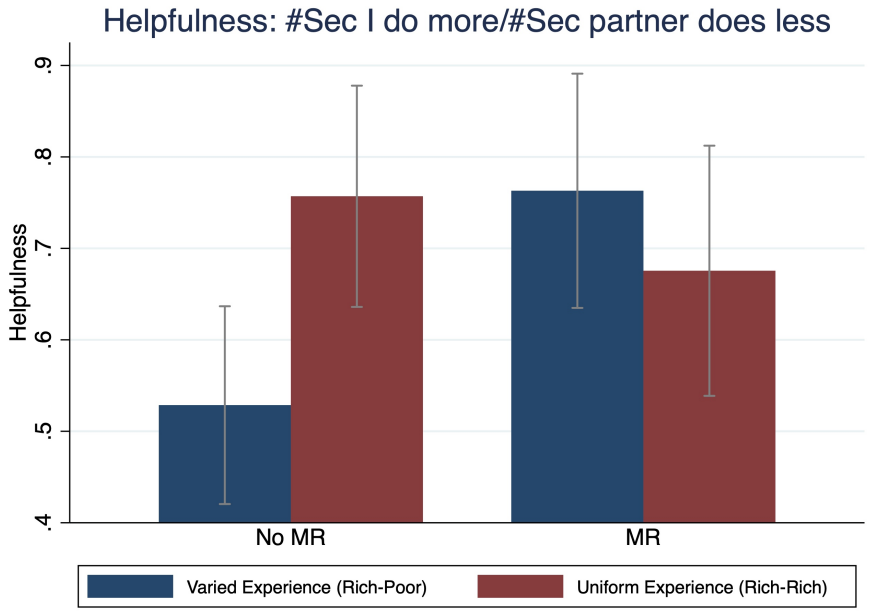
Starting from the left two bars where subjects are in the baseline no motivated reasoning treatment, subjects who have always been wealthy in the two previous stages are more generous than those who have once been poor. This is in line with my analysis in Section 5.2.1 that subjects with uniform experience get a higher pay per hour in this experiment than those with varied experience and should be more generous due to wealth effect. The baseline difference is an underestimation of the effect of varied experience on redistribution through empathy due to this confound, and I use it as a benchmark to compare to the case with motivated reasoning.

Comparing the right two bars to the left two bars, I am able to zoom in to the interaction effect between motivated reasoning and experience variation. The difference between the red and blue bars reverses signs under the motivated reasoning treatment compared to that under the no motivated reasoning treatment. The wealthy individual with uniform experience is now less helpful than those with varied experience once I turn on the motivated reasoning channel. The result is consistent with the model proposition that the wealthy individual with uniformly wealthy experience before are more subject to selfish motivated reasoning in their redistribution decision than those with varied experience.

As a side note, one may question why the wealthy individual with varied experience are becoming more generous when they are in the motivated reasoning treatment than when motivated reasoning is shut down. In line with the analysis in Section 5.2.1, the baseline effect of motivated reasoning is confounded with experimenter demand effect. This further stresses the importance of focusing on the difference in difference effect–the difference between the two right bars of around 0.1 decrease in helpfulness is an *underestimation* of the effect of empathy on redistribution due to the offsetting experimenter demand effect. The actual effect should be the decrease in right bars from the left to

the right, *plus* the increase in blue bars from the left to the right, totalling around 0.3 in helpfulness that can be attributed to the lack of empathy arising from selfish motivated reasoning and less varied experience with clicking.

Figure 2: Helpfulness in Redistribution: The Interaction Effect between Motivated Reasoning and Experience Variation



Helpfulness is defined as the exchange rate in between the wealthy individual's extra clicking time and the minimum reduction in their poor partner's clicking time at which the wealthy individual is willing to help. The numerator is fixed at 50 seconds, whereas the denominator is the reserve value in the reduction of partner's clicking time and ranges from 25 seconds to 350 seconds. Therefore, the helpfulness measure ranges from 0.14 to 2. The larger it is, the more generous the wealthy individual is in redistribution.

I then input the data into a simple difference-in-difference regression model as below.

$$\text{Helpfulness} = a_1\text{MR} + a_2\text{Uniform Experience} + a_3\text{MR} \times \text{Uniform Experience}$$

$$+ a_4\text{Altruism control} + a_5\text{Demographic Control}$$

The dependent variable is the same as in Figure 2, helpfulness, that ranges from 0.14 to 2, and the larger it is, the more generous the wealthy individual is in redistribution. MR is a dummy variable with a value of either 0 or 1. If MR=0, subjects are in the CDA treatment group and they do not know about redistribution when I measure their beliefs; if MR=1, subjects are in the CDA control group and they are informed about redistribution at the beginning of Stage 3. Uniform Experience is also a dummy variable taking the value of either 0 or 1. If Uniform experience=1, subjects are "wealthy-wealthy" in the previous two experience stages, and have only done the easy 100 seconds clicking task; if Uniform experience=0, subjects are "wealthy-poor" in the two experience stages and did

both the 100 seconds clicking and the 500 seconds one. MR×Uniform Experience is the interaction term between the motivated reasoning treatment and the experience variation treatment.

Results of the regression are in Table 2. Columns (1) and (2) show the baseline effects of motivated reasoning treatment and experience variation treatment, respectively. As expected, neither of the two baseline effects is statistically significant due to the confounds cancelling out the empathy effect. In fact, the baseline effects in Column (3) show that subjects in the motivated reasoning treatment redistribute more than those in the no motivated reasoning treatment, and that subjects with uniform "wealthy-wealthy" experience are more generous than those with "wealthy-poor" experience. The former is due to experimenter demand effect, and the latter is due to wealth effect. See these confounds analyzed in more details in Section 5.2.1.

The interaction term between the two treatment is key for identifying the empathy channel. Once I apply the difference-in-difference approach in Column (3), the interaction term between the two treatment is now significantly negative as expected. That is, compared to subjects with varied experience with the clicking task, subjects with uniform experience is more subject to selfish motivated reasoning and redistribute less generously as a result. The interaction effect remains robust after I add in the altruism measure and demographics as control.[13] I stress this as the primary finding of this paper.

Also, adding the altruism control in does not subsume the empathy effect. This demonstrates why it is necessary that I separate empathy from altruism in the first place. In fact, the empathy effect captured by the interaction term is nearly half as large as the altruism effect, making empathy a sizable economic effect worthy of more research. Still, taking a step back, one would need to design another set of exercise quantifying the effects of empathy versus altruism in redistribution should they want to make a claim on their relative magnitudes. This paper only serves as a lead in this direction.

Besides, younger subjects are on average more helpful than older subjects in my sample, which is not surprising given the physical nature of the clicking task. Younger subjects are more able to click quickly for an extended period of time than older subjects, making redistribution more affordable to them than to the older subjects. Finally, female are more helpful than male in my sample, redistributing more of their partner's tasks to themselves than men, and the difference is marginally significant.

[13] I measure altruism using incentivized dictator games at the end of the multi-stage experiment. Section 4.5.3 documents how I construct the altruism measure in more detail.

Table 2: Helpfulness in Redistribution: The Interaction Effect between Motivated Reasoning and Experience Variation

	Helpfulness			
	(1)	(2)	(3)	(4)
MR	0.07		0.23***	0.20**
	(0.06)		(0.09)	(0.08)
Uniform Experience		0.07	0.23***	0.18**
		(0.06)	(0.08)	(0.08)
MR×Uniform Experience			-0.32**	-0.24*
			(0.13)	(0.13)
Altruism				0.51***
				(0.18)
Young				0.13*
				(0.07)
Male				-0.09
				(0.07)
Observations	333	333	333	322

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Helpfulness is defined as the exchange rate in between the wealthy individual's extra clicking time and the minimum reduction in their poor partner's clicking time at which the wealthy individual is willing to help. The helpfulness measure ranges from 0.14 to 2. The larger it is, the more generous the wealthy individual is in redistribution. MR is a dummy variable with a value of either 0 or 1. If MR=0, subjects are in the no motivated reasoning treatment; if MR=1, subjects are in the motivated reasoning treatment. Uniform Experience is also a dummy variable taking the value of either 0 or 1. If Uniform experience=1, subjects are "wealthy-wealthy" in the previous two experience stages, and have only done the easy 100 seconds clicking task; if Uniform experience=0, subjects are "wealthy-poor" in the two experience stages and did both the 100 seconds clicking and the 500 seconds one. MR×Uniform Experience is the interaction term between the motivated reasoning treatment and the experience variation treatment. Altruism ranges from 0 to 1, the higher it is, the more altruistic the subject is. Young is a dummy variable taking the value of 0 or 1. If Young=1, subject's age is below the median age of the sample (younger than 40); if Young=0, subjects are older than 40. Male is a dummy variable taking the value of 0 or 1. If Male=1, subject is men, otherwise, subject is woman.

Furthermore, I show that the results are driven almost entirely by subjects with median or below median level of altruism. In Table 3, when I only include subjects whose altruism level (measured in a separate dictator game at the end of the study) is equal to or smaller than the median altruism in the sample, the detected interaction effect between experience variation and motivated reasoning remains unchanged with slightly larger economic magnitude.

Table 3: Helpfulness for Subjects with Median or Below Median Altruism

	Helpfulness for Subjects with Median or Below Median Altruism			
	(1)	(2)	(3)	(4)
MR	0.11		0.27***	0.27***
	(0.07)		(0.10)	(0.09)
Uniform Experience		0.02	0.17*	0.17*
		(0.07)	(0.09)	(0.09)
MR×Uniform Experience			-0.30**	-0.27*
			(0.15)	(0.15)
Altruism				0.76***
				(0.23)
Young				0.20**
				(0.08)
Male				-0.05
				(0.08)
Observations	258	258	258	258

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

However, when I look at subjects with median or above medium level of altruism, see Table 4, the results no longer exist. This result is intuitive as empathy is naturally correlated with altruism, and one anticipates less altruistic people exhibit more excuse-seeking behaviors to suit their self interests.[14]

Table 4: Helpfulness for Subjects with Median or Above Median Altruism

	Helpfulness for Subjects with Median or Above Median Altruism			
	(1)	(2)	(3)	(4)
MR	-0.004		0.10	0.06
	(0.08)		(0.11)	(0.11)
Uniform Experience		0.09	0.18*	0.11
		(0.08)	(0.11)	(0.11)
MR×Uniform Experience			-0.19	-0.05
			(0.16)	(0.17)
Altruism				-0.43
				(0.41)
Young				0.07
				(0.09)
Male				-0.05
				(0.09)
Observations	221	221	221	210

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[14] Exley (2016) contains a related result that generous types are less likely to exhibit excuse-seeking behaviors when giving to charity.

5.4 Motivated Beliefs Results

The primary results of the paper is on redistribution, as those choices can be measured more precisely than beliefs and have solid policy implications. In this section I show results on beliefs because they can supplement the redistribution results and provide more insights into the mechanisms.

Figure 3 shows how experience variation impacts empathy when the wealthy individual engages in motivated reasoning about how unpleasant it would be for someone else to click from 100 seconds to 150 seconds. Note that since subjects are assigned to click 100 seconds themselves in Stage 3, and that to shorten partner's clicking time they'd need to click 50 seconds more, going from 100 seconds to 150 seconds would be exactly what it takes for them to offer help to their poor partner.

Starting from the left two bars where subjects are in the no motivated reasoning treatment, subjects with varied experience with clicking guess a slightly higher marginal cost of clicking from 100 seconds to 150 seconds than those with uniform experience. However, when motivated reasoning is turned on, the difference between those two groups flips signs and widens. Compared to subjects who have varied experience, those with uniform experience tend to think doing 50 seconds more is more unpleasant when they have a selfish motivation. Also, note that the difference between the two red bars is an underestimation of the motivated reasoning effect–the decrease in the blue bar from left to right shows the existence of experimenter demand effect, that is, subjects want to appear more helpful once they know about the redistributional purpose of Stage 3. It is the difference in the difference between the red and blue bars from left to right that showcases the effect of a lack of empathy arising from a lack of experience variation.

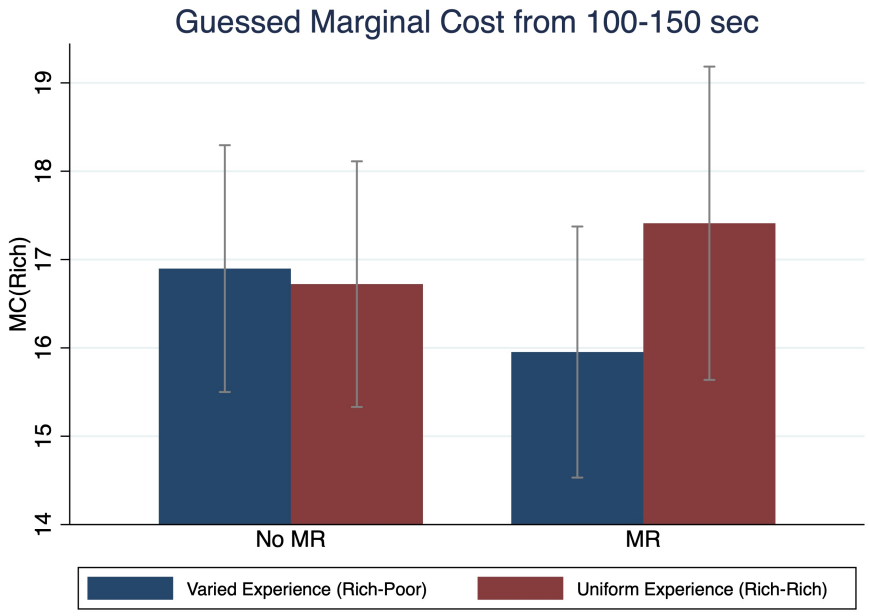
Circling back to the notion of empathy, I want to stress that the selfish motivated reasoning by those with uniform experience in Figure 3 indeed is a lack of empathy. When they are asked about how someone else assigned 150 seconds should feel about the last 50 seconds, they are projecting themselves into that person's situation. The accuracy of this projection is subject to self-serving motivated reasoning. The larger this self-serving bias is in their beliefs, the lower their empathy is towards the poor that need help.

To be more specific, in this paper I model empathy as how much a wealthy individual believes that the utility of consumption is concave with decreasing marginal utility when they are in a redistribution situation and need to form an opinion.[15] By choosing to believe that the marginal utility of consumption at a high level of consumption is large, the wealthy individual denies the concavity of the utility function, believing that themselves are in need of help as much as the poor do, and chooses not to be empathetic toward the poor. Mapping the theory into the experiment result, by choosing to believe that the marginal cost at a low level of clicking is high, those wealthy individuals with uniform

[15] See Section 3 for details.

experience allege that they need help as much as their unfortunate partner, and therefore, choose not to be empathetic when they have a selfish motivation.

Figure 3: Subject's guesses about how unpleasant it is for someone else assigned 150 seconds of clicking to complete their final 50 seconds

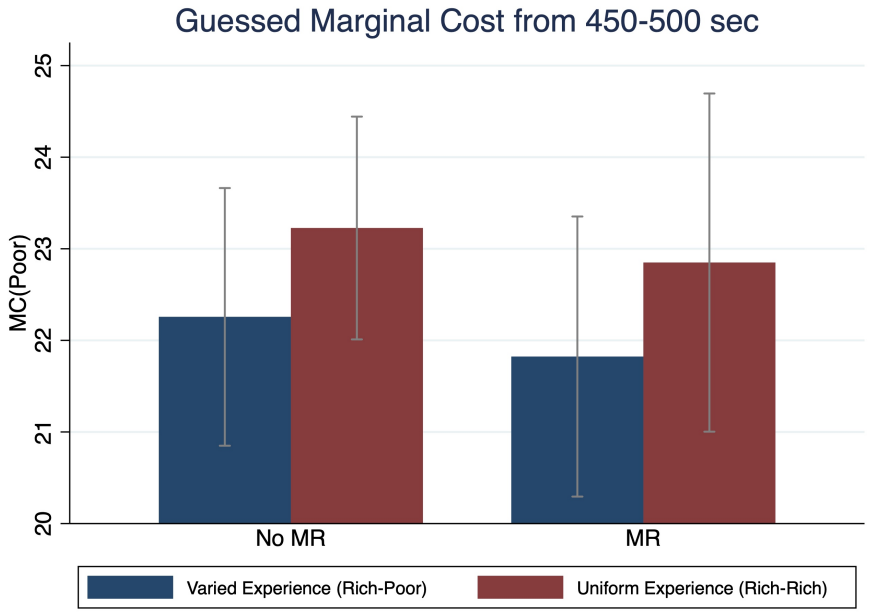


The height of the bars stands for subject's guesses for how unpleasant it for someone else assigned 150 seconds to complete the last 50 seconds from 100 seconds to 150 seconds. The range of this unpleasantness measure is from 0 to 30. The higher the bar is, the more subjects acknowledge that clicking from 100 seconds to 150 second is unpleasant.

In Figure 4, I also show the effect of experience variation and motivated reasoning on subject's guessing about the marginal unpleasantness from 450 seconds to 500 seconds. Compare the difference between the red and blue bars on the left, no motivated reasoning treatment, to that between the two bars on the right, the motivated reasoning treatment, there is not much a difference. Also, when I compare in between the two blue bars or between the two red bars, I find the beliefs about clicking at 500 seconds with motivated reasoning is slightly lower than those without motivated reasoning, but the differences are very small. It seems that motivated reasoning does not have a strong effect on either the varied experience group or the uniform experience group when they are predicting about the cost at 500 seconds. Contrasting this result with that in Figure 3, I have an interesting finding that when one can engage in motivated reasoning about *both* how much in need the poor is *and* how costly it would be for them to help, subjects in my sample choose the latter over the former. That is, the lack of empathy for the group with uniform "wealthy-wealthy" experience manifests itself mostly through alleged higher marginal cost for the wealthy. This observation might be related to the particular feature of the task–clicking 500 seconds is something few people have done before and there are a lot of cross-subject heterogeneity in how well people handle it. Given the noisiness of how difficult it is

clicking for 500 seconds, when I take within group averages for the bar graph, most of the effect is shown through reasoning about the difficulty of clicking from 100 seconds to 150 seconds instead.

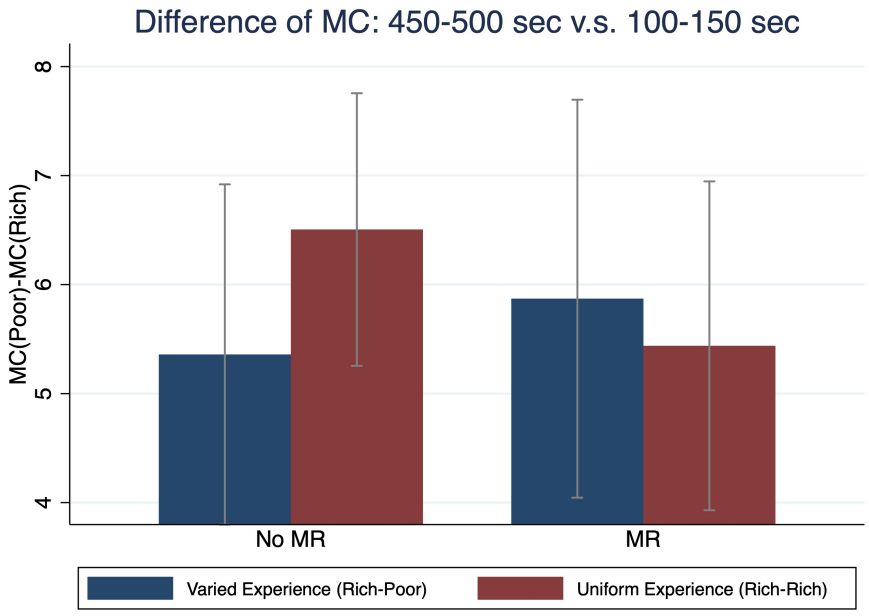
Figure 4: Subject's guesses about how unpleasant it is for someone else assigned 500 seconds to complete their final 50 seconds



The height of the bars stands for subject's guesses for how unpleasant it for someone else assigned 500 seconds to complete the last 50 seconds from 450 seconds to 500 seconds. The range of this unpleasantness measure is from 0 to 30. The higher the bar is, the more subjects acknowledge that clicking from 450 seconds to 500 second is unpleasant.

Summarizing Figure 3 and Figure 4, I show the difference between subject's beliefs of the marginal cost at 500 seconds and that at 150 seconds in Figure 5. For the baseline no motivated reasoning treatment, it is the group with uniform experience, instead of the group with varied experience, that believes in a larger marginal cost difference between clicking at 500 seconds and at 150 seconds. However, when I turn on motivated reasoning, the difference flips. Now the group with uniform clicking experience believes in a smaller marginal cost difference than the group with varied experience. This flipping difference is evidence of self-serving motivated reasoning: for those who have never been assigned the hard task themselves, empathy vanishes right when one most needs it, that is, when one is faced with a redistribution decision. On the other hand, for those who have experienced clicking 500 seconds before, empathy sustains itself even when I turn on selfish motivations for the subjects.

Figure 5: Differences between subject's guesses about the marginal unpleasantness of a 500 seconds clicking assignment versus a 150 seconds one



The height of the bars stands for differences in subject's guess about the unpleasantness of clicking from 450 seconds to 500 seconds for someone assigned 500 seconds and their guess about the unpleasantness of clicking from 100 seconds to 150 seconds for someone assigned 150 seconds. The range of this difference in unpleasantness measure is from -30 to 30. The higher the bar is, the more subjects acknowledge that clicking from 450 seconds to 500 seconds is more unpleasant than clicking from 100 seconds to 150 seconds.

I then show the regression results on empathetic beliefs in Table 5 for the following econometric model.

$$\frac{\hat{\text{MC}}(\text{Poor})}{\hat{\text{MC}}(\text{wealthy})} = a_1\text{MR} + a_2\text{Uniform Experience} + a_3\text{MR} \times \text{Uniform Experience}$$

$$+ a_4\text{Altruism control} + a_5\text{Demographic Control}$$

The dependant variable is the within-subject ratio of the subject's guess of unpleasantness of clicking the last 50 seconds for someone assigned 500 seconds to that for someone assigned 150 seconds. The larger it is, the more the subject acknowledges that clicking from 450 seconds to 500 seconds is more unpleasant than clicking from 100 seconds to 150 seconds. Due to noisiness of the beliefs data, the results lack power in general. Still, the key interaction effect between motivated reasoning and experience variation is still statistically significant with the hypothesized negative sign, see Column (3).

The negative interaction effect in Column (3) means that compared to wealthy individuals with more varied experience, those with uniform experience are more subject to self-serving motivated reasoning in their beliefs. In other words, once subjects know about the redistributional purpose of Stage 3 and that they are in the privileged position to redistribute, those with uniform experience are more likely to believe that the marginal

cost of clicking at 500 seconds is not much higher than that at 100 seconds than those with varied experience. That is, once I turn on the motivated reasoning channel, subjects with uniform experience are less empathetic towards their poor partner than subjects with varied experience.

In Column (4), I control for the individual level altruism measure as well as demographics. Since I add in three additional variables, the interaction term is now only marginally significant with almost the same magnitude as that in Column (3). Altruism level and age don't seem to be related to empathetic beliefs. Women on average are more empathetic than men, acknowledging that the poor needs more help than themselves even when there is a selfish motivations. Still, the gender effect is only marginally significant, and a separate investigation into gender differences in empathy, or motivated reasoning in general, will be left to future research.

This result that empathy is increased by more varied experience is consistent with the redistribution results in Table 2. Recall that in Table 2, I show that redistribution is increased by more varied experience when there is motivated reasoning. Putting the two pieces of results together, I find through the laboratory experiment that self-serving motivated reasoning prevents the wealthy individual from being empathetic towards their poor partner. More varied experience constrains this kind of motivated reasoning, and thus increases empathy, which increases redistribution. The laboratory experiment corroborates the main propositions of the theory.

Table 5: Empathy in Beliefs: The Interaction Effect between Motivated Reasoning and Experience Variation

	Guessed Subjective Marginal Costs Ratios MC at 450-500sec divided by MC at 100-150sec			
	(1)	(2)	(3)	(4)
MR	0.20		0.57	0.59
	(0.25)		(0.44)	(0.45)
Uniform Experience		-0.27	0.12	0.11
		(0.24)	(0.21)	(0.22)
MR×Uniform Experience			-0.79*	-0.73
			(0.47)	(0.47)
Altruism				-0.02
				(0.35)
Young				-0.27
				(0.26)
Male				-0.40
				(0.30)
Observations	331	331	331	319

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependant variable is the within-subject ratio of the subject's guess of unpleasantness of clicking the last 50 seconds for someone assigned 500 seconds to that for someone assigned 150 seconds. The larger it is, the more the subject acknowledges that clicking from 450 seconds to 500 seconds is more unpleasant than clicking from 100 seconds to 150 seconds. MR is a dummy variable with a value of either 0 or 1. If MR=0, subjects are in the no motivated reasoning treatment; if MR=1, subjects are in the motivated reasoning treatment. Uniform Experience is also a dummy variable taking the value of either 0 or 1. If Uniform experience=1, subjects are "wealthy-wealthy" in the previous two experience stages, and have only done the easy 100 seconds clicking task; if Uniform experience=0, subjects are "wealthy-poor" in the two experience stages and did both the 100 seconds clicking and the 500 seconds one. MR×Uniform Experience is the interaction term between the motivated reasoning treatment and the experience variation treatment. Altruism ranges from 0 to 1, the higher it is, the more altruistic the subject is. Young is a dummy variable taking the value of 0 or 1. If Young=1, the subject's age is below the median age of the sample (younger than 40); if Young=0, subjects are older than 40. Male is a dummy variable taking the value of 0 or 1. If Male=1, the subject is men, otherwise, the subject is woman.

6 Concluding Remarks

In this paper, I study the under-researched economic area, empathy, and investigate its implication for redistribution. Due to motivated reasoning, the wealthy lacks empathy towards the poor. More varied personal experience of the wealthy constrains this self-serving bias in beliefs, promotes empathy towards the poor, and therefore, increases redistribution. I build a model of empathy that considers these mechanisms, and test the model predictions with a laboratory experiment. I find that for subjects who have always been assigned the easy task, those who know they are going to redistribute to others believe in a smaller difference between the difficulties of the hard versus easy tasks and redistribute less than those who do *not* know about redistribution–the former engage in motivated reasoning and has lower empathy, and therefore are less generous in redistri-

bution. This effect is reversed for subjects who have done both the hard and the easy tasks–for them, knowing about the redistribution does not make them think more selfishly about the relative difficulty of the easy and hard tasks to justify less redistribution, nor does it make them more selfish in redistribution.

This paper opens path for future research on the formation and implications of empathy. First, it would be enlightening if one can quantify the effects of altruism versus empathy in other-regarding decisions. This paper separates empathy from altruism conceptually and theoretically, and the experimental results confirm that the existence of altruism does not subsume empathy. Yet the magnitudes of their relative importance are unknown. It would be especially interesting if one can endogenize *both* empathy *and* altruism simultaneously and investigate their mechanisms. Second, this paper shows that women are on average more helpful and more empathetic than men. Yet, the statistical power of the gender results in this paper are not strong enough to overcome the lack of consensus in previous research. It would be very helpful if one can zoom in the mechanisms behind gender differences in empathy, and quantify their contributions to the gender differences in pro-social behaviors. Finally, this paper focuses on the impact of direct personal experience treated with a laboratory experiment. However, it remains silent about how indirect experience via social networks shapes empathy. One may want to understand to what degree indirect experience acquired from family and friends impact empathy and prosocialness, and discuss its potential implication for economic inequality and cultural polarization.

References

Ager, P., L. Bursztyn, and H.-J. Voth (2017). Killer incentives: Status competition and pilot performance during world war ii. Technical report, National Bureau of Economic Research.

Alesina, A. and P. Giuliano (2011). Preferences for redistribution. In *Handbook of social economics*, Volume 1, pp. 93–131. Elsevier.

Ambuehl, S. and B. D. Bernheim (2021). Interpreting the will of the people: a positive analysis of ordinal preference aggregation. Technical report, National Bureau of Economic Research.

Ambuehl, S., B. D. Bernheim, and A. Ockenfels (2021). What motivates paternalism? an experimental study. *American economic review 111*(3), 787–830.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of political Economy 97*(6), 1447–1458.

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, 891–904.

Andreoni, J., D. Aydin, B. Barton, B. D. Bernheim, and J. Naecker (2020). When fair isn't fair: Understanding choice reversals involving social preferences. *Journal of Political Economy 128*(5), 1673–1711.

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological inquiry 3*(4), 303–311.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of economic literature 49*(1), 3–71.

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena.

Becker, G. S. (1974). A theory of social interactions. *Journal of political economy 82*(6), 1063–1093.

Benabou, R. and J. Tirole (2006). Belief in a just world and redistributive politics. *The Quarterly journal of economics 121*(2), 699–746.

Bernheim, B. D., L. Braghieri, A. Martínez-Marquina, and D. Zuckerman (2021). A theory of chosen preferences. *American Economic Review 111*(2), 720–54.

Brunnermeier, M. K. and J. A. Parker (2005). Optimal expectations. *American Economic Review 95*(4), 1092–1118.

Chancel, L. and T. Piketty (2021). Global income inequality, 1820–2020: the persistence and mutation of extreme inequality. *Journal of the European Economic Association 19*(6), 3025–3062.

Cherkaoui, M. (1992). Mobilité. In *Traité de sociologie*, pp. 153–193.

Cohn, A., L. J. Jessen, M. Klasnja, and P. Smeets (2019). Why do the rich oppose redistribution? an experiment with america's top 5%. *An experiment with America's top 5*.

Cohn, A., L. J. Jessen, M. Klasnja, and P. Smeets (2021). Why do the rich oppose redistribution? an experiment with america's top 5%. *An experiment with America's top 5*.

Corneo, G. and H. P. Grüner (2002). Individual preferences for political redistribution. *Journal of public Economics 83*(1), 83–107.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature 47*(2), 448–74.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Deaton, A. (2021). Covid-19 and global income inequality. Technical report, National Bureau of Economic Research.

Di Tella, R., R. Perez-Truglia, A. Babino, and M. Sigman (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review 105*(11), 3416–42.

Eil, D. and J. M. Rao (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–38.

Elliott, R., A. C. Bohart, J. C. Watson, and L. S. Greenberg (2011). Empathy. *Psychotherapy 48*(1), 43.

Erdelyi, M. H. (1974). A new look at the new look: perceptual defense and vigilance. *Psychological review 81*(1), 1.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies 83*(2), 587–628.

Exley, C. L. (2020). Using charity performance metrics as an excuse not to give. *Management Science 66*(2), 553–563.

Exley, C. L. and J. B. Kessler (2019). Motivated errors. Technical report, National Bureau of Economic Research.

Festinger, L. (1962). *A theory of cognitive dissonance*, Volume 2. Stanford university press.

Fisman, R., P. Jakiela, and S. Kariv (2017). Distributional preferences and political behavior. *Journal of Public Economics 155*, 1–10.

Fisman, R., P. Jakiela, S. Kariv, and D. Markovits (2015). The distributional preferences of an elite. *Science 349*(6254).

Fisman, R., I. Kuziemko, and S. Vannutelli (2021). Distributional preferences in larger groups: Keeping up with the joneses and keeping track of the tails. *Journal of the European Economic Association 19*(2), 1407–1438.

Giuliano, P. and A. Spilimbergo (2014). Growing up in a recession. *Review of Economic Studies 81*(2), 787–817.

Gul, F. and W. Pesendorfer (2001). Temptation and self-control. *Econometrica 69*(6), 1403–1435.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American economic review 90*(4), 1072–1091.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin 108*(3), 480.

Lee, W. and J. E. Roemer (2006). Racism and redistribution in the united states: A solution to the problem of american exceptionalism. *Journal of public Economics 90*(6-7), 1027–1052.

Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health psychology 24*(4S), S49.

Lönnqvist, J.-E. and G. Walkowitz (2019). Experimentally induced empathy has no impact on generosity in a monetarily incentivized dictator game. *Frontiers in Psychology 10*, 337.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review 111*(6), 1807–44.

Malmendier, U. and S. Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *The quarterly journal of economics 126*(1), 373–416.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics 131*(1), 53–87.

Margalit, Y. (2013). Explaining social policy preferences: Evidence from the great recession. *American Political Science Review 107*(1), 80–103.

Mazar, N., B. Koszegi, and D. Ariely (2014). True context-dependent preferences? the causes of market-dependent valuations. *Journal of Behavioral Decision Making 27*(3), 200–208.

Parker, K. (2012). Yes, the rich are different. *Pew Research Center Report*.

Piketty, T. (1995). Social mobility and redistributive politics. *The Quarterly journal of economics 110*(3), 551–584.

Piketty, T. (2003). Income inequality in france, 1901–1998. *Journal of political economy 111*(5), 1004–1042.

Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913–1998. *The Quarterly journal of economics 118*(1), 1–41.

Piketty, T. and E. Saez (2014). Inequality in the long run. *Science 344*(6186), 838–843.

Piketty, T., L. Yang, and G. Zucman (2019). Capital accumulation, private property, and rising inequality in china, 1978–2015. *American Economic Review 109*(7), 2469–96.

Roth, C. and J. Wohlfart (2018). Experienced inequality and preferences for redistribution. *Journal of Public Economics 167*, 251–262.

Stark, O. and I. Falk (2000). Transfers, empathy formation, and reverse transfers. In *The economics of reciprocity, giving and altruism*, pp. 174–181. Springer.

Stiglitz, J. (2020). Point of view: Conquering the great divide. *Finance & Development 57*(003).

Stocks, E. and D. Lishner (2012). Empathy. In V. Ramachandran (Ed.), *Encyclopedia of Human Behavior (Second Edition)* (Second Edition ed.)., pp. 32–37. San Diego: Academic Press.

Thaler, M. (2020). The'fake news' effect: Experimentally identifying motivated reasoning using trust in news. *Available at SSRN 3717381*.

Zaki, J. (2014). Empathy: a motivated account. *Psychological bulletin 140*(6), 1608.

A Theories and proofs

A.1 Propositions

Proposition A.1. *Assume the cognitive dissonance term D increases whenever* $|\hat{\gamma}-\gamma|$ *increases, when only temptation is turned off, the optimal chosen belief of the consumption utility curvature (empathy) is downward biased with* $\hat{\gamma}^* < \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) < t^*(\gamma)$*.*

Proposition A.2. *Assume the cognitive dissonance term D increases whenever* $|\hat{\gamma}-\gamma|$ *increases, when only guilt is turned off, the optimal chosen belief of the consumption utility curvature (empathy) is upward biased with* $\hat{\gamma}^* > \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) > t^*(\gamma)$*.*

Proposition A.3. *Assume the cognitive dissonance term D increases whenever* $|\hat{\gamma}-\gamma|$ *increases, when the wealthy individual has the same sensitivities to their visceral feelings before and during redistribution, i.e.* $\lambda_1 = \lambda_2$ *and* $\nu_1 = \nu_2$*, the optimal chosen belief of the consumption utility curvature (empathy) is unbiased with* $\hat{\gamma}^* = \gamma$*. Accordingly,* $t^*(\hat{\gamma}^*) = t^*(\gamma)$*.*

Proposition A.4. *When* $\nu_1 = \nu_2 = 0$*,* $\hat{\gamma}^*$ *is increasing in* $var(c)$*,* $t(\hat{\gamma}^*)$ *is increasing in* $var(c)$*.*

Proposition A.5. *When* $\lambda_1 = \lambda_2 = 0$*,* $\hat{\gamma}^*$ *is decreasing in* $var(c)$*,* $t(\hat{\gamma}^*)$ *is decreasing in* $var(c)$*.*

A.2 Basis for reducing parameters

Take logarithm of the data generating process for the consumption-utility experience, Equation (**??**), I get:

$$\log u_i = [\beta - \log(1-\gamma)] + (1-\gamma)\log c_i + \varepsilon_i \tag{A.1}$$

where $i \in \{1,2,...,n\}$ enumerates his experiences.

Now, I can reduce the two subjective slope and curvature parameters, $\hat{\beta}$ and $\hat{\gamma}$, into just one subjective curvature parameter, $\hat{\gamma}$. The reason is that given any consumption-utility data set, $\{(u,c)\}$, and any subjective curvature parameter, $\hat{\gamma}$, the subjective slope parameter $\hat{\beta}$ can auto-adjust to achieve a best fit for the data. That is, I can write $\hat{\beta}$ as $\hat{\beta}(\hat{\gamma})$. In this way, the cognitive dissonance term, D, that measures deviation of the chosen belief from experience is a function of the chosen belief about curvature, $\hat{\gamma}$, alone.

A.3 Recap of the 2-Period Multi-Self Model

In Period 1:

$$\max_{\hat{\theta}} W(t(\hat{\theta}),\theta) - \lambda_1 G(t(\hat{\theta}),\theta) - \nu_1 T(t(\hat{\theta}),\theta) - \phi D(\hat{\theta}|\{(u,c)\})$$

where $\theta = (\gamma,\beta)$;
$W(t,\hat{\theta}) = u(X-t,\hat{\theta}) + au(t,\hat{\theta})$, $a \in (0,1)$;
$G(t,\theta) = \max_{\tilde{t}} V(\tilde{t},\theta) - V(t,\theta)$; $V(t,\theta) = u(X-t,\theta) + \hat{a}u(t,\theta)$, $\hat{a} > a$;
and
$T(t,\theta) = \max_{\tilde{t}} u((X-\tilde{t}),\theta) - u((X-t),\theta)$.

In Period 2:

$$\max_{t} W(t,\hat{\theta}) - \lambda_2 G(t,\hat{\theta}) - \nu_2 T(t,\hat{\theta})$$

where $\lambda_2 > \lambda_1 \geq 0$, $\nu_2 > \nu_1 \geq 0$

A.4 Optimal transfer given empathy

After dropping terms that is independent from t in the Period 2 objective function, I get:

$$\max_{t} (1+\lambda_2+\nu_2)u(X-t,\hat{\theta}) + (a+\lambda_2\hat{a})u(t,\hat{\theta})$$

The F.O.C. gives:

$$-(1+\lambda_2+\nu_2)(X-t)^{-\hat{\gamma}} \cdot (1-\hat{\gamma}) \cdot \exp\hat{\beta}(\hat{\gamma}) + (a+\lambda_2\hat{a})t^{-\hat{\gamma}} \cdot (1-\hat{\gamma}) \cdot \exp\hat{\beta}(\hat{\gamma}) = 0$$

Drop off the $(1-\hat{\gamma}) \cdot \exp\hat{\beta}(\hat{\gamma})$ terms, and let

$$A \equiv \frac{a+\lambda_2\hat{a}}{1+\lambda_2+\nu_2} \in (0,1) \tag{A.2}$$

[16] I have:

$$\frac{t}{X-t} = A^{1/\hat{\gamma}}$$

[16] Since both $a, \hat{a} \in [0,1]$ and $\lambda_2, \nu_2 \in [0,+\infty)$, technically, $A \in [0,1]$. For analysis purposes, here I do not consider the case where $A = 0$, i.e. the wealthy individual does not care about the poor at all, or $A = 1$, i.e. the wealthy individual cares about the poor as much as they care about themselves and they don't feel any temptations to be selfish in redistribution. I call A an effective altruism parameter that summarizes intrinsic altruism and sensitivity to visceral emotions such as guilt and temptations when the wealthy individual is making a redistribution. Also note that when $\nu_2 \to 0$, guilt dominates and the effective altruism $A > a$. When $\lambda_2 \to 0$, temptation takes over and the effective altruism $A < a$.

i.e.

$$t^* = \frac{A^{1/\hat{\gamma}}}{A^{1/\hat{\gamma}}+1} \cdot X \tag{A.3}$$

$$(X-t)^* = \frac{1}{A^{1/\hat{\gamma}}+1} \cdot X$$

[17]

A.5 Proof of Proposition A.1

Proof. Substitute the solved redistribution function $t^*(\hat{\gamma})$ into the first period objective function:

$$\max_{\hat{\gamma}} \underbrace{(1+\lambda_1+\nu_1)e^{\beta}(\frac{1}{A^{1/\hat{\gamma}}+1})^{1-\gamma}X^{1-\gamma} + (a+\lambda_1\hat{a})e^{\beta}(\frac{A^{1/\hat{\gamma}}}{A^{1/\hat{\gamma}}+1})^{1-\gamma}X^{1-\gamma}}_{\text{"U"}} - \phi D(\hat{\gamma}, \hat{\beta}(\hat{\gamma})|\{(u,c)\})$$

[18] Rearrange

$$U = e^{\beta}X^{1-\gamma}(A^{1/\hat{\gamma}}+1)^{\gamma-1} \cdot [1+\lambda_1+\nu_1+(a+\lambda_1\hat{a}) \cdot A^{(1-\gamma)/\hat{\gamma}}] > 0$$

Hence,

$$\text{sign}(\frac{\partial U}{\partial \hat{\gamma}}) = \text{sign}(\frac{\partial \log U}{\partial \hat{\gamma}})$$

where

$$\frac{\partial \log U}{\partial \hat{\gamma}} = \frac{\partial \beta + (1-\gamma)\log X + (\gamma-1)\log(A^{1/\hat{\gamma}}+1) + \log[1+\lambda_1+\nu_1+(a+\lambda_1\hat{a}) \cdot A^{(1-\gamma)/\hat{\gamma}}]}{\partial \hat{\gamma}}$$

Dropping the terms independent from the chosen curvature parameter, $\hat{\gamma}$:

$$\frac{\partial \log U}{\partial \hat{\gamma}} = \underbrace{(\gamma-1)}_{<0}\underbrace{\frac{\partial \log(A^{1/\hat{\gamma}}+1)}{\partial \hat{\gamma}}}_{>0} + \underbrace{\frac{\partial \log[1+\lambda_1+\nu_1+(a+\lambda_1\hat{a}) \cdot A^{(1-\gamma)/\hat{\gamma}}]}{\partial \hat{\gamma}}}_{>0}$$

[19] Writing out and combining the two parts on the RHS with opposing signs below. Also let

$$A_0 = \frac{a+\lambda_1\hat{a}}{1+\lambda_1+\nu_1} \in (0,1) \tag{A.4}$$

[17] Intuitively, transfers are increasing in the compound altruism parameter, A. Also, note that optimal redistribution t^* is only a function of the subjective curvature parameter, $\hat{\gamma}$, i.e. $t^*(\hat{\gamma})$.

[18] I call the bracketed function a compound utility function, U, that summarizes not only utilities from one's own and other's material consumption, but also the visceral feelings such as guilt and temptation in deciding such consumption redistribution. Also, notice that only the true level parameter, β, enters U, not the subjective $\hat{\beta}(\hat{\gamma})$, which provides tractability in the analysis to follow.

[19] The two log-terms are non-decreasing in $\hat{\gamma}$ since $A \in [0,1]$ and the curvature parameter $\gamma \in (0,1)$

[20] I have:

$$\begin{aligned}\frac{\partial \log U}{\partial \hat{\gamma}} &= (\gamma-1)\frac{A^{1/\hat{\gamma}}\ln A\cdot\frac{-1}{\hat{\gamma}^2}}{A^{1/\hat{\gamma}}+1}+\frac{(a+\lambda_1\hat{a})\cdot A^{(1-\gamma)/\hat{\gamma}}\ln A\cdot\frac{-(1-\gamma)}{\hat{\gamma}^2}}{1+\lambda_1+\nu_1+(a+\lambda_1\hat{a})\cdot A^{(1-\gamma)/\hat{\gamma}}} \\ &= -\frac{A^{1/\hat{\gamma}}\ln A\cdot\frac{(\gamma-1)}{\hat{\gamma}^2}}{A^{1/\hat{\gamma}}+1}+\frac{(a+\lambda_1\hat{a})A^{1/\hat{\gamma}}\cdot\ln A\cdot\frac{(\gamma-1)}{\hat{\gamma}^2}}{(1+\lambda_1+\nu_1)\cdot A^{\gamma/\hat{\gamma}}+(a+\lambda_1\hat{a})A^{1/\hat{\gamma}}} \\ &= \underbrace{A^{1/\hat{\gamma}}\ln A\cdot\frac{\gamma-1}{\hat{\gamma}^2}}_{>0}\cdot\Big(\underbrace{\frac{-1}{A^{1/\hat{\gamma}}+1}}_{<0}+\underbrace{\frac{1}{A^{1/\hat{\gamma}}+\frac{1}{A_0}\cdot A^{\gamma/\hat{\gamma}}}}_{>0}\Big)\end{aligned} \tag{A.5}$$

Since $\lambda_2 \geq \lambda_1 \geq 0, \nu_2 \geq \nu_1 \geq 0$, $\hat{a} > a > 0$, when there is guilt but no temptation, i.e. $\nu_1 = \nu_2 = 0$ and $\lambda_2 > \lambda_1$, from Equations (A.2) and (A.4), I have $A > A_0$.

I now prove the proposition by contradiction. Suppose $\hat{\gamma}^* \geq \gamma$, then:

$\Rightarrow 0 < \gamma/\hat{\gamma}^* \leq 1$, given $0 < A_0 \leq A < 1$,

$\Rightarrow A^{\gamma/\hat{\gamma}^*} \geq A > A_0$

$\Rightarrow \frac{1}{A_0}\cdot A^{\gamma/\hat{\gamma}^*} > 1$

$\Rightarrow \frac{1}{A^{1/\hat{\gamma}^*}+\frac{1}{A_0}\cdot A^{\gamma/\hat{\gamma}^*}} < \frac{1}{A^{\frac{1}{\hat{\gamma}^*}}+1} \Rightarrow \frac{\partial \log U}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} < 0 \Rightarrow \frac{\partial U}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} < 0$ when $\hat{\gamma}^* \geq \gamma$

In addition, since the subjective curvature parameter would fit the data best when it equals to the true parameter in the data generation process, I have:

$\frac{\partial -D}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} \leq 0, \quad$ *when* $\hat{\gamma}^* \geq \gamma$.

Putting the U and D terms together, the objective function (with endogenous action) in period 1 is strictly decreasing when $\hat{\gamma}^* \geq \gamma$, which contradicts the optimality of $\hat{\gamma}^*$. Therefore, $\hat{\gamma}^* < \gamma$.

The redistribution result follows from Equation (A.3). Since t^* is increasing in $\hat{\gamma}$, $\hat{\gamma}^* < \gamma$, I have $t^*(\hat{\gamma}^*) < t^*(\gamma)$. □

A.6 Proof for Proposition A.2

Proof. When there is temptation but no guilt, i.e. $\lambda_1 = \lambda_2 = 0$ and $\nu_2 > \nu_1$, from Equations (A.2) and (A.4), I have $A < A_0$.

[20] Same as for A, here I do not consider the case where $A_0 = 0$ or $A_0 = 1$. I call A_0 an effective ex-ante altruism parameter that takes place when the wealthy individual is considering a future redistribution.

Again, I prove through contradiction. Suppose $\hat{\gamma}^* \leq \gamma$, following Equation (A.5), I then have:

$$\Rightarrow \gamma/\hat{\gamma}^* \geq 1, \text{ given } 0 < A < A_0 < 1,$$

$$\Rightarrow A^{\gamma/\hat{\gamma}^*} \leq A < A_0$$

$$\Rightarrow 0 < \tfrac{1}{A_0} \cdot A^{\gamma/\hat{\gamma}^*} < 1$$

$$\Rightarrow \tfrac{1}{A^{1/\hat{\gamma}^*} + \frac{1}{A_0} \cdot A^{\gamma/\hat{\gamma}^*}} > \tfrac{1}{A^{\frac{1}{\hat{\gamma}^*}} + 1} \Rightarrow \tfrac{\partial \log U}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} > 0 \Rightarrow \tfrac{\partial U}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} > 0 \text{ when } \hat{\gamma}^* \leq \gamma$$

In addition, since the subjective curvature parameter would fit the data best when it equals to the true parameter in the data generation process, I have:

$$\tfrac{\partial -D}{\partial \hat{\gamma}}|_{\hat{\gamma}^*} \geq 0, \quad \textit{when } \hat{\gamma}^* \leq \gamma.$$

Putting the U and D terms together, the objective function (with endogenous action) in period 1 is strictly increasing when $\hat{\gamma}^* \leq \gamma$, which contradicts the optimality of $\hat{\gamma}^*$. Therefore, $\hat{\gamma}^* > \gamma$.

The redistribution result follows from Equation (A.3). Since t^* is increasing in $\hat{\gamma}$, $\hat{\gamma}^* > \gamma$, I have $t^*(\hat{\gamma}^*) > t^*(\gamma)$.

□

A.7 Proof for Proposition A.3

Proof. When $\lambda_1 = \lambda_2$ and $\nu_1 = \nu_2$, I have $A = A_0$.

Prove by contradiction. First suppose $\hat{\gamma}^* > \gamma$, then given $A \in (0,1)$, $A^{(\gamma - \hat{\gamma}^*)/\hat{\gamma}^*} > 1$, so from Equation (A.5), I have $\tfrac{\partial U}{\partial \gamma}|_{\hat{\gamma}^*} < 0$. In addition, the goodness of $\hat{\gamma}$ fitting the data set $\{(u,c)\}$ is non-increasing with $\hat{\gamma}$ when $\hat{\gamma} > \gamma$. Putting the two together, the objective function is strictly decreasing when at $\hat{\gamma}^* > \gamma$, a contraction with its supposed optimality. Therefore, $\hat{\gamma}^* \leq \gamma$.

Now suppose $\hat{\gamma}^* < \gamma$, then $A^{(\gamma - \hat{\gamma}^*)/\hat{\gamma}^*} = \tfrac{A^{\gamma/\hat{\gamma}*}}{A} < 1$, then from Equation (A.5), I have $\tfrac{\partial U}{\partial \gamma}|_{\hat{\gamma}^*} > 0$. In addition, $-\tfrac{\partial D}{\partial \hat{\gamma}} > 0$ when $\hat{\gamma} < \gamma$. Therefore, the objective function is strictly increasing at $\hat{\gamma}^* < \gamma$, contradicting its optimality. Therefore, $\hat{\gamma}^* = \gamma$.[21]

The redistribution result automatically follows from Equation (A.3).

□

[21] It is not needed for the proof, but it is also easy to verify that $\tfrac{\partial U}{\partial \gamma}|_{\hat{\gamma}^*} - \phi \tfrac{\partial D}{\partial \hat{\gamma}} = 0$ when $\hat{\gamma}^* = \gamma$.

A.8 Proof for Proposition A.4 and A.5

Proof. Notice that the consumption-utility experience dataset only enters the objective function through the cognitive dissonance term, and is not related to the intrinsic utility term. It is sufficient that I only look at the D term. Using the OLS formula for intercepts, $\hat{\beta} = \overline{\log u} - (1-\hat{\gamma})\overline{\log c}$. Now I compute the MSE of the estimation and take that as the cognitive dissonance cost D.

$$
\begin{aligned}
& u(c) = e^{\beta} c^{1-\gamma} e^{\varepsilon} \\
\Rightarrow & \log u_i = \beta + (1-\gamma)\log c_i + \varepsilon_i \\
& = [\overline{\log u} - (1-\hat{\gamma})\overline{\log c}] + (1-\hat{\gamma})\log c_i + \hat{\varepsilon}_i \\
\Rightarrow & \hat{\varepsilon}_i = \log u_i - \overline{\log u} + (1-\hat{\gamma})(\overline{\log c} - \log c_i) \\
\Rightarrow & D = \frac{1}{n\sigma^2}\sum_{i=1}^{n}[\log u_i - \overline{\log u} + (1-\hat{\gamma})(\overline{\log c} - \log c_i)]^2 \\
\Rightarrow & -\frac{\partial \phi D}{\partial \hat{\gamma}} = \frac{2}{\sigma^2}(1-\hat{\gamma})\underbrace{\frac{1}{n}\sum_{i=1}^{n}(\log c_i - \overline{\log c})^2}_{var(c)} - \frac{2}{\sigma^2}\underbrace{\frac{1}{n}\sum_{i=1}^{n}[(\log c_i - \overline{\log c})\cdot(\log u_i - \overline{\log u})]}_{corr(\log c, \log u)}
\end{aligned}
$$

Since the correlation between consumption and utility is generated by the true utility function, it is fixed. Therefore, $-\frac{\partial \phi D}{\partial \hat{\gamma}}$ is increasing in $var(c)$.

Also, based on the construction of $-D < 0$, it measures how much fitness to data is lost with the subjective $\hat{\gamma}$ parameter, so $-\frac{\partial \phi D}{\partial \hat{\gamma}} > 0$ when $\hat{\gamma} < \gamma$ and $-\frac{\partial \phi D}{\partial \hat{\gamma}} < 0$ when $\hat{\gamma} > \gamma$.

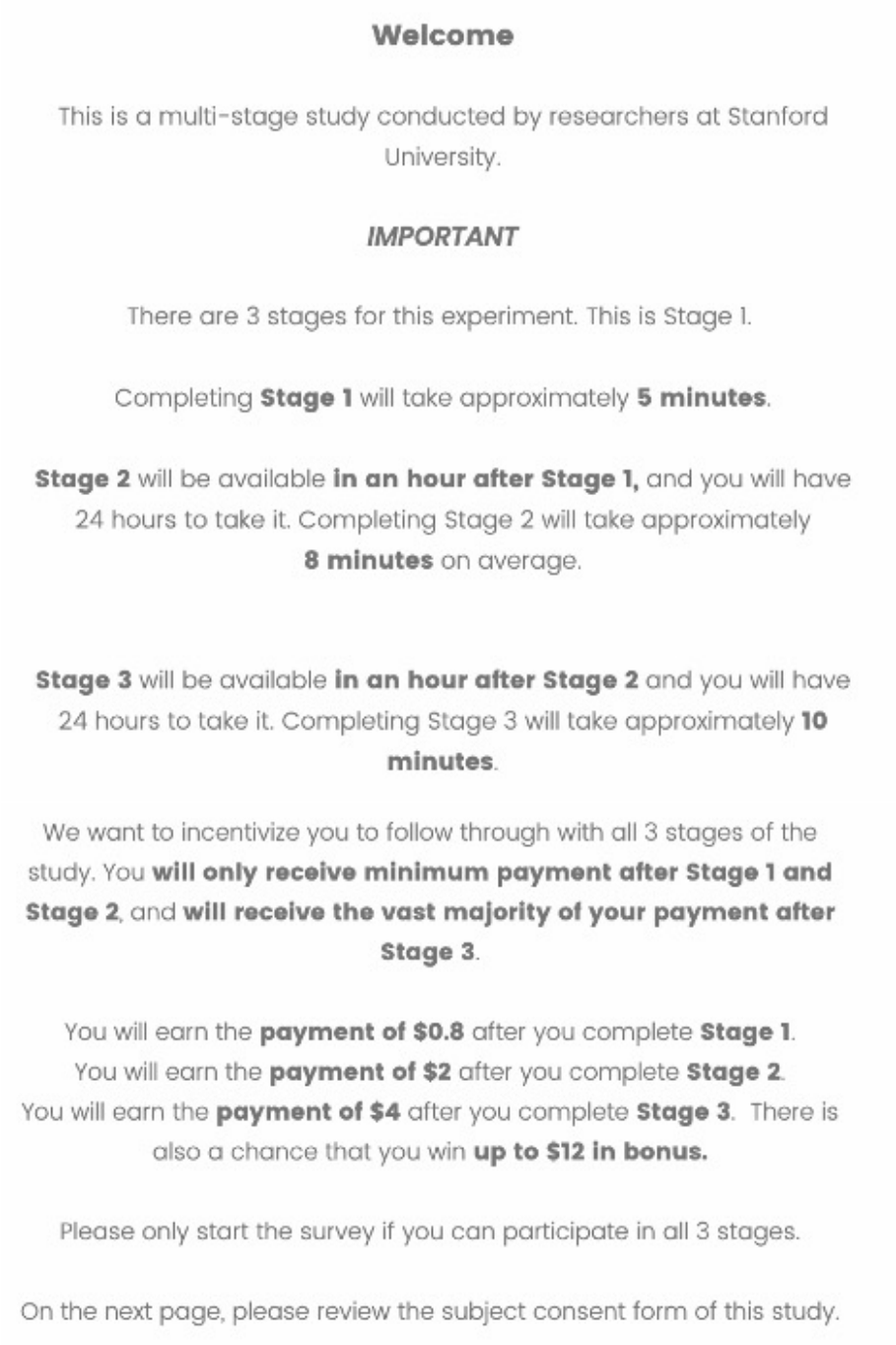
I first prove Proposition A.4. Proposition A.1 shows that when $\nu_1 = \nu_2 = 0$, $0 < \hat{\gamma}^* < \gamma < 1$, so the larger the consumption variation is, the steeper is the slope of cognitive dissonance relative to any chosen belief in curvature $\hat{\gamma}^* < \gamma$. Therefore, optimal $\hat{\gamma}^*$ will be closer to the true curvature γ when the variation of consumption, $var(c)$, is larger. The rest follows from Equation (A.3).

I now prove Proposition A.5. Proposition A.2 shows that when $\lambda_1 = \lambda_2 = 0$, $0 < \gamma < \hat{\gamma}^* < 1$. So the larger $var(c)$ is, the less negative is $-\frac{\partial \phi D}{\partial \hat{\gamma}}$ for any $\hat{\gamma}^* > \gamma$. Therefore, optimal $\hat{\gamma}^*$ will be closer to the true curvature γ when the variation of consumption, $var(c)$, is larger. The rest follows from Equation (A.3). □

B Experiment Instructions

B.1 Stage 1

On the welcome page, subjects were first notified about the multi-stage nature of the experiment. See the following screenshot for the message.



Welcome

This is a multi-stage study conducted by researchers at Stanford University.

IMPORTANT

There are 3 stages for this experiment. This is Stage 1.

Completing **Stage 1** will take approximately **5 minutes**.

Stage 2 will be available **in an hour after Stage 1,** and you will have 24 hours to take it. Completing Stage 2 will take approximately **8 minutes** on average.

Stage 3 will be available **in an hour after Stage 2** and you will have 24 hours to take it. Completing Stage 3 will take approximately **10 minutes**.

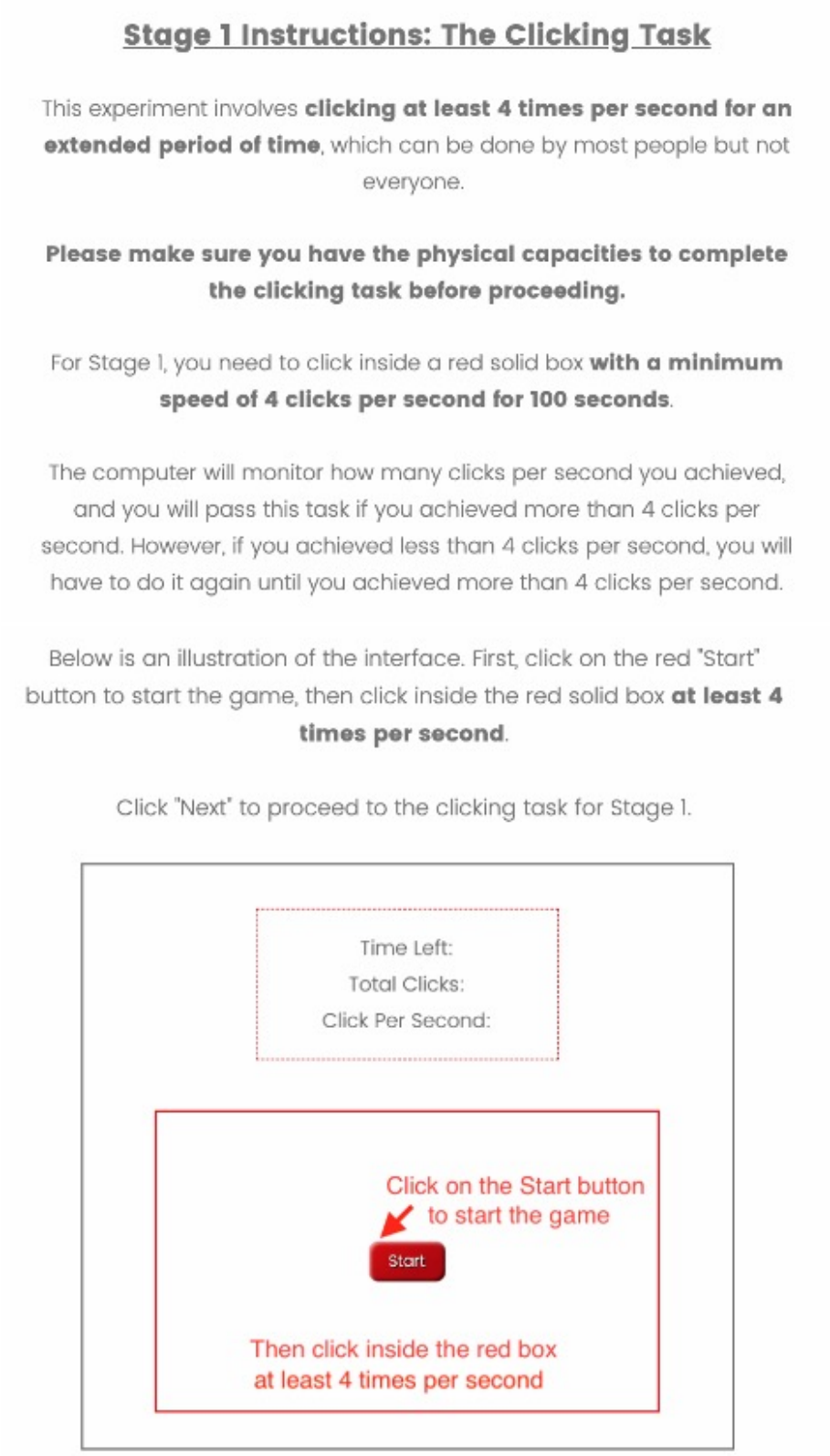
We want to incentivize you to follow through with all 3 stages of the study. You **will only receive minimum payment after Stage 1 and Stage 2**, and **will receive the vast majority of your payment after Stage 3**.

You will earn the **payment of $0.8** after you complete **Stage 1**.
You will earn the **payment of $2** after you complete **Stage 2**.
You will earn the **payment of $4** after you complete **Stage 3**. There is also a chance that you win **up to $12 in bonus.**

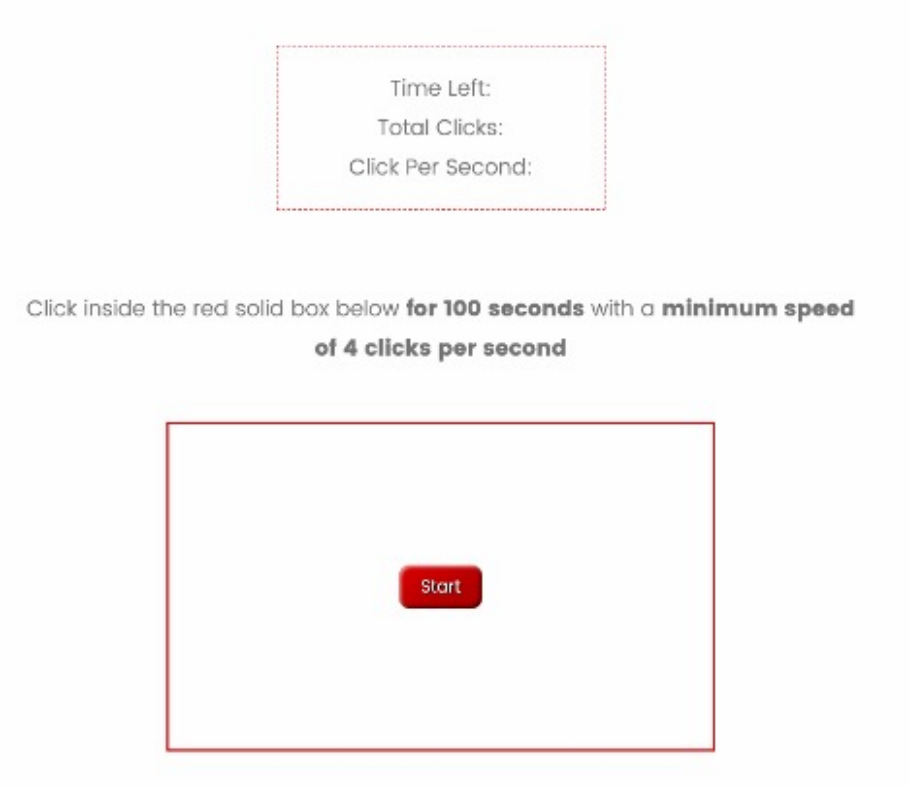
Please only start the survey if you can participate in all 3 stages.

On the next page, please review the subject consent form of this study.

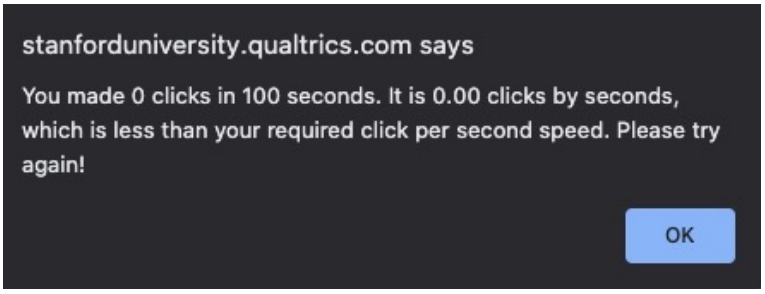
Then I introduce the real effort task to subjects. See the following screenshot for the instruction of the task.



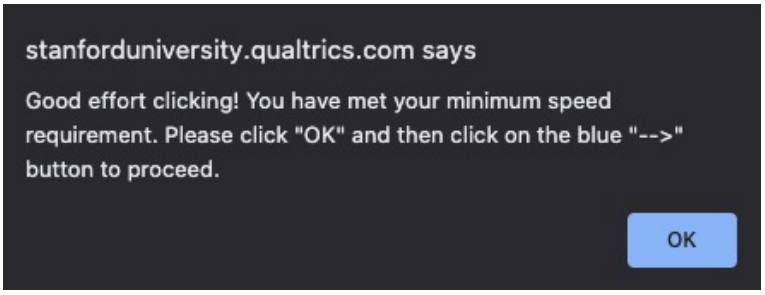
All subjects are assigned the easy task, i.e. 100 seconds of clicking, for Stage 1. See the following screenshot for the interface.



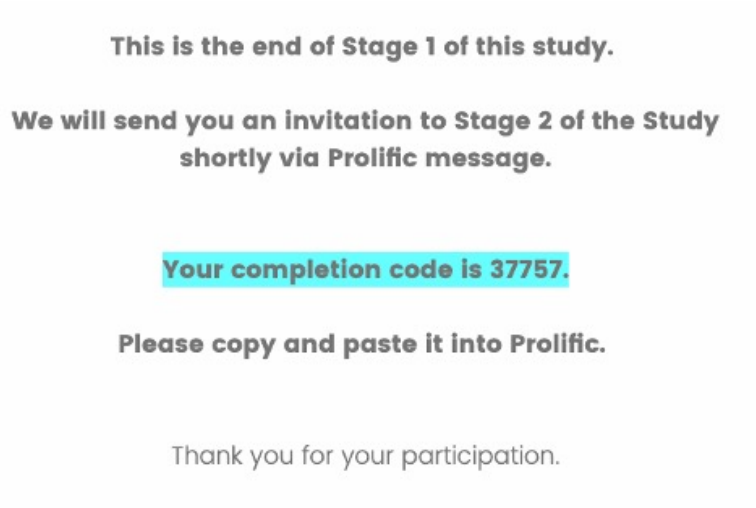
Subjects have to meet the 4 clicks/sec minimum speed requirement, if not, they will be told to try again until they meet the lower speed limit. See the following screenshot for the alert message in that case.



Once they meet the minimum speed requirement, they will see the message below and be allowed to proceed.

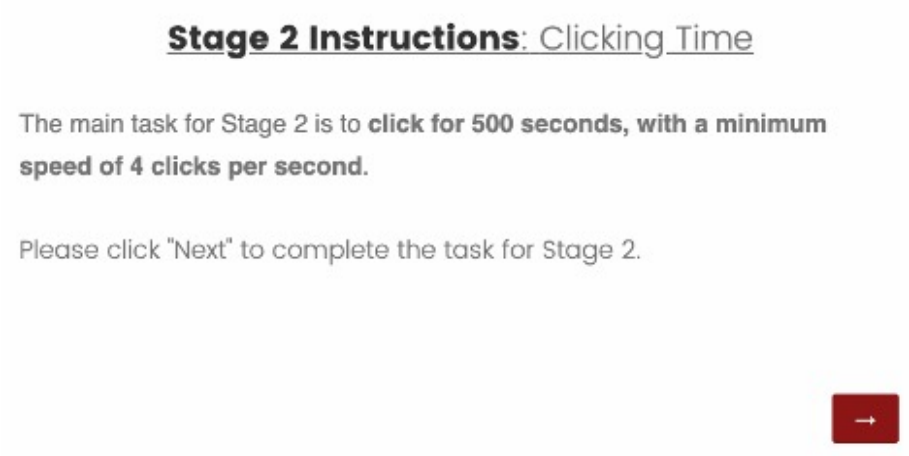


Finally, subjects were told that they were done with Stage 1 and would receive invite to Stage 2 on Prolific.

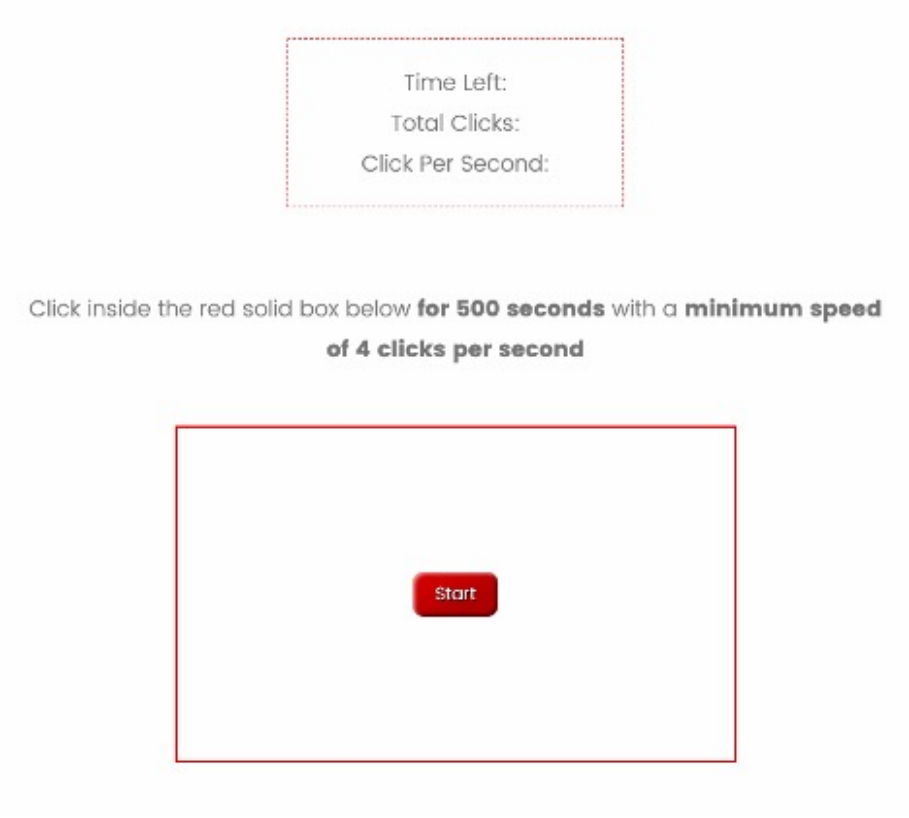


B.2 Stage 2

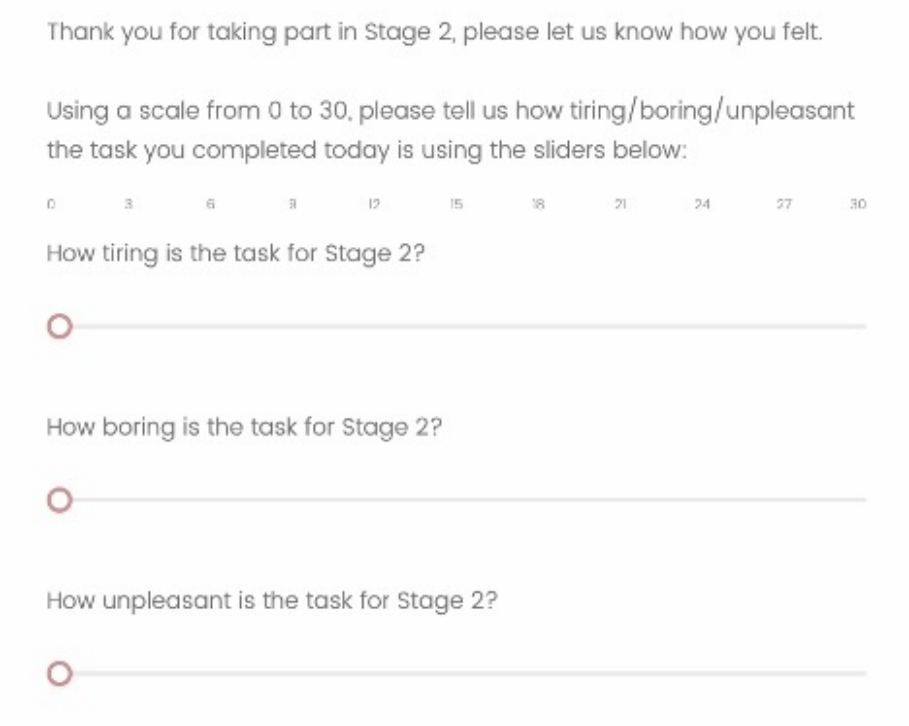
Half of the subjects are assigned the easy task (100 seconds of clicking) again, and the other half are assigned the hard task (500 seconds of clicking). The following screenshot shows the page where they were assigned 500 seconds.



The task interface is the same throughout the three stages.

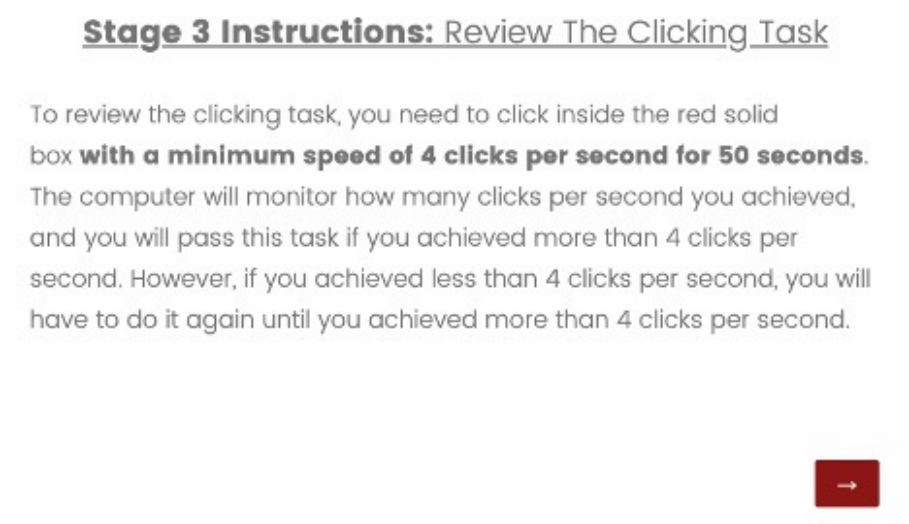


After they were done with either the 500 seconds of clicking or the 100 seconds of clicking, subjects were asked to evaluate their experience in Stage 2. This is to help them reflect on the experience and remember it better. See the following screenshot for the questions.

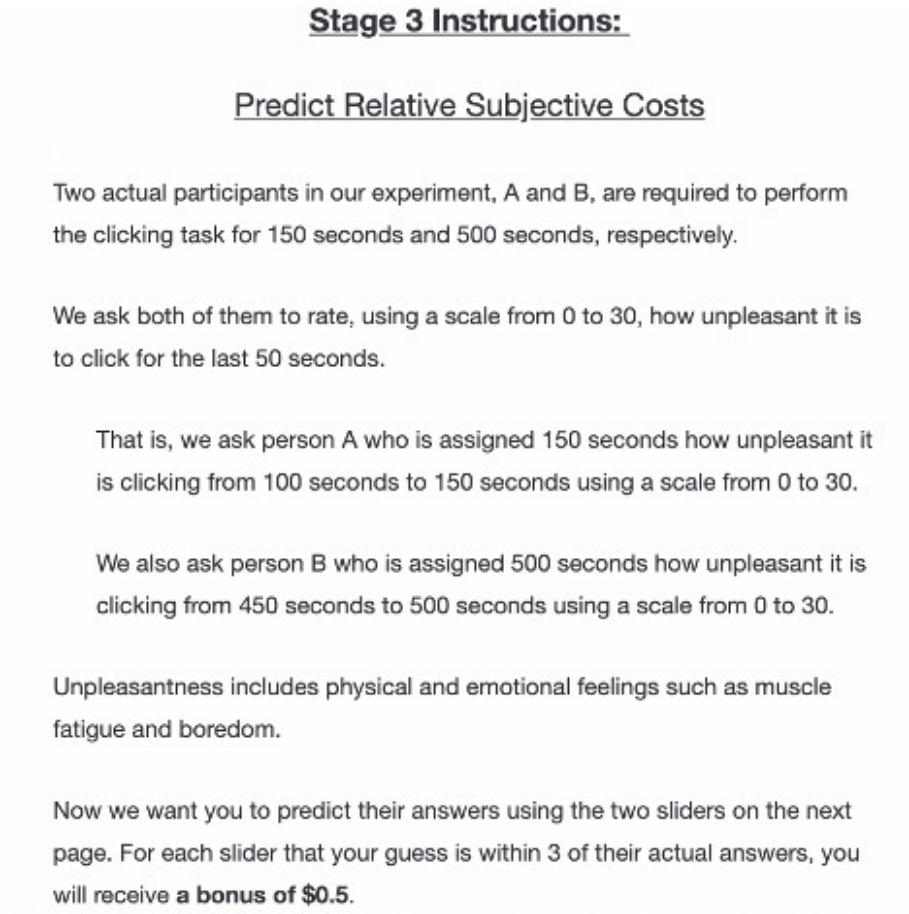


B.3 Stage 3

Stage 3 begins by having everyone review the clicking task for 50 seconds.



In Stage 3, half of the subjects were in the no motivated reasoning treatment, whereas the other half were in the motivated reasoning treatment. The only difference between these two groups is whether they were first shown the belief elicitation block or the position revelation block. For those in the no motivated reasoning treatment, they first saw the belief elicitation block shown in the following screenshots.



Stage 3 Instructions:

Predict Relative Subjective Costs

Two actual participants in our experiment, A and B, are required to perform the clicking task for 150 seconds and 500 seconds, respectively.

We ask both of them to rate, using a scale from 0 to 30, how unpleasant it is to click for the last 50 seconds.

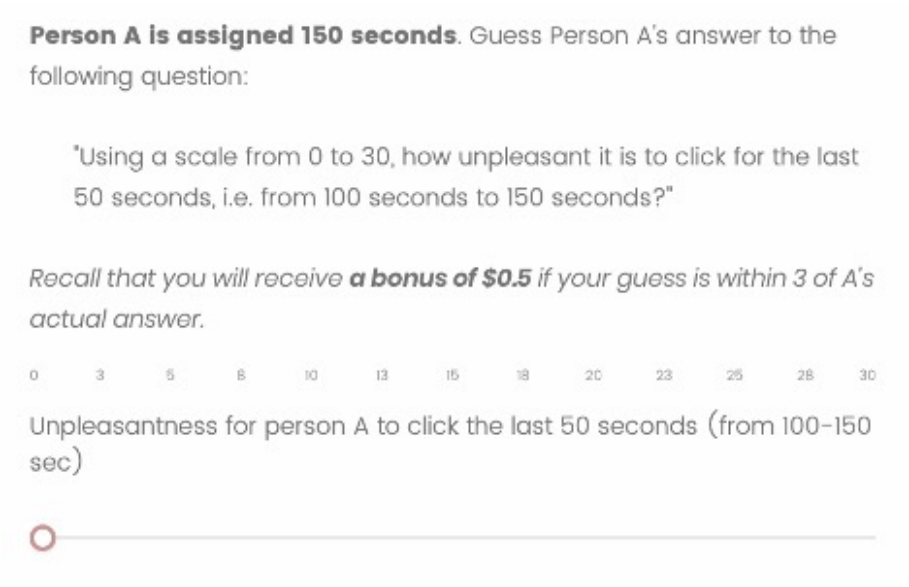
That is, we ask person A who is assigned 150 seconds how unpleasant it is clicking from 100 seconds to 150 seconds using a scale from 0 to 30.

We also ask person B who is assigned 500 seconds how unpleasant it is clicking from 450 seconds to 500 seconds using a scale from 0 to 30.

Unpleasantness includes physical and emotional feelings such as muscle fatigue and boredom.

Now we want you to predict their answers using the two sliders on the next page. For each slider that your guess is within 3 of their actual answers, you will receive **a bonus of $0.5**.

On the next page, they were asked to guess the relative unpleasantness of two other subjects, A and B, doing the hard (500 seconds) task versus the easy task (100 seconds).



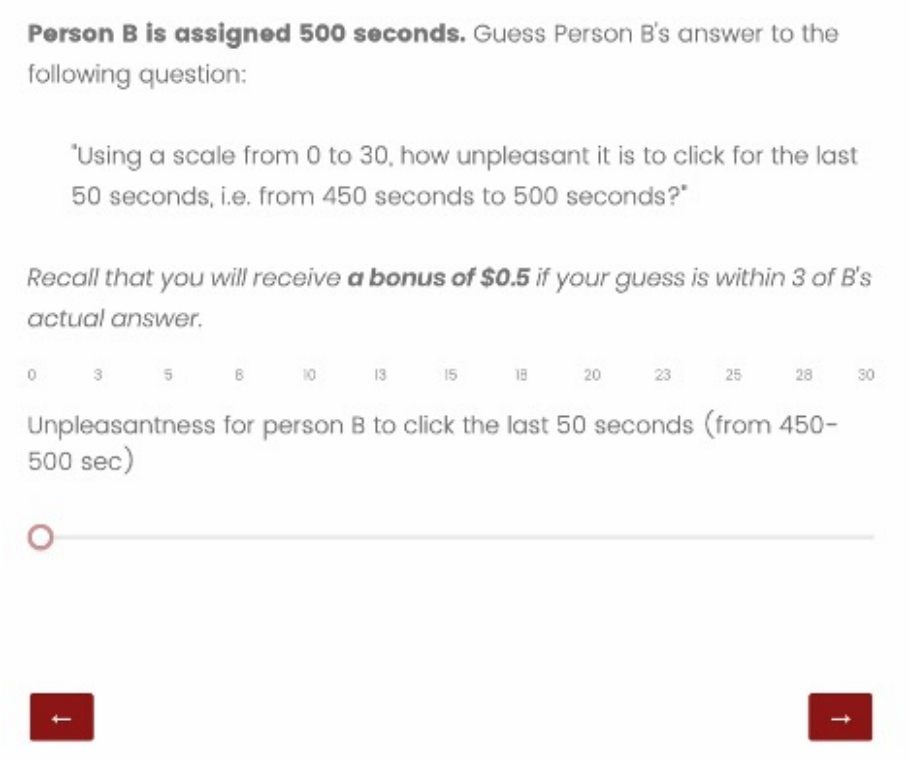
Person A is assigned 150 seconds. Guess Person A's answer to the following question:

"Using a scale from 0 to 30, how unpleasant it is to click for the last 50 seconds, i.e. from 100 seconds to 150 seconds?"

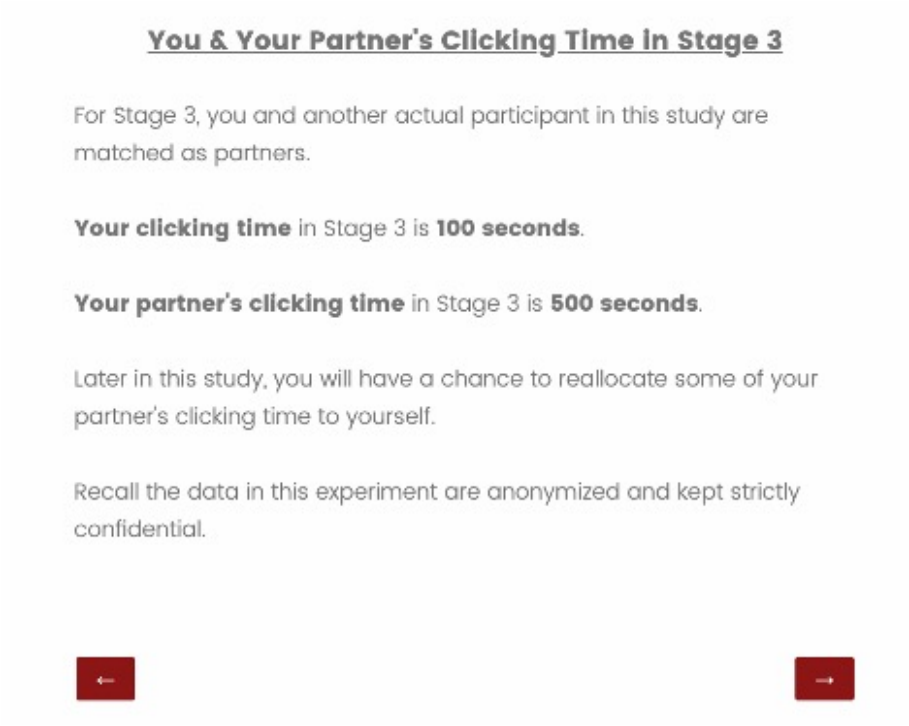
*Recall that you will receive **a bonus of $0.5** if your guess is within 3 of A's actual answer.*

0 3 5 8 10 13 15 18 20 23 25 28 30

Unpleasantness for person A to click the last 50 seconds (from 100-150 sec)



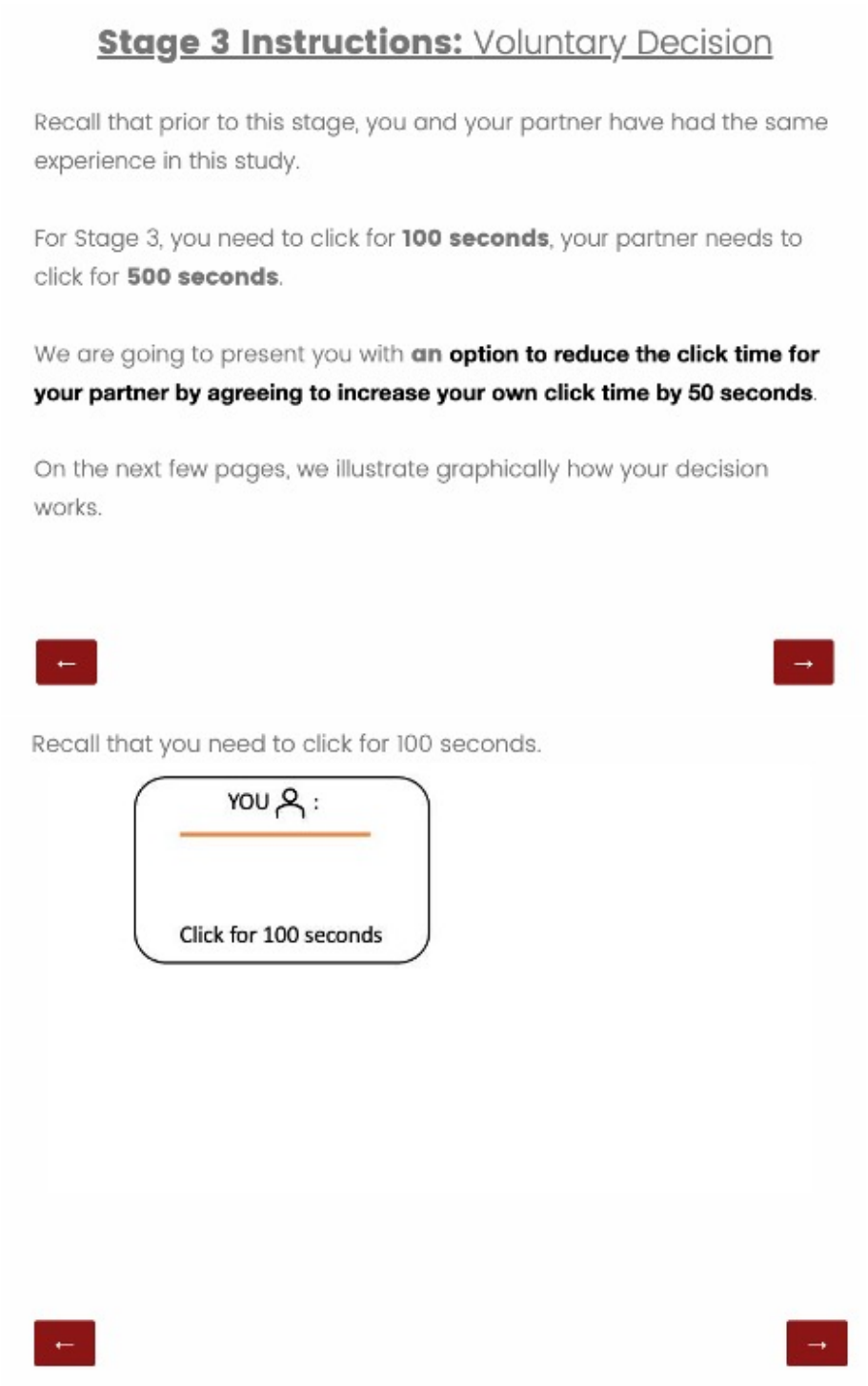
For these subjects in the no motivated reasoning treatment, after they were done with the belief questions shown above, they were told about their positions in Stage 3 and notified about the redistribution task. See the following screenshot for this revelation block.



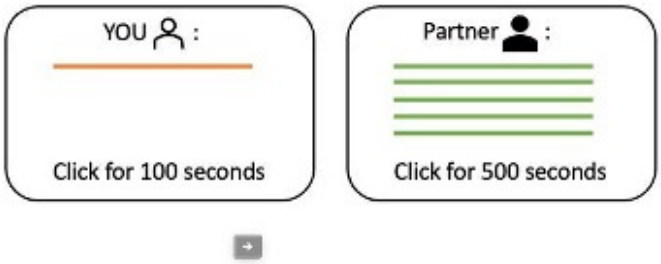
For the other half of the subjects in the motivated reasoning treatment, they were first shown the revelation block as in the screenshot titled "You & Your Partner's Clicking

Time in Stage 3", and then elicited beliefs about unpleasantness of the hard versus easy tasks as in the screenshot titled "Stage 3 Instructions: Predict Relative Subjective Costs". All the contents of the two blocks remain the same, only the sequence in between the two blocks was switched.

Then all subjects come to the redistribution block. They were first introduced to the redistribution task with the following instructions.



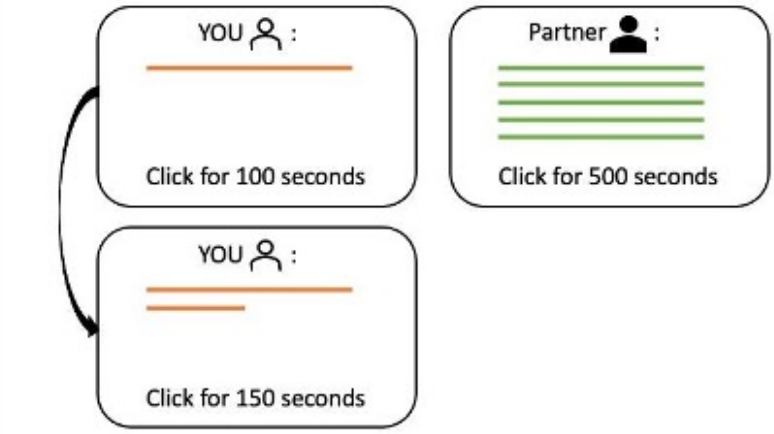
Your partner, on the other hand, needs to click for 500 seconds.



←

→

You could choose to volunteer to click for another 50 seconds. If you volunteer, you will have to click for 150 seconds.



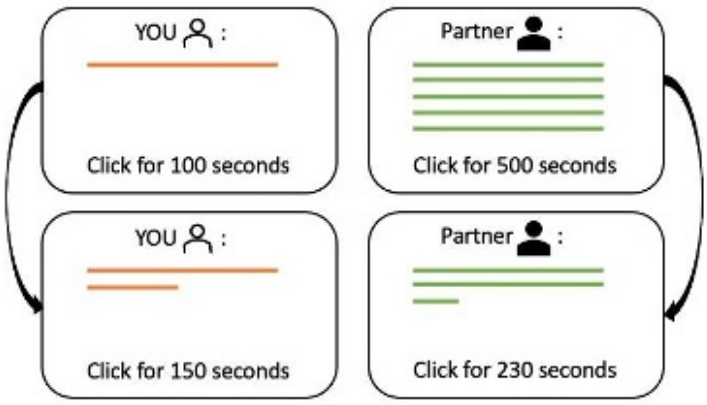
←

→

As an illustrative example, suppose the reduction in your partner's clicking time is set to be 270 seconds. In this case, if you decide to volunteer for an extra 50 seconds of clicking, your partner's click time will be shortened from 500 seconds to 230 seconds. This situation is shown in the figure below.

If you choose not to volunteer, you will have to click for 100 seconds. Your partner will end up clicking for 500 seconds.

If you choose to volunteer, you will have to click for 150 seconds. Your partner's click time will be shortened to 230 seconds.

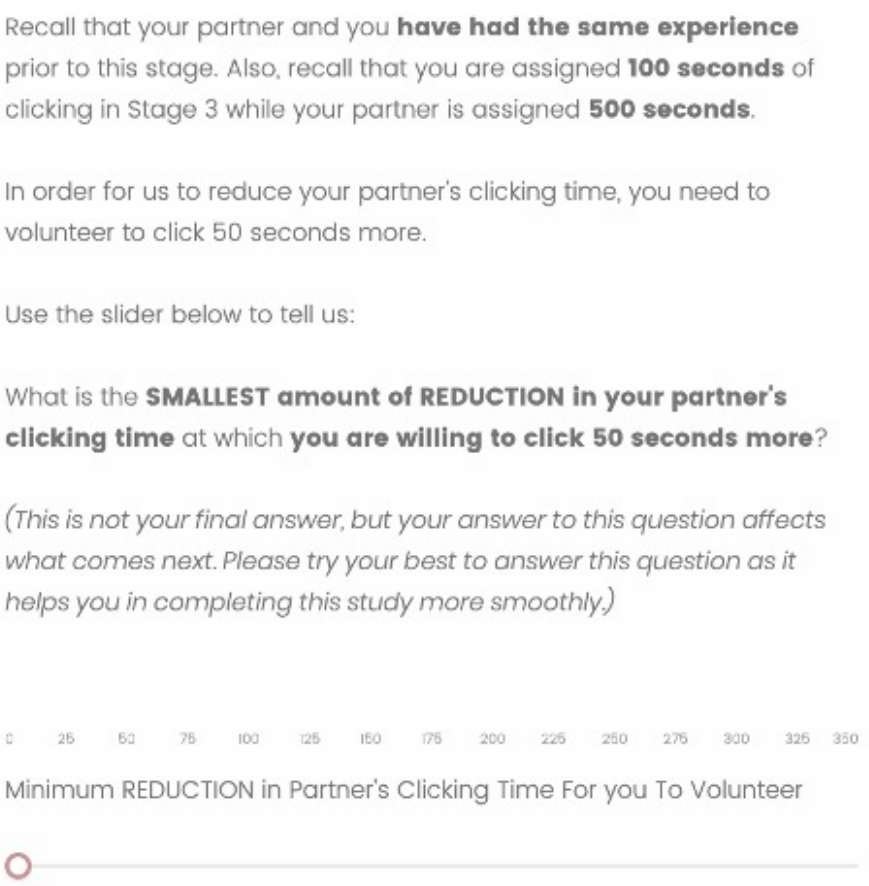


We haven't set the actual reduction in your partner's clicking time yet. Intuitively, *the higher the reduction* in your partner's clicking time is, *the more worthwhile it is for you to volunteer* to click for 50 seconds more.

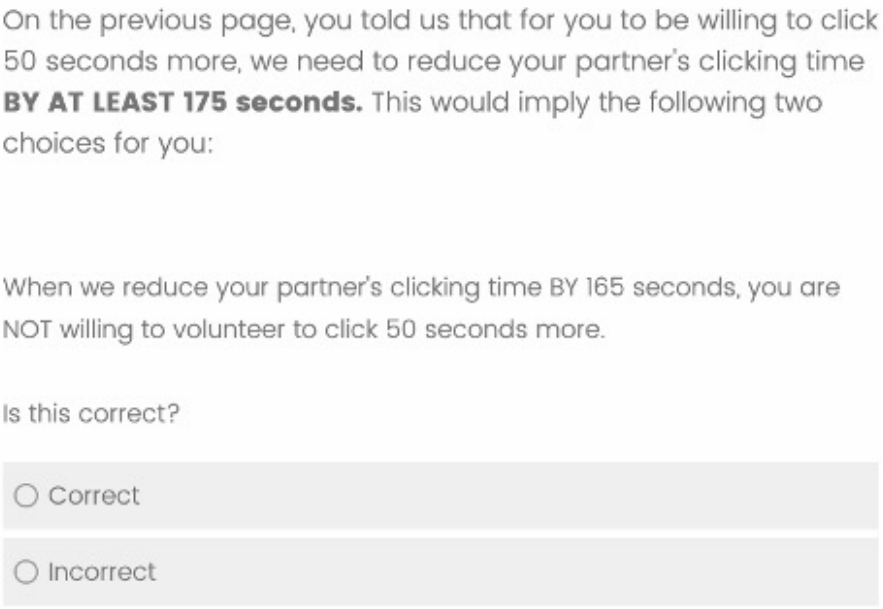
Starting from the next page, we will work with you to **find the smallest amount of reduction in your partner's clicking time for you to be willing to volunteer to click 50 seconds more.**

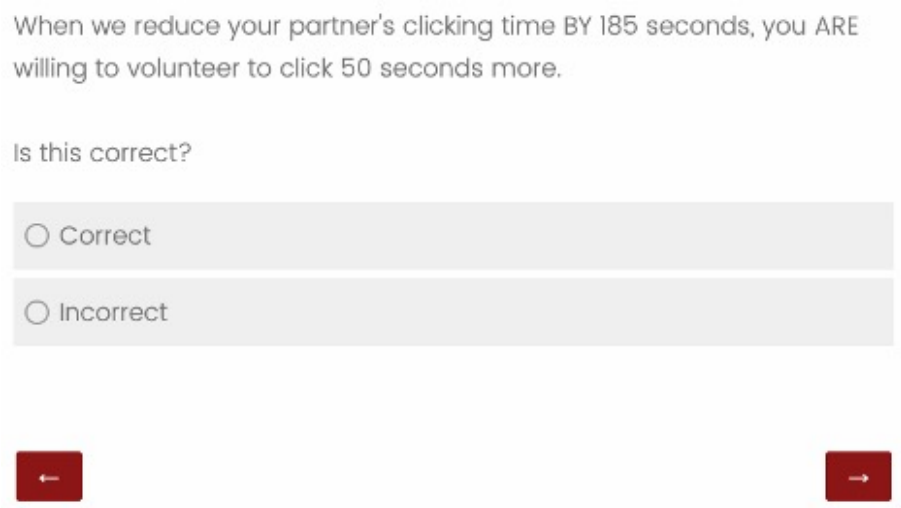
Now you will make real decisions whether or not to help your partner. Click next to begin with the first question.

They then make their redistribution decisions using a Titration-BDM mechanism. They first enter their reserve value to offer help in terms of partner's time reduction. See the following screenshots for the question.

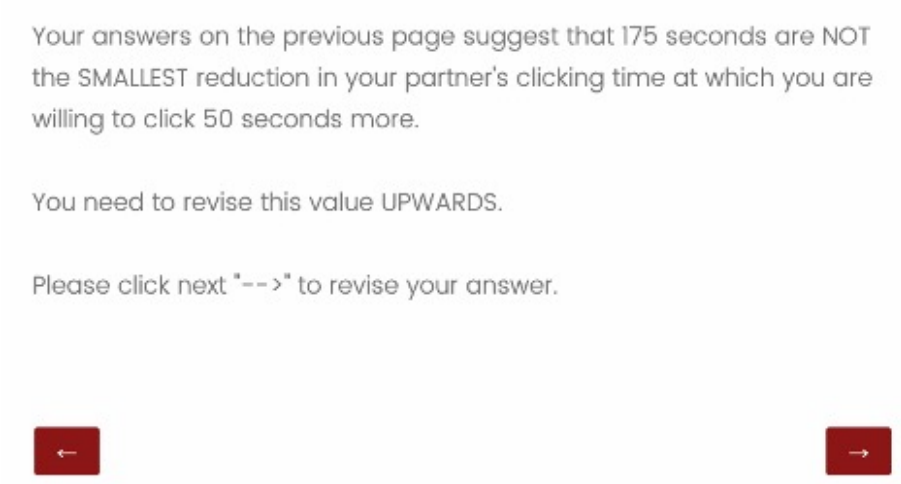


Based on their answer, they were then asked two confirmation questions as below.





Suppose they answered no to either one of the two questions, they were prompted to reconsider their reserve value in the message below. After seeing the message, they were redirected to the reserve value page.



Once they answered yes to both confirmation questions, they were shown a Multiple Price List (MPL) about whether or not to help their partner that was pre-filled for them based on their reserve value. They need to review the prefilled table and can make revisions if necessary. See the interface below.

Below is a table of action plans that we filled out for you based on your answers to the previous slider and yes/no questions. The table is about whether or not you will volunteer to click 50 seconds more for your partner to click certain seconds less.

Each row of the table corresponds to a certain reduction in your partner's clicking time.

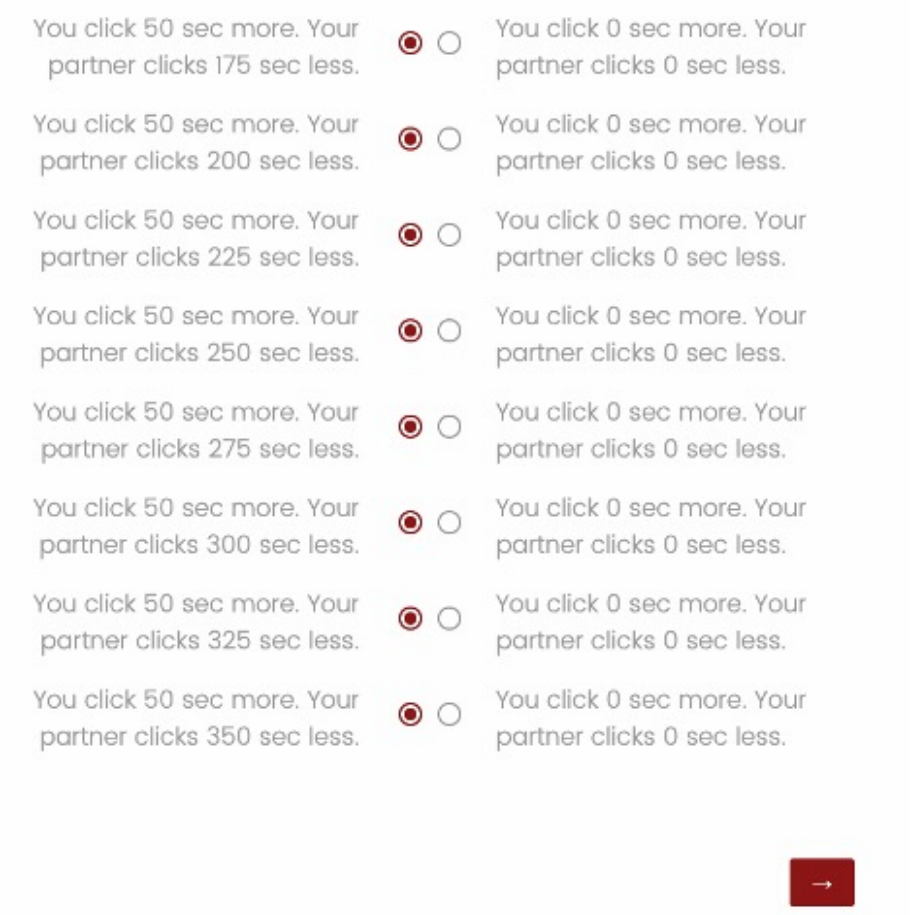
If the *left* button is selected on that line, for the specified reduction in your partner's clicking time, you AGREE to click 50 seconds more. Your partner's clicking time will be reduced by the number stated on that line.

If the *right* button is selected on that line, you DISAGREE to click 50 seconds more. And your partner will click the entire 500 seconds.

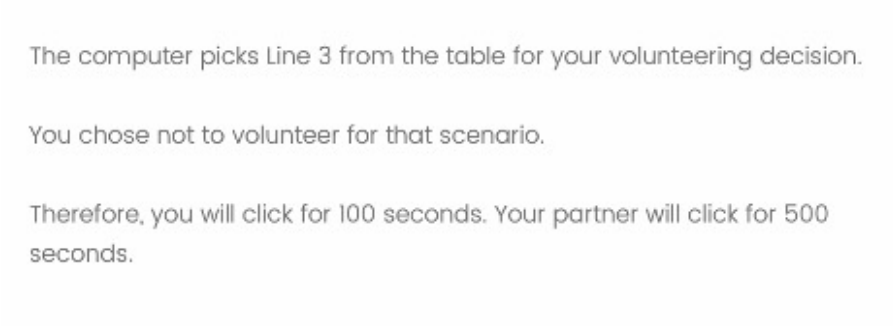
Please try your best to review the table below to make sure it is what you want to be implemented. You can edit the table if necessary. Once you confirm the choices, please click next "-->" to continue.

The computer will then draw a random line from the table, and implement your choice on that line. Please review the table carefully as it is going to determine you and your partner's clicking time in this stage.

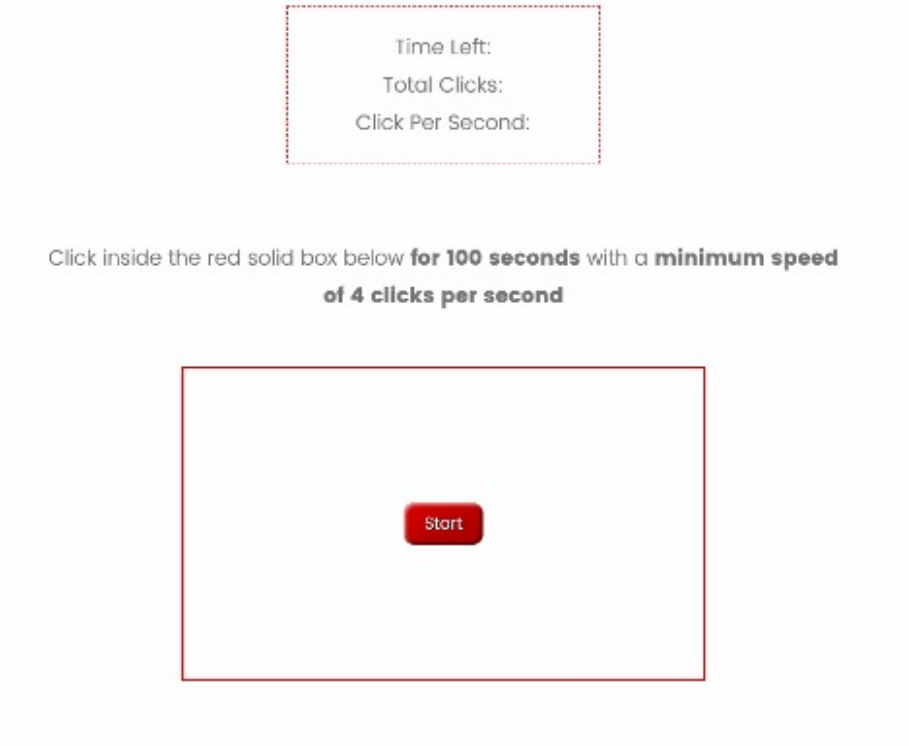
You click 50 sec more. Your partner clicks 25 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.
You click 50 sec more. Your partner clicks 50 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.
You click 50 sec more. Your partner clicks 75 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.
You click 50 sec more. Your partner clicks 100 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.
You click 50 sec more. Your partner clicks 125 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.
You click 50 sec more. Your partner clicks 150 sec less.	○ ◉	You click 0 sec more. Your partner clicks 0 sec less.



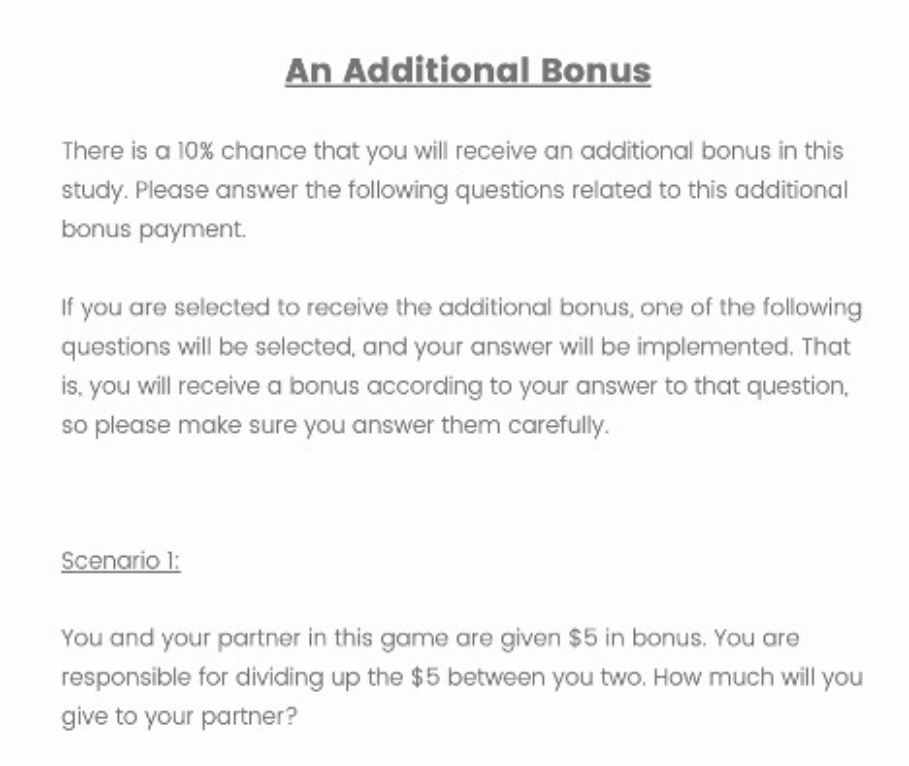
Then the computer randomly picks a line from the MPL and implement subject's choice. See the following screenshot.



Subjects were asked to complete either 100 seconds of clicking or 150 seconds of clicking, depending on whether they've chosen to help their partner and whether the computer picks a line where they chose to help.



Finally, after subjects are done with everything, they were surprised with a dictator game with another participant in this study. This is to measure and control for their altruism level. See the following screenshots for the questions.



Remember, if this scenario is selected, we will implement your decision below.

0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

How much of the $5 will you give to your partner?

Scenario 2:

You and your partner in this game are given $10 in bonus. You are responsible for dividing up the $10 between you two. How much will you give to your partner?

Remember, if this scenario is selected, we will implement your decision below.

0 1 2 3 4 5 6 7 8 9 10

How much of the $10 will you give to your partner?

Scenario 3:

There are 10 lottery tickets to win $5 in bonus. Each lottery ticket has a 10% chance of winning. You are responsible for dividing up the lottery tickets between you and your partner. How many lottery tickets will you give to your partner?

Remember, if this scenario is selected, we will implement your decision below.

0 1 2 3 4 5 6 7 8 9 10

How many lottery tickets will you give to your partner?

Scenario 4:

There are 10 lottery tickets to win $10 in bonus. Each lottery ticket has a 10% chance of winning. You are responsible for dividing up the lottery tickets between you and your partner. How many lottery tickets will you give to your partner?

Remember, if this scenario is selected, we will implement your decision below.

0 1 2 3 4 5 6 7 8 9 10

How many lottery tickets will you give to your partner?