What You Don't Know May Hurt You: A Revealed Preferences Approach*

Gonzalo R. Arrieta† Lukas Bolte‡

Preliminary Draft - Please do not circulate

Click here for the most recent version

October 12, 2023

Abstract

The dominant welfare approach is based on revealed preferences, which is restricted to settings where the individual knows their preferences have been fulfilled. We use a choosing-for-others framework to experimentally study welfare when what the individual believes to be true differs from what is actually true. We find substantial heterogeneity. About 40% of participants see welfare as independent of beliefs; 15% see welfare impact only via beliefs; and 45% exhibit mixed behavior. Our results suggest most people support the idea that welfare goes beyond awareness, which may inform media regulation, informational policies, and government communication.

I. Introduction

Welfare considerations are at the core of modern economic policy evaluations: A policy is a good policy if it enhances welfare. Hence, normative questions are central to good policy design. The dominant approach to those questions is rooted in the paradigm of revealed preference: Economists defer to choice and use revealed preferences as the welfare criterion. By this criterion, alternative x is deemed to be better than alternative y if and only if, given

*We are especially grateful to B. Douglas Bernheim for his guidance and encouragement and to Paul Milgrom and Alvin Roth for collaborating in the implementation of our experimental design. We also thank Muriel Niederle, Kirby Nielsen, and seminar participants at Stanford University for their helpful comments. This study is covered under Stanford University's IRB Protocol 44866 and Carnegie Mellon University IRB's #IRB-STUDY2015_00000482. The study was registered on the AEA RCT registry under ID AEARCTR-0011851 under the title "Red or Blue Pill? A Positive Welfare Analysis."

†Stanford University. E-mail: garrieta@stanford.edu.

‡Carnegie Mellon University. E-mail: lukas.bolte@outlook.com.

the opportunity, the individual would choose x over y. This paradigm equates well-being with preference fulfillment, which intrinsically relies on choice data.

However, a welfare criterion that has such a strong reliance on choice data is lacking as an account of well-being. In particular, using choice data to assess individual well-being essentially restricts our criterion to settings in which the individual *knows* those preferences have been fulfilled. This is by virtue of the fact that an individual cannot consciously make a choice for themselves without, in tandem, adjusting their beliefs about the fulfillment of the preferences underlying that choice. For example, when an individual chooses to consume alternative x, their beliefs naturally converge to believing they consume alternative x, and it is not possible to have them consciously choose to consume x at the same time that they delude themselves into the belief that they are consuming y. The act of choosing alerts the individual to the fact that the preference is being satisfied or not. This gives rise to the question: How do we think about welfare when the individual remains unaware that their preferences have been fulfilled? The relevance of this distinction is evident in the behavioral economics literature: We often study environments where individuals may misunderstand the consequences of their choices. In those cases, does well-being depend on what they believe the consequences to be, what the real consequences are, or both? The answer to this question fundamentally shapes normative economic analyses.[1]

This paper provides an answer to the welfare effects in situations where what the individual believes to be true differs from what is actually true. To overcome the aforementioned limitation of choice data, we resort to a choosing-for-others framework: an altruistic other party chooses between options affecting the individual but ones that hold what the individual believes fixed. Thus, we maintain the revealed preference paradigm as a basis for welfare assessment but extend the domain of applicable welfare questions through the choosing-for-others framework.

We create a binary state of the world over which a particular participant (the Receiver) has preferences; if asked, they would like the state of the world to be 1 (as opposed to 0). The state is constructed so that it is impossible for the Receiver to know which state it is unless he is told about it. We also minimize that anyone else but the Receiver cares about the state directly. We then ask the other parties to trade off a random surprise bonus for the

[1]One illustrative example is "The parable of the oblivious altruist" by Bernheim and Taubinsky (2018). *A small town in Arkansas experiences massive flooding, leaving many families homeless. To provide financial assistance for the impacted families, the government raises taxes, including a $100 levy on Norman. As a general matter, Norman thinks government spending is wasteful, but he is also an altruist, and would gladly contribute $100 to the fund if he knew about it. However, he never learns about the flood or the relief effort. Does the government's policy make him better off or worse off?* Of course, it is impossible for Norman's choices to inform the planner in this setting: if he chooses the government policy, he would also learn that the government has such policy, contrary to what the thought experiment's premise.

Receiver with whether the state is 1 or not. Since the state is constructed to be only (directly) meaningful to the Receiver, the (altruistic) other party makes choices that, according to the revealed-preference paradigm, maximize the Receiver's welfare. Crucially, we elicit these welfare assessments for two cases: one, where the Receiver learns about the state—which is a relatively standard elicitation—and another, where they do not.

Our main finding is that roughly a third of participants make the same welfare assessment whether or not the receiver learns about the state; a third acts as if there is no welfare impact of the state on the receiver unless the Receiver learns about the state. And a third think that the welfare impact is less, but not zero, when the Receiver does not learn. Secondly, by exogenously varying what the receiver believes when they do not learn about the state—they could think the state is most likely to be 0 or most likely to be 1—we also show that some participants act as if welfare came from beliefs being as accurate as possible. These findings hold both in the full sample and in a subsample screened for quality.

We validate our welfare measure, asking participants to make unincentivized welfare judgments. We find a positive correlation between the unincentivized questions and the incentivized elicitation, giving us confidence in our welfare types classification and suggesting, to some extent, the context independence of welfare notions. Our unincentivized welfare judgments also suggest that there may be a cheap and quick way to elicit welfare types, which may be useful in assessing policy impacts for particular populations. The individual unincentivized measures are also interesting in themselves: our incentivized measure pairwise correlates with the experience machine thought experiment and with a real-world welfare judgment, but not with the policy assessment (see details in section III.A.4).

The plethora of implications of the composition of welfare has been acknowledged by philosophers for thousands of years, which has given rise to a diversity of theories of well-being, making this a fundamental philosophical question with little to no robust evidence to answer it. There are two broad classes of theories of well-being as suggested in the philosophical literature.[2] On the one hand, *Welfare Hedonism* proposes that "well-being consists solely in the presence of pleasure and the absence of pain" (e.g., Bentham 1789; Mill 1861). A common variant of this notion of welfare is *Mental Statism*, which postulates that well-being is exclusively a reflection of mental states. On the other hand, *Preference Theory* postulates that "well-being consists in having one's preferences satisfied." This welfare notion asks whether the world is as the individual would like it to be rather than whether they

[2]Classification from Parfit (1984) and descriptions taken from Kagan (1998) and Bernheim and Taubinsky (2018). Bernheim and Taubinsky (2018) describes a third class of "Objective theories" that we don't directly connect to, by which "well-being is a matter of having certain goods in one's life, goods that are simply worth having, objectively speaking," irrespective of whether one prefers them or not (Aristotle 2011 translation; Sen 1985).

believe this to be the case. Generalized versions of preference theory allow for the possibility that the individual's preferences encompass their own mental states, which may depend on their understanding of outcomes. Generalized versions of preference theory posit empirical complications: eliciting preferences over mental states of mind can be challenging, in particular over deluded ones. Preference theory is the welfare notion predominant in modern welfare economics for its natural relation to choice data. We resort to a choosing-for-others framework to make progress in this respect.[3]

There is a vast literature in experimental philosophy and psychology that discusses these notions of welfare. A lot of this work focuses on the experience machine, a famous thought experiment by Nozick (1974). The literature has extensively discussed what this thought experiment is conceptually capturing, and the experimental work has progressively refined experiments that try to account for potential biases when asking participants what they would do if facing the machine, with inconclusive evidence about the role of mental states (Baber, 2008; De Brigard, 2010; Smith, 2011; Weijers, 2013; Rowland, 2017; Hindriks and Douven, 2018). For example, Weijers (2014) finds that 16% of participants would enter the experience machine, and Löhr (2019) find that 28% would enter the machine permanently. We contribute to this discussion about the role of mental and external states in welfare with an incentivized controlled experiment that tests whether mental states are all that matters for welfare and finds that the vast majority of participants' responses suggest external states play a relevant role.

Our paper contributes to a growing literature on positive welfare economics, in particular to the branch which aims to determine how people evaluate the well-being of other individuals and groups, mostly from a paternalistic lens (e.g., Uhl 2011; Ambuehl et al. 2021, 2023; Bartling et al. 2023). Notably, Ambuehl et al. (2021) experimentally studies when, why, and how people intervene in others' choices and find that choice architects intervene in others' choices as if they seek to align others' choices with their own aspirations. This finding supports an interpretation of our choosing-for-others results as reflective of how individuals think of their own welfare in terms of mental and external states. While understanding how individuals think about others' welfare is important to interpret policy, to *inform* policy, we want to more directly get at what welfare *is*, which is better captured as a measure for the self.[4]

[3]Relatedly, it is worth noting that assuming a preference theoretical approach to welfare does not force the planner to take a stance regarding the role of mental states but rather defers this stance to the individual. This conceptual feature of preference theory is, however, difficult to implement since it is often not feasible to observe the choice that the planner might need the agent to make to discover their preferences.

[4]We run a survey on the Social Science Predictions Platform (DellaVigna et al. 2019; Public Study ID sspp-2023-0032-v1 at www.socialscienceprediction.org) to capture our current understanding about the role of external and mental states for welfare. Predictions, while heterogenous themselves, to a large extent capture

A large literature documents individuals motivated to hold particular beliefs—i.e., mental states—(Bénabou, 2015). For instance, individuals may be motivated to think highly of themselves as a source of ego utility (Köszegi, 2006), or about a state of the world to derive utility through anticipation (Bénabou and Tirole, 2002; Brunnermeier and Parker, 2005; Caplin and Leahy, 2001). However, these models are typically silent on whether such desire to manage one's mental state is welfare-enhancing or a mistake; is the decision-maker's objective normative or not? By showing that, for most people, external states are not all that drives welfare, our results open the door for beliefs in general and biased ones in particular to drive welfare.

The paper proceeds as follows. Section II presents a conceptual framework which makes the distinction between mental and external states, highlights the limitations of the existing revealed choice paradigm as an underpinning of welfare economics, and shows which type of welfare statements we study. Section III presents our experimental design, Section IV the results, and Section V concludes.

II. Conceptual Framework

The revealed preference paradigm typically used to make welfare statements is only applicable when choice data exists or can plausibly be gathered. However, there are some choice problems that are impossible to state: we cannot elicit someone's choice between alternatives, *while holding their beliefs about which alternative realizes fixed*.[5] We consider a framework of such choice problems, i.e., where choices consist of "mental states" (e.g., what the person believes) and "External states" (what is actually true), to highlight the limits of the standard paradigm. We then use this framework to define types of preferences over this larger choice set (welfare types), whose prevalence we measure in our experiment in a choosing-for-others framework.

Let $x \in X$ be a set of goods. Let $\mu \in \Delta(X)$ be a distribution over the set of goods. We refer to x as the *external state* and μ as the *mental state* (of the individual). We want to know how the bundle (x, μ) affects the individual's welfare, i.e., we want to learn about $\mathcal{W} : X \times \Delta(X) \to \mathbb{R}$.

The revealed preference paradigm underlying much of welfare economics assumes that individuals know what is best for them, and so $\mathcal{W}$ is estimated by giving the individual choice problems. Many choice problems involve choosing from a set $A_{aux} \equiv \{(x, \delta_x) \in A \times \Delta(A)\}$, for some $A \subseteq X$, where δ_x places all weight on x. Here, the individual chooses her preferred

the heterogeneity in welfare notions that we find in our data. Predictors underestimate, however, the degree to which participants behave as if external states drive welfare, conditional on beliefs.

[5] e.g., in the exercise described in footnote 1, we cannot have Norman choose between the government offering disaster relief or not, *while holding Norman's beliefs fixed*.

good, and whatever she chooses, she must believe.[6]

However, such choice problems only give us the value of $\mathscr{W}$ for a restricted domain. To illustrate, suppose $X = \{0,1\}$. Our focus is on choices over $\{(1,\mu),(0,\mu)\}$, where μ denotes the probability of the external state being 1. Giving this choice problem to someone is not feasible—when the DM chooses $(1,\mu)$, she knows that the external state is 1, and so her mental state cannot be μ (unless $\mu = 1$, but then the DM cannot choose $(0,\mu)$).

To overcome this problem of the relevant choice data not existing, we consider a choosing-for-others framework. Suppose an altruistic and otherwise disinterested third party makes the choice on behalf of the individual. In particular, we assume that they, too, maximize $\mathscr{W}$—our experimental design aims to minimize all other considerations. Then we can, in fact, study such choice problems. This is because it is perfectly feasible for a third party to make choices over someone else's $\{(1,\mu),(0,\mu)\}$, with $\mu \neq 1$ (i.e., the third party can choose x for another person while keeping the other person's mental state fixed).

What could preference look like here? We focus on the case $\mathscr{W}(1,1) > \mathscr{W}(0,0)$, i.e., the individual would choose $(1,1)$ over $(0,0)$, which, given the assumed revealed-preference paradigm, is equivalent to $(1,1)$ increasing the individual's welfare relative to $(0,0)$.

But where does the increase in welfare come from? Two arguments are changing—the individual's external state and the individual's mental state. Below, we discuss a few reasonable cases that will form the basis of the types we elicit later. These cases strongly relate to the philosophical welfare notions we mention in the introduction.

Mental state is all that matters. One such type is that $\mathscr{W}(x,\mu)$ is independent of x, i.e,. only μ, the individual's mental state, matters. The intuition is simple: *what you don't know can't hurt you*. This corresponds to mental statism, as discussed in the introduction.

External state is all that matters. Conversely, it may be that $\mathscr{W}(x,\mu)$ is independent of μ, i.e., only x, the individual's external state, matters. In this case, the individual's welfare is affected by what is actually true and not what the individual believes to be true.

It could also be that both the external and the mental state matter for the individual's welfare. Let us distinguish this case further. A simple case would be $\mathscr{W}(x,\mu) = ES(x) + MS(\mu)$, where both ES and MS are increasing. Here, the DM's welfare is affected by both states and independently.

Another possibility is that there is a dependence between external and mental states for their impact on welfare. For instance, it might be that external and mental states are

[6]More generally, the DM chooses from a set $\lambda \in \mathscr{C} \subseteq \Delta(\{(x,\mu) \in A \times \Delta(A)\})$ such that the marginal distribution of λ over x equals the expected value of μ, i.e., the DM chooses a distribution over mental states and external states, where the expected mental state must coincide with the distribution over external states. Note that this allows for the option of the individual's realized mental state not being degenerate, i.e., the individual does not learn the external state although it is realized.

complements, i.e., $\mathcal{W}$ has increasing differences.[7] One interpretation here is that "incorrect beliefs," e.g., believing the external state is 1 whereas it is actually 0, are bad for welfare (it as bad to "live a lie"). Given this welfare consideration, it even makes sense to have $\mathcal{W}(1,\mu) < \mathcal{W}(0,\mu)$ for low μ. Of course, the welfare impact of a mismatch between external and mental state could be asymmetric: e.g., overly optimistic beliefs could be more or less harmful than pessimistic beliefs relative to accurate ones.

Our experiment allows us to shed light on how x and μ enter $\mathcal{W}$. In particular, we go beyond the usual elicitation measuring $\mathcal{W}(1,1) - \mathcal{W}(0,0)$ and also consider, $\mathcal{W}(1,\mu) - \mathcal{W}(0,\mu)$, varying μ. These preference considerations guide our analysis of the experimental results in informing our classification of participants into types, as discussed in section IV.

III. Experimental design

We measure which welfare notions individuals adhere to using an incentivized online experiment. To get at welfare notions, it is essential that participants make decisions driven by particular considerations (e.g., altruism) that can only reliably be triggered in a controlled experiment. Naturally occurring data does not allow us to measure welfare notions because of the myriad of confounders that take place outside the lab. Moreover, our experiment uses a choosing-for-others framework for the reasons explained in Section II. In this section, we describe our experiment in detail and discuss the role of our design choices, which enable us to use our experiment as a proper measuring tool to get at participants' welfare notions.

III.A. Environment and treatments

Our experiment consists of a group of participants who make altruistic surrogate choices for one individual participant whom we call the "Receiver." Different welfare notions make different predictions for the choices that participants might make on behalf of the Receiver. The main outcome of our experiment is a classification of participants into types defined by the welfare notions their choices are consistent with.

The main structure of the experiment has an external state (ES) over which the Receiver has preferences. To study the role of the ES and the mental state (MS) in welfare, we vary whether the Receiver's MS changes or remains fixed (i.e., we vary whether the Receiver will learn whether their preferences got satisfied or not) and elicit our participants' preference to change the ES (i.e., to satisfy the Receiver's preferences) in each case.

In our baseline experiment, the Receiver's MS *level* remains uncontrolled (i.e., the Receiver has some prior belief on the likelihood that their preferences are satisfied, which our

[7] For $\mu' > \mu$, $\mathcal{W}(1,\mu') + \mathcal{W}(0,\mu) \geq \mathcal{W}(1,\mu) + \mathcal{W}(0,\mu')$.

participants do not know). In two additional treatments, we test the extent to which participants want to match the ES to the MS by fixing the Receiver's MS to a level that our participants know and comparing the participants' preference to change the ES for the two different levels at which we fix the MS.

III.A.1. The External State

To enable us to assess which welfare notion participants adhere to, our choice of ES must satisfy the following three requirements. Firstly, we need an ES over which the Receiver is not indifferent. This introduces altruistic motives into the participants' objective. Altruistic motives are essential in that we cannot study which welfare notion participants adhere to if the decisions they make do not concern the Receiver's welfare.

Secondly, we need an ES over which *only* the Receiver is not indifferent. This guarantees that *only* altruistic motives enter the participants' objectives. Otherwise, there would be a confounding motive in interpreting the participants' decisions. For example, to the extent that the ES created consequences over which individuals might have strict preferences that operate independently of the impact they have on the Receiver's welfare (i.e., selfish considerations), this would introduce a confounding motive in interpreting the participants' decisions. This allows for the ES to have consequences over which participants (or anyone else) have strict preferences, as long as these preferences operate via the Receiver's welfare (e.g., "warm-glow"). Note that it is important to communicate to participants that the Receiver has strict preferences over the ES for them to even want to satisfy their preferences in the first place. In doing so, we need to be careful not to make suggestions about whether the Receiver's preferences exclude or include the MS induced by knowing the ES is satisfied since this could bias responses. Hence, instead of explicitly saying the Receiver prefers the original notes to the fake notes, we implicitly suggest it by conveying that the Receiver is a fan of the authors and enjoys their work.[8]

Thirdly, we need an ES that we can implement while keeping the Receiver's MS fixed. Detaching changes in ES from changes in MS is the basic distinction that we study in this paper, and hence, it is essential for the interpretation of our results to convince participants that the Receiver will never know, unless we tell them, what the ES is (i.e., the Receiver will never know whether their preferences have been satisfied unless we tell them).

Our choice of ES reasonably satisfies these requirements. In our experiment, there are four books by two Nobel Laureates in economics: two copies of "Discovering Prices" by Paul Milgrom and two copies of "Who Gets What and Why" by Alvin Roth. Each book has a

[8]See instructions in the appendix section B for the exact implementation.

handwritten note dedicated to the Receiver.[9] In one copy of each of these books, the note was handwritten by the author: we say the note is original. In the second copy of each of these books, the handwritten note was copied from the original note: we say the note is fake. The fake note was copied from the original in such a way that makes them indistinguishable.[10] The Receiver will receive two books, one copy of each. They know that the handwritten notes that come with the book might be original or fake.

The ES is defined by whether the notes are original or fake.[11] The person in the role of the Receiver strictly prefers the books with the original notes, so we satisfy our first requirement by design.[12] Our choice of ES makes it implausible that individuals other than the Receiver have strict preferences over the notes in the books the Receiver gets, other than those operating via the Receiver's welfare. This satisfies our second requirement. To ensure the requirement is satisfied, we will return the copies of the books that the Receiver does not get to the authors. This minimizes concerns about the existence of strict preferences over the outcome of these other two copies. For example, participants could have strict preferences over the notes in the two other copies if we were to destroy or donate them, but it is unlikely they have preferences over the books the authors receive being original or fake, since they can reproduce the original notes for free. Our third requirement is satisfied by virtue of the original and the fake notes being indistinguishable.

III.A.2. Preference elicitation

Our primary outcome is a measure of the participants' intensity of preferences to give the Receiver the books with the original notes, as opposed to those with the fake notes. We get at that measure by means of a willingness-to-pay elicitation utilizing the Becker–DeGroot–Marschak method (BDM) in the form of a Multiple Price List (MPL). We ask participants how much they are willing to pay for the Receiver to get the books with original notes, as opposed to receiving the ones with fake notes. Importantly, this WTP elicitation is implemented using the Receiver's money. Specifically, we elicit the monetary bonus amount that we would have to give to the Receiver to make participants indifferent between having the Receiver get the books with the original notes or the books with the fake notes and the bonus. Appendix section B shows the specific questions we ask participants in our MPL. For the rest of the

[9] The notes on the books by Milgrom and Roth say "To Alex: I hope you enjoy reading about auctions! Paul Milgrom" and "For Alex: I hope you enjoy reading about market design. Alvin E. Roth," respectively. The real person participating in the role of Receiver is called Alex.

[10] We ensured this by having Paul Milgrom and Alvin Roth themselves "copy" their respective notes.

[11] More precisely, the ES is defined by a bundle of whether the notes in the books that the Receiver gets are original or fake and a monetary bonus amount. We ignore the latter for clarity in this section and discuss its —minimal— role in section III.A.2.

[12] Refer to the experimental instructions in appendix B to see how we communicate this to participants.

paper, when we say "WTP" we always refer to participants' WTP using the Receiver's bonus. Our participants do not receive any bonus and never make decisions that affect their own monetary payment.

Implementing a WTP elicitation using the Receiver's bonus payment minimizes selfish considerations and keeps the participants' decisions as pertaining to the Receiver's welfare (e.g., tradeoffs between money for themselves and notes for the Receiver could give rise to "altered responses" as those documented in Exley (2016)). However, note that, in principle, the Receiver could make inferences from the bonus payment they receive about the ES that is realized in a way that jeopardizes our control over their MS.[13] To prevent this from happening, we make the bonus the Receiver gets a *surprise* bonus. This obfuscates the mapping the Receiver might have from bonus payments to the realized ES.

We implement our MPL in a three-step procedure. First, we ask participants whether they would rather give the Receiver the books with original notes, the ones with fake notes, or whether they are indifferent. Next, for participants who do not select indifference, we ask them to pick between the notes they selected in the first step and the notes they did not select with an added $1 bonus. This makes for an easy-to-understand couple of questions that identify participants who are indifferent, those who prefer to give the original notes and those who, for some reason, prefer to give the fake notes. Finally, for participants who prefer the original notes over the fake notes and a $1 bonus, we present them with a MPL to elicit how much they prefer to give the Receiver the original notes.[14] We implement this three-step procedure as opposed to implementing the MPL in one step because, to identify the types we care about, we are particularly interested in whether participants strictly prefer the original notes, are indifferent, or strictly prefer the fake notes. The intensity of such preferences is relevant yet second-order to its direction. Thus, we designed an adaptive elicitation procedure that first elicits this and then the extent of the strict preference.

We conduct our WTP three-step elicitation for two cases for each participant, presented in random order. In *Learns*, we elicit the WTP to give the Receiver the books with the original notes over the ones with fake notes for the case in which we will tell the Receiver which notes they are getting. In *NotLearns*, we elicit the WTP to give the Receiver the books with the original notes over the ones with fake notes for the case in which we will *not* tell the Receiver which notes they are getting. Participants see both cases on the same page. Both the cases and the options for each question in the first two steps of the three-step procedure are presented in random order.

[13] More precisely, this would be a concern to the extent that our participants believe it to be the case.

[14] The bonuses in the MPL span the following values: $2, $3, $5, $7, $10, $15, $25, $45, $70, $100, $140, $200. See appendix section B for details on the elicitation.

NotLearns is the main elicitation that allows us to test whether participants see welfare value in changing the ES while the MS is fixed. *Learns* serves multiple auxiliary purposes. Firstly, it allows us to identify participants who don't value satisfying the Receiver's preferences even when the Receiver learns about it. We need participants to see a welfare increase when both the ES and the MS change in order to test which welfare notion they adhere to. Secondly, the WTP elicitation when both the ES and the MS change serves as a benchmark for the WTP to change the ES when the MS is fixed, as elicited in *NotLearns*.

III.A.3. Varying the Mental State

In our baseline experiment, the Receiver's MS level remains uncontrolled (i.e., the Receiver has some prior belief on the likelihood that their preferences are satisfied, which our participants do not know). Types are determined by how they value changes in ES when MS are fixed, regardless of at which level they are fixed (i.e., regardless of how likely the Receiver believes it is that their preferences are satisfied). However, controlling the level of the MS allows us to refine our classification into types by testing whether participants' preferences over the ES depend on the level of the MS. For example, to test for a preference for the ES to match the MS (i.e., the Receiver's beliefs to be accurate) as much as possible, we need to know the level of the MS.

Therefore, in two additional treatments, we test the extent to which participants want to match the ES to the MS by fixing the Receiver's MS to a level that our participants know and comparing the participants' preference to change the ES for the two different levels at which we fix the MS. We randomly assign participants to the *HighMS* or the *LowMS* treatment. In the *HighMS* treatment, we tell participants that the Receiver "... knows that if we don't tell him which books he got, there is at least a 75% chance that they are the ones with the original notes." In the *LowMS* treatment, we tell participants that the Receiver "... knows that if we don't tell him which books he got, there is at least a 75% chance that they are the ones with the fake notes."[15]

Varying the level of the Receiver's MS in this way allows us to test for types that attempt to make the ES and the MS close to each other. For example, such a participant would have a higher WTP to give the Receiver the books with the original notes in the *HighMS* treatment than in the *LowMS* treatment. We discuss this further in section IV.B.

[15]We implemented these two treatments and the baseline treatment by randomly selecting one of them before approaching Alex, the Receiver. We truthfully informed Alex that with, e.g., 75% chance, he will have gotten the books with the original notes if we do not reveal the books, and with complementary probability which books he will have gotten will have been determined in some other way.

III.A.4. Unincentivized questions

At the end of the experiment, we ask participants three unincentivized questions. Two of these questions allow us to assess the external validity of the types that we identify in our experiment by correlating those types with responses in more naturalistic settings. The third question allows us to relate our types identified in the experiment in a choosing-for-others framework to how participants would make choices for themselves when considering their own welfare (Ambuehl et al. 2021 provide evidence that participants project their own preferences into others, suggestive that they might pick for others as they would pick for themselves).

Our first hypothetical question presents participants with the Experience Machine, a canonical thought experiment long discussed in the psychology and philosophy literatures on welfare notions and the role of mental states (Nozick, 1974). To our knowledge, this is the first and most studied attempt to assess mental-state theories of well-being (Baber, 2008; De Brigard, 2010; Smith, 2011; Weijers, 2013; Rowland, 2017; Hindriks and Douven, 2018). Our second question most directly relates to policy preferences by asking participants whether the policy described in "The parable of the oblivious altruist" by Bernheim and Taubinsky (2018) leaves the protagonist better or worse off. Our third question takes advantage of a real scenario where 1000 individuals received a drawing, knowing only one had the original copy and the rest had indistinguishable fakes. We ask participants whether the person who has the original drawing is better off getting the original one instead of a fake, even if they —or anyone else— will never know which they have.[16] Note that the Experience Machine can be interpreted as considering changes in MS for a fixed ES, while the other two scenarios vary the ES for a fixed MS, as does our experiment.

Participants answer these questions by selecting one of two answers to each question, where one answer is always more aligned towards a welfare notion that assigns more value to the MS and the other to the ES. Below we show the exact descriptions of the three settings we present to participants.

John A small town in Arkansas experiences massive flooding, leaving many families homeless. To provide financial relief to the impacted families, the government temporarily increases taxes, including a $100 levy on John. John lives far away and *will never learn about the flooding or the relief effort*. However, he cares about helping others and would gladly contribute $100 to the relief effort if he knew about the flood.

[16] See www.moforgeries.org for a detailed description.

Experience Machine If given the option, would you choose to plug into an experience machine that could provide you with an entirely immersive, simulated reality where you can experience any desirable scenario despite not being real? Keep in mind that while plugged in, you would never be aware that you are in the experience machine and would believe that the simulated reality is real.

Art collective Hundreds of Andy Warhol fakes, and one original drawing worth $20k, sold for $250 each. An art collective purchased an original Warhol drawing and copied it 999 times. The copies are carefully created so that not even their creators can tell them apart from the original drawing. They then mixed the original together with the copies and sold the 1000 drawings.[17]

III.B. Implementation and Recruitment Details

We recruited our main sample of 1478 participants on Prolific, an online platform frequently used for research studies, and randomized them into one of our three treatments.[18] Each participant received a $3 completion payment, and the median completion time was around 16.2 minutes. We pre-registered this study in the AEA RCT Registry (AEARCTR-0011851).

Participants receive ample instructions and are required to correctly answer understanding questions before proceeding to the main parts of our study. Rather than excluding participants, they are given as many times as needed to correctly answer the understanding questions. For full experimental instructions of all study versions that we run, see the Appendix B. [19]

IV. Results

IV.A. Empirical types

In this section, we analyze the experimental data to uncover the distribution of welfare types. We find evidence of substantial individual-level heterogeneity. Notably, a large share of participants report the same WTP for both the *Learns* (when the Receiver learns about the good) and *NotLearns* (when the Receiver does not learn).

[17] We further give participants a chance to watch the video on this website, describing the art collective's initiative, as well access to the website itself: `https://moforgeries.org/`

[18] We recruited all participants on August 23^{rd} of 2023. In order to qualify for our study, participants were required to be located in the USA and have a minimum of 100 prior submissions on Prolific, with a perfect approval rate. The experiment was implemented using the oTree platform (Chen et al., 2016).

[19] We ensure the truthfulness of our procedures by giving the books, as determined by the decisions of a randomly selected participant, to the Receiver, who satisfies the description in our instructions (e.g., is named Alex) and wishes to remain anonymous.

Table I shows our main classification into the types described in Section II.

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
Baseline N=497	0.20%	0.20%	0.20%	0.41%	14.40%	3.04%	3.65%	10.14%	29.01%	33.67%	5.07%
Low N=497	0.80%	0.00%	1.01%	0.80%	17.91%	2.21%	6.44%	8.45%	24.75%	31.99%	5.63%
High N=488	0.20%	0.00%	0.41%	0.20%	12.70%	3.48%	4.51%	7.58%	27.25%	35.86%	7.79%

Table I: Share of responses by type for the quality-restricted sample
Note: WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

We classify participants by treatment for our full sample. The results look the same using a high-quality sample where we use participants' open-ended responses to a question about why they responded the same or differently for both cases (i.e., *Learns* and *NotLearns*) to flag those who exhibit an evident misunderstanding of instructions (in particular, a misunderstanding of the difference between *Learns* and *NotLearns*). We begin with *Baseline* (the top row). Our main interest is in participants who have a strictly positive WTP in *Learns*, the last super column. This is because we need participants to see the books with real notes as welfare-improving for the Receiver to study which welfare notion they have.

Briefly, a negligible share (0.60%) of our participants have a negative WTP in *Learns*, which means they strictly prefer the Receiver to receive the books with fake notes over the books with real notes when they learn about it. Relatedly, 17.85% of our participants have a WTP of 0 even in *Learns*, which suggests they don't value giving Receiver the real notes enough for us to identify their welfare notion; we classify them as "Indifferent." Even if they don't care about the notes being original or not, participants who see welfare gains in ES and MS *matching* might be willing to pay in *NotLearns* even if they aren't in *Learns*, depending on their beliefs about Receiver's beliefs. We find 0.41% and 3.04% of participants with valuations consistent with this type for levels of their beliefs about the Receiver's beliefs higher and lower than 50%, respectively. Note that this group exhibits such preferences even if they think satisfying the Receiver's preferences is worthless in this case, conditional on the Receiver knowing about it, which suggests they see welfare value in beliefs matching the external state.

Now to our group of interest. Most participants (81.54%) have a strictly positive WTP in *Learns*, meaning they see the book with real notes as welfare-improving to the Receiver. We find that 10.14% of our participants have $ES = 0$ (and $WTP > 0$) and 33.67% have $ES = WTP$ (and $WTP > 0$). For illustrative purposes, assuming participants have a wel-

fare function in which Mental and External States are additive and monotonic, these types can be interpreted as *only* seeing welfare gains in variations of Mental States or External States, respectively ("Pure Mental Statists", and "Pure External Statist").

The modal type in our classification emphasizes the importance of the distinction between External and Mental States for welfare analysis. Modern economics firmly embraces preference theory, including generalized versions of it where preferences encompass mental states, which may depend on their understanding of what the external state is (Bernheim and Taubinsky, 2018). While most applications have mental and external states move in tandem, making the distinction between external and mental states moot, in many cases, individuals are not perfectly informed of the external state. The disconnect between Mental and External states is particularly relevant in cases in which the consequences of the planner's actions for a particular individual are not reproducible as consequences of actions when that individual is the decision maker: For example, it is impossible to elicit preferences over deluded states of mind without identifying and hence removing the delusion. In these cases, we need further refinements of the paradigm of revealed preference for welfare analysis. Our results provide evidence that suggests individuals support a paradigm that sees welfare as a matter of preference satisfaction, even when preferences do not encompass mental states.[20]

34.08% of participants are willing to pay positive different amounts to give the Receiver the books with original notes in both cases. These participants might be of the types that care about the MS and ES matching (i.e., beliefs being accurate), or they might simply see welfare value in both ES and MS, for example, consistent with a generalized version of Preference Theory where participants perceive the Receiver to have preferences over mental states. Even participants who are willing to pay the same in both cases might see welfare gains in ES and MS *matching*, depending on their beliefs about the Receiver's beliefs. In the next section, we use our additional treatments to shed light on how large each of these types could be.

IV.B. Testing independence

In this section, we study the extent to which the perceived level of the Receiver's initial beliefs (i.e., the Receiver's initial MS) matters for our WTP valuations. In *Learns*, varying the Receiver's MS (i.e., what the Receiver expects) can be interpreted as a reference point manipulation in a standard WTP elicitation. We find that such MS manipulations still

[20]This is not to say that the Receiver's preferences do not encompass their Mental States or that participants believed that to be the case, but rather that even if this was the case, more than a third of participants "still" see welfare gains in satisfying the Receiver's preferences.

affect the ES-WTP, even for *NotLearns*. In particular, in aggregate, participants exhibit some preference to match the ES to the MS. Thus, in this setting, loss aversion still exists, even if losses are not (consciously) experienced.

Table I's second and third rows show our types classification for our two other treatments. Recall that in *LowMS* (*HighMS*), participants know the Receiver is told there is at least a 75% chance (at most a 25% chance) they receive the books with fake notes. The share of participants classified into each type is mostly similar across treatments, suggesting our classification is robust and relatively stable to the perceived level of the Receiver's beliefs about his preferences being satisfied. One caveat to this stability is that participants value ES more in *HighMS* (in a FOSD sense, ordering types by ES-WTP), suggesting a preference to match ES with MS.[21]

Next, we explore whether the perceived level of the Receiver's beliefs affects the participant's WTP valuations beyond shifts in the share of types.

Figure I displays the distribution of ES valuations, conditional on $WTP > 0$, by treatment. We observe a similar pattern on the intensive margin as we did on the extensive margin (type classification): participants' willingness to pay to give Receiver the books with the real notes in both *Learns* and *NotLearns* is increasing in the perceived level of Receiver's beliefs about his preferences being satisfied, as manipulated in *LowMS* and *HighMS*.[22]

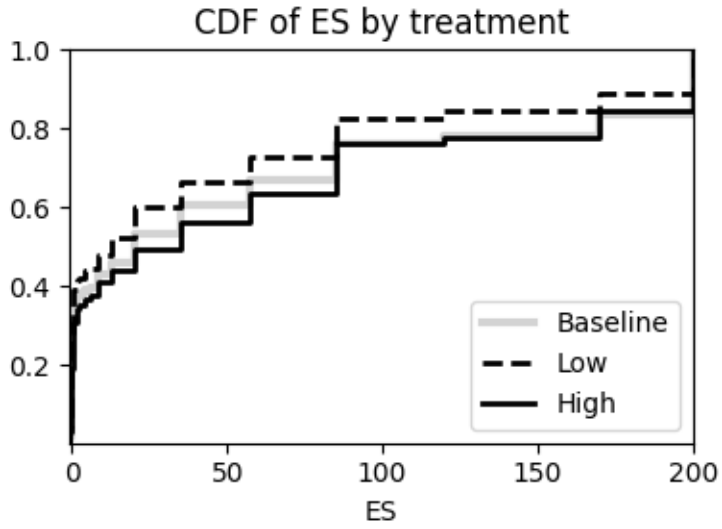


Figure I: CDF of ES-WTP by treatment.
Note: CDF of ES-WTP by treatment, conditional on $WTP > 0$ and $ES - WTP > 0$.

[21]This pattern holds regardless of whether one conditions on $WTP > 0$.

[22]Note that it is hard to interpret the baseline treatment in this sense since we have no control over the perceived level of the Receiver's beliefs.

This pattern, suggesting that participants have some preference to match the external state to the mental state, further supports the idea that some participants' welfare notions trade-off the satisfaction of preferences with the accuracy of the beliefs about those preferences being satisfied. This suggests participants identify a welfare cost in the mismatch between preference satisfaction and the beliefs about the satisfaction of those preferences.[23]

Note that while decisions in *NotLearns* directly affect the mismatch between preference satisfaction and the Receiver's beliefs, in *Learns*, we tell the Receiver whether their preferences are being satisfied, so no such mismatch exists.

Note that figure II shows we still observe the pattern by which participants are more willing to satisfy the Receiver's preferences if the Receiver expected that to be the case (i.e., in *HighMS*).

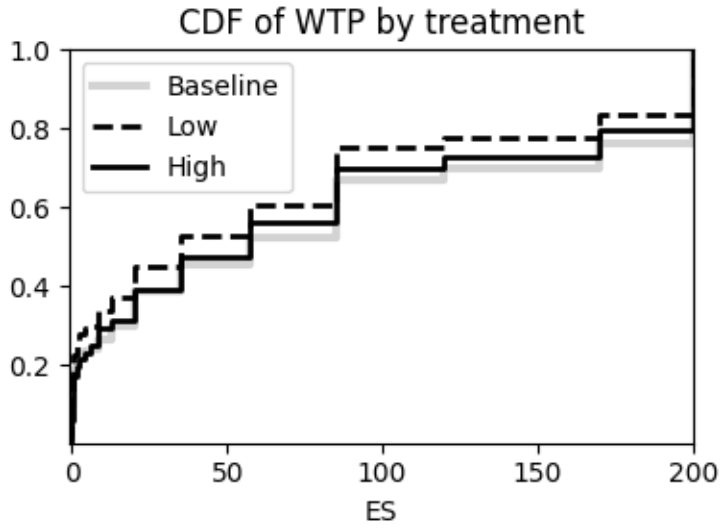


Figure II: CDF of WTP by treatment.

In *Learns*, ES and MS move in tandem, which means the elicitation is incapable of generating a mismatch, yet participants value matching the Receiver's prior MS. Manipulating the Receiver's prior MS as we do is plausibly affecting their reference point by changing their expectations about the books they'll receive. Hence, the preferences we identify are similar to those studied in the guilt aversion (Charness and Dufwenberg, 2006) and loss aversion (Tversky and Kahneman, 1991) literatures. In *Learns*, loss aversion can manifest itself in a standard way. However, the pattern consistent with loss aversion that we

[23]These welfare costs of a mismatch, together with a general preference to provide the real notes, could imply a net ES-WTP of 0 if the participant assigns the receiver a relatively low prior MS. Thus, some participants we classify as ES=0 may, in fact, have more nuanced preferences, in particular, preferences that depend on the ES. This case, however, is nongeneric.

observe in *Learns* is also present in *NotLearns*. This calls for a refinement in our interpretation of reference-dependence preferences: do references only pertain to what happens in the mind? This further speaks to in which settings we would or would not expect to observe reference-dependent preferences in general and loss aversion in particular. We assert that loss aversion, in particular, and reference-dependent preferences, in general, are typically understood as psychological phenomenons: e.g., the decision-maker has some expectation and then experiences "loss" or "gain" relative to those expectations. Under such an interpretation of reference-dependent preferences, they would predict no asymmetry in ES-WTP across our *LowMS* and *HighMS* treatments since there cannot plausibly be a psychological response to what the Receiver never learns about. Hence, how can we make sense of the reference-dependence-looking pattern we observe in figure I?

One way of making sense of this pattern is that reference-dependent preferences and loss aversion (LA) are not purely psychological phenomena, unlike how they are commonly understood. Under such an interpretation of reference-dependent preferences, "loss" or "gain" relative to the reference point can be "experienced" even if the reference point is psychologically driven (e.g., driven by beliefs) and without changing it.[24]

Another way of making sense of the pattern in figure I is that it is driven by "matching preferences," in which participants want the external state to match the Receiver's beliefs. Such preferences lead exactly to the pattern we observe, in which participants are WTP more to give the Receiver what they expect, even if the Receiver will never learn what they receive. Such preferences cannot be manifested in *Learns* since mental and external states always match by design, so even though we observe the same pattern in *Learns* and *NoLearns*, under this interpretation, they are driven by different reasons, which apply only to each particular domain. Crucially, our conjectured explanation for why we observe a pattern consistent with loss aversion in *NoLearns*, a setting in which loss aversion as typically understood has no bite, happens to, in turn, have no bite in *Learns*. Alternative explanations for this pattern that have a bite in *Learns* would act as confounders in our regular estimations of loss aversion.

The implications of these results are illustrated by "The parable of the oblivious altruist" mentioned at the beginning of the paper. Would the welfare assessment change, depending on what Norman expects the government to do in such a situation? Our findings suggest a lot of heterogeneity. Many individuals judge Norman's beliefs as irrelevant in assessing the impact of offering disaster relief on their welfare, and many deem the welfare impact to be larger the more Norman expects such a relief effort.

[24] In principle, this interpretation can allow for non-psychologically driven reference points. We don't engage in that discussion in this paper.

Next, we use our types classification to shed light on the LA pattern we find in figure II. Most participants are WTP for varying both MS and ES, so, naturally, the LA pattern that we observe in aggregate holds for this group; see appendix figure A.7. Next, we look at our two most extreme types: Figures III and IV show the CDF of WTP for the subset of participants that we identify as Pure MS and Pure ES. Interestingly, Figure III shows a reversed LA pattern. One interpretation for this pattern is that among the type that *only* cares about mental states, gains loom larger than losses. This makes evident the importance of exploring the heterogeneity of our results: isolating the more extreme mental statist types reverses the LA pattern that we observe for the whole sample and for participants who value both MS and ES, which is contrary to what we would have expected under the assumption that LA is a purely psychological phenomenon. Naturally, the extreme types shown in Figures III and IV are small, making the comparisons less reliable (e.g., a t-test of differences in means between our treatments does not allow us to reject equality of means). However, the evidence for the Pure MS group in figure III suggests it is unlikely that a LA pattern holds for this group.[25] Similarly surprising is the LA pattern we observe in figure IV. This evidence can be cautiously interpreted as suggestive that LA is not a purely psychological phenomenon.[26]

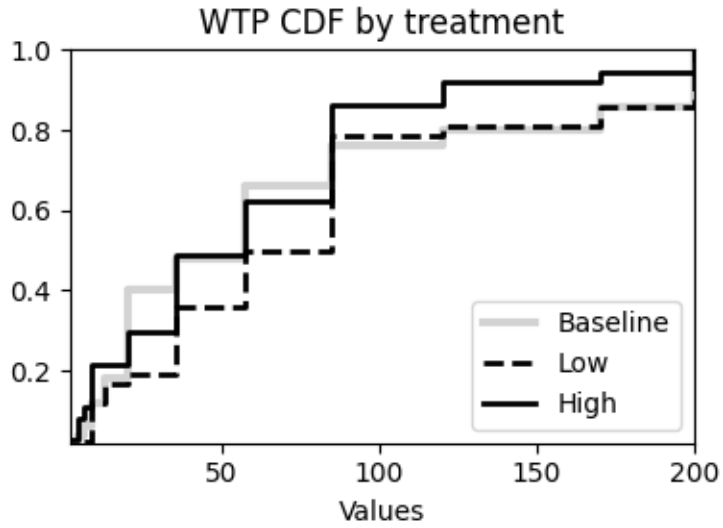


Figure III: Trad WTP for Pure MS

[25]A one-sided t-test testing whether the mean WTP in the *High* treatment is higher than the mean WTP in the *Low* treatment yields a p-value of 0.06. Thus, the LA pattern, the aforementioned comparison between the mean WTPs, is unlikely to hold in this subsample.

[26]Caution is in order because a one-sided t-test testing whether the mean WTP in the *High* treatment is lower than the mean WTP in the *Low* treatment yields a p-value of 0.09. Thus, the LA pattern, the aforementioned comparison between the mean WTPs, is relatively likely to hold in this subsample.

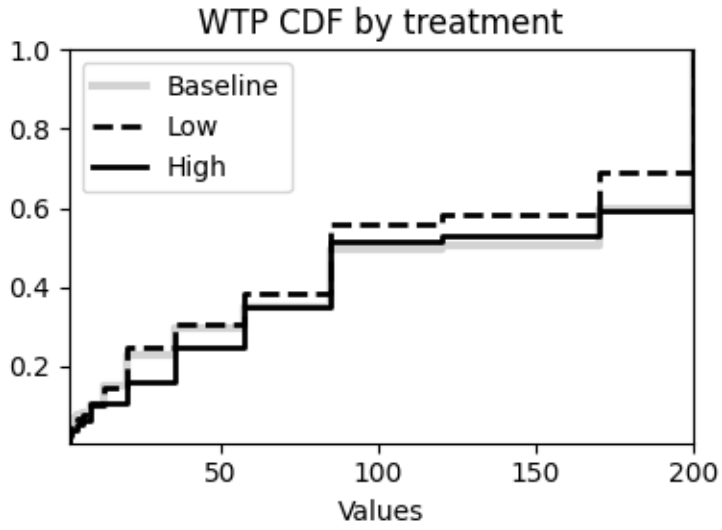


Figure IV: Trad WTP for Pure ES

IV.C. External relevance and choosing-for-self

In this section, we test whether our type classification correlates with unincentivized measures of participants' types. Indeed, we find that our type classification is predicted by our unincentivized measures jointly.

We use three hypothetical questions asked of all participants at the end of the experiment. Our first hypothetical question presents participants with the Experience Machine thought experiment (Nozick, 1974). Our second one asks participants whether the policy described in 'The parable of the oblivious altruist" by Bernheim and Taubinsky (2018) leaves the protagonist better or worse off. Our third question presents participants with a real scenario where 1000 individuals received a drawing knowing only one had the original copy and the rest have indistinguishable fakes and asks them whether this person is better off by getting the original one instead of a fake.[27] Note that the Experience Machine can be interpreted as considering changes in MS for a fixed ES, while the other two scenarios vary the ES for a fixed MS, as does our experiment.

Participants answer these questions by selecting one of two answers to each question, where one answer is always more aligned towards a welfare notion that assigns more value to the MS and the other to the ES. Because responses to any individual question are noisy, we focus on the subset of participants who consistently respond to all three questions, picking the option that better aligns with either MS or ES.[28] Figure V shows the CDF of the ratio

[27] See www.moforgeries.org/ for a detailed description.

[28] In particular, we find that participants' answers to the Experience Machine question and to the welfare

of ES-WTP to Trad WTP for these two groups. We condition on participants whose WTP is strictly greater than zero because these are the ones for which our type classification has bite, as we explain in section IV.A. Participants who consistently give unincentivized responses that align with a welfare notion that puts value in the ES exhibit a larger ratio across the whole of the distribution, suggesting they value changing the ES more than participants whose unincentivized responses align with a welfare notion that puts value in the MS.

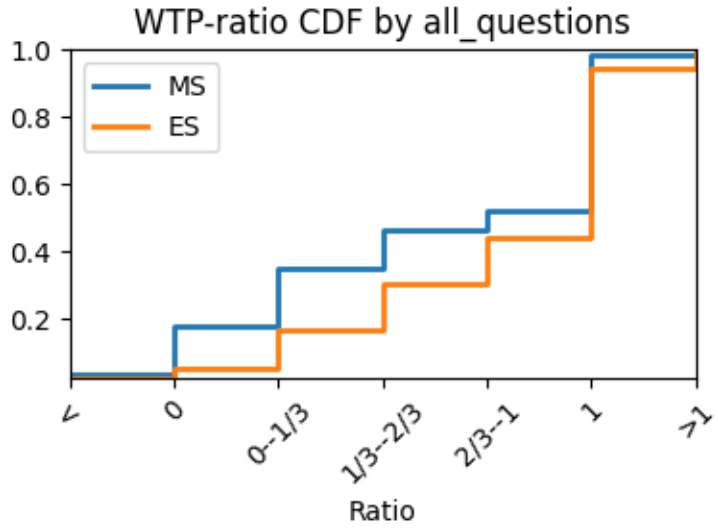


Figure V: Caption

Note: CDF of the ratio of ES-WTP to Trad WTP, conditional on $WTP > 0$, for participants in baseline who give responses to unincentivized questions that are consistent with welfare notions that assign more value to the MS or to the ES. 28.2% of participants in baseline who have $WTP > 0$ give such responses.

Note that the analysis in this section tests, to some extent, the stability of welfare notions. If participants' welfare notions were very domain-dependent, then we would expect little or no correlation between our experiment and the unincentivized questions. The evidence we present suggests that there might be some specificity to the welfare notions individuals adhere to in that the correlation we estimate is far from perfect, yet figure V strongly suggests we are capturing, at least for some participants, stable welfare notions that transport across the domains we consider.

judgment about the drawing predict their WTP elicitations, although we find that answers to a policy question are not predictive. See appendix figures A.1, A.2, and A.3 for the distribution of responses to each individual question.

V. Conclusion

This paper uses a choosing-for-others framework to test the role of mental and external states in welfare assessments. Understanding what composes welfare has ubiquitous implications. Economists usually think of individuals as solving problems in which they maximize their own welfare and policy-makers as maximizing others' welfare. Understanding what they are maximizing is of first-order importance, particularly in settings in which mental states and external states of the world do not move in tandem. People's awareness is limited, and what people are unaware of cannot affect mental states. Our results suggest most people support the idea that welfare goes beyond awareness.

Shedding light on the role of beliefs for welfare plays a key role in our understanding of the role that news media play on welfare and the value of non-instrumental information. Our findings speak about the welfare consequences of what happens "under-the-hood," and how (mis-)perceptions about these affect welfare. Thus, the evidence we provide can prove fundamental to media regulation, informational policies, and government communication. For example, our findings allow for a re-interpretation of the welfare impact of surveillance programs: while the surveillance versus privacy discussion usually focuses on its instrumental implications, much more is at stake if one takes a welfare notion that goes beyond mental states. More generally, the same point applies to a broader set of government policies that rely on an oblivious population, like those practiced by intelligence agencies and the military.

Similarly: can deceptive policies be optimal? For example, the literature on tax salience shows that an obfuscated tax can be less distortionary, suggesting that the government should obfuscate taxes (Chetty et al., 2009). However, our findings suggest that, to many, external states (e.g., what the taxes actually are), matter per se, independent of what the perceived taxes are, beyond budgetary concerns. We also find that a substantial share of people value knowing the truth regardless of what the truth is (i.e., even if it is undesirable), which speaks to the welfare consequences of transparency and obfuscation of government policy. Moreover, future work can address a more direct way to study the role of deception and policy obfuscation by considering a setting where the external state is held fixed, and testing the welfare consequences of changing the Receiver's mental state, for example, via the provision of non-instrumental information. If individuals have a preference for not believing falsehoods, policy prescriptions might become, at the very least, more nuanced, in terms of the use of deception.

We broadly speak to the value of information in the functioning of society: altruistic individuals who can benefit others at a cost might or might not do this depending on their

welfare notion. Mental statists would only take the costly action if the beneficiary learns about it, whereas external statists would take the costly action irrespective. Our results seem to suggest heterogeneous support for such behaviors.

Finally, the heterogeneity we observe in our results provokes questions about the stability of welfare notions and their determinants. Future work could explore these dimensions, for example, considering how culture and norms determine support for different welfare notions.

References

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

Ambuehl, S., Blesse, S., Doerrenberg, P., Feldhaus, C., and Ockenfels, A. (2023). Politicians' social welfare criteria: An experiment with german legislators.

Aristotle (2011). *Aristotle's Nicomachean ethics*. Translation by Bartlett, Robert C and Collins, Susan D and others. University of Chicago Press.

Baber, H. (2008). The experience machine deconstructed. *Philosophy in the Contemporary World*, 15(1):132–137.

Bartling, B., Cappelen, A. W., Hermes, H., and Tungodden, B. (2023). Free to fail? paternalistic preferences in the united states. *NHH Dept. of Economics Discussion Paper*, (09).

Bentham, J. (1789). Pml. *An Introduction to the Principles of Morals and Legislation*.

Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 381–516. Elsevier.

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

Bénabou, R. (2015). The Economics of Motivated Beliefs. *Revue d'économie politique*, 125(5):665–685.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Caplin, A. and Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55–79.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

De Brigard, F. (2010). If you like it, does it matter if it's real? *Philosophical Psychology*, 23(1):43–57.

DellaVigna, S., Pope, D., and Vivalt, E. (2019). Predict science to improve science. *Science*, 366(6464):428–429.

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

Hindriks, F. and Douven, I. (2018). Nozick's experience machine: An empirical study. *Philosophical Psychology*, 31(2):278–298.

Kagan, S. (1998). Normative ethics. *New York: McGraw Hill publishers*.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707.

Löhr, G. (2019). The experience machine and the expertise defense. *Philosophical Psychology*, 32(2):257–273.

Mill, J. S. (1861). *Utilitarianism*. Oxford University Press UK.

Nozick, R. (1974). *Anarchy, state, and utopia*. John Wiley & Sons.

Parfit, D. (1984). *Reasons and persons*. OUP Oxford.

Rowland, R. (2017). Our intuitions about the experience machine. *J. Ethics & Soc. Phil.*, 12:110.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland, Amsterdam. New Delhi: Oxford University Press, 1987; Italian translation: Giuffre Editore, 1988; Japanese translation: Iwanami, 1988.

Smith, B. (2011). Can we test the experience machine? *Ethical Perspectives*, 18(1):29–51.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 106(4):1039–1061.

Uhl, M. (2011). Do self-committers mind other-imposed commitment? an experiment on weak paternalism.

Weijers, D. (2013). Intuitive biases in judgements about thought experiments: The experience machine revisited. *Philosophical Writings*, 41(1).

Weijers, D. (2014). Nozick's experience machine is dead, long live the experience machine! *Philosophical Psychology*, 27(4):513–535.

A. Additional Figures and Tables

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
High-quality data											
Baseline N=406	0.00%	0.00%	0.00%	0.25%	10.10%	0.99%	2.46%	10.84%	33.50%	36.95%	4.93%
Low N=393	0.25%	0.00%	0.25%	0.25%	12.72%	1.53%	6.62%	9.16%	28.24%	35.62%	5.34%
High N=399	0.00%	0.00%	0.00%	0.00%	10.53%	1.50%	3.26%	8.77%	30.08%	40.10%	5.76%

Table A.1: Share of responses by type for the quality-restricted sample
Note: WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

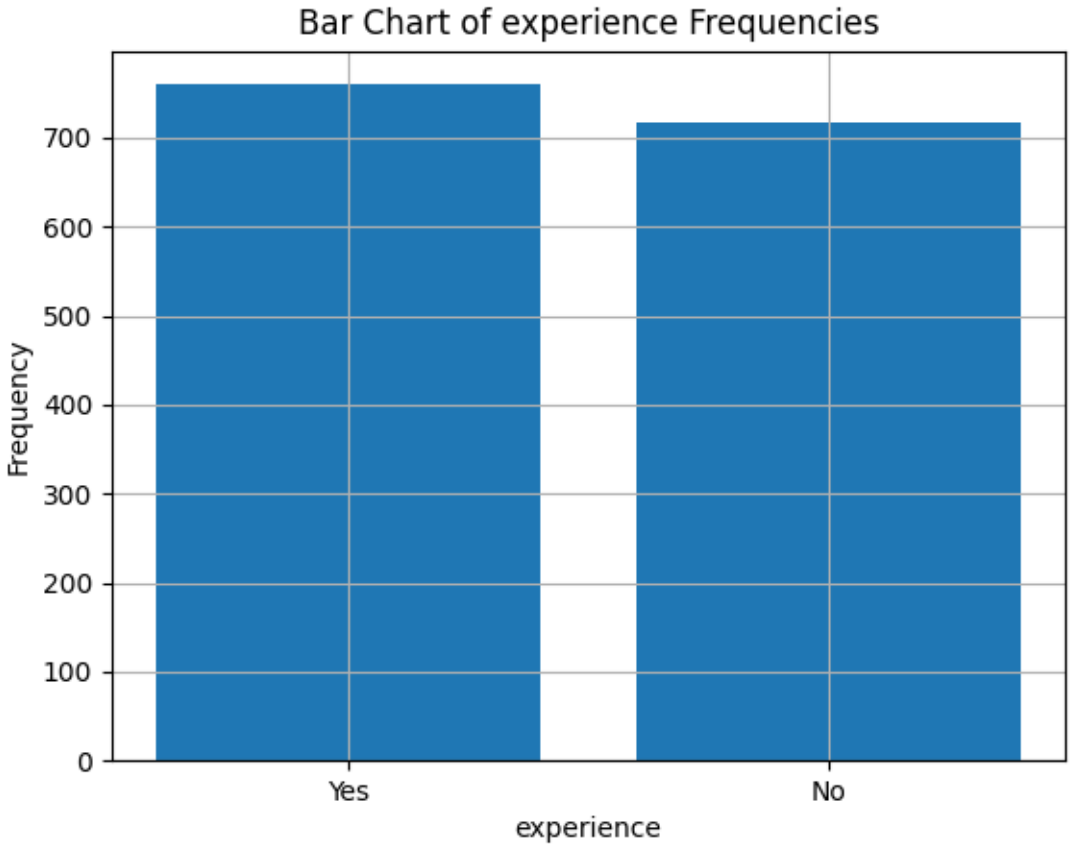


Figure A.1: Distribution of responses `Experience Machine`

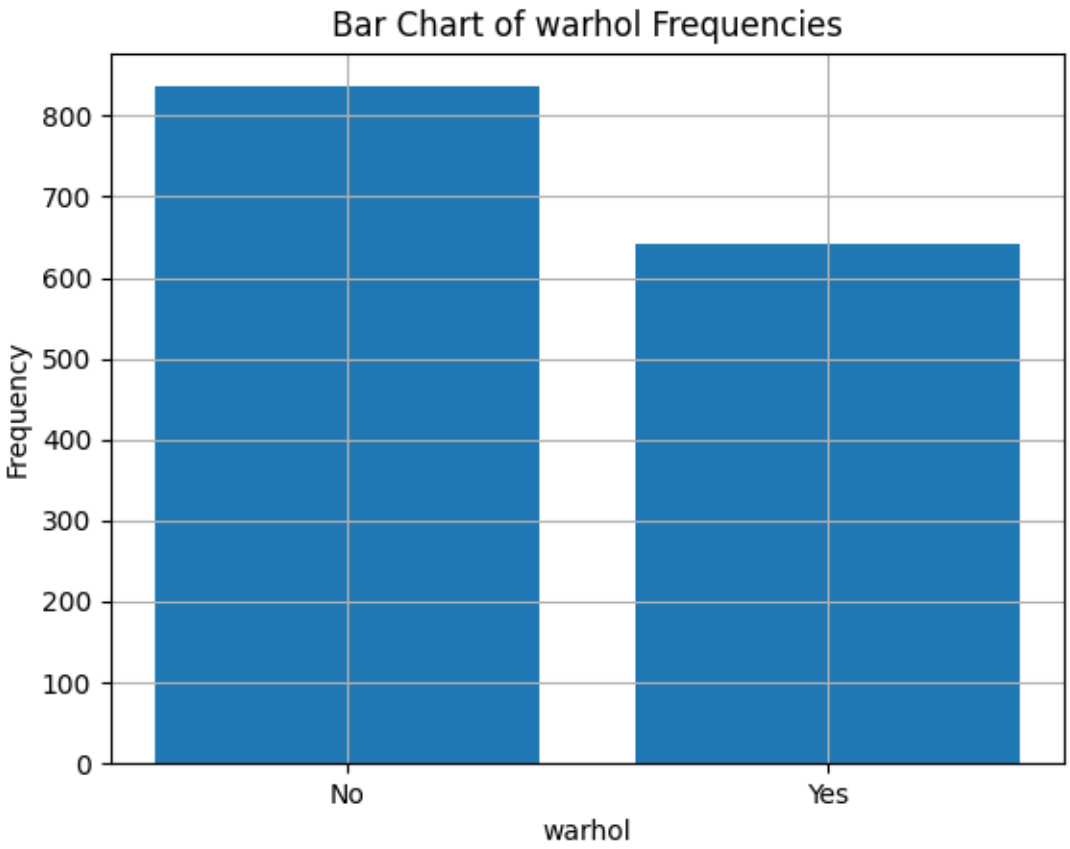


Figure A.2: Distribution of responses `Welfare Judgement`

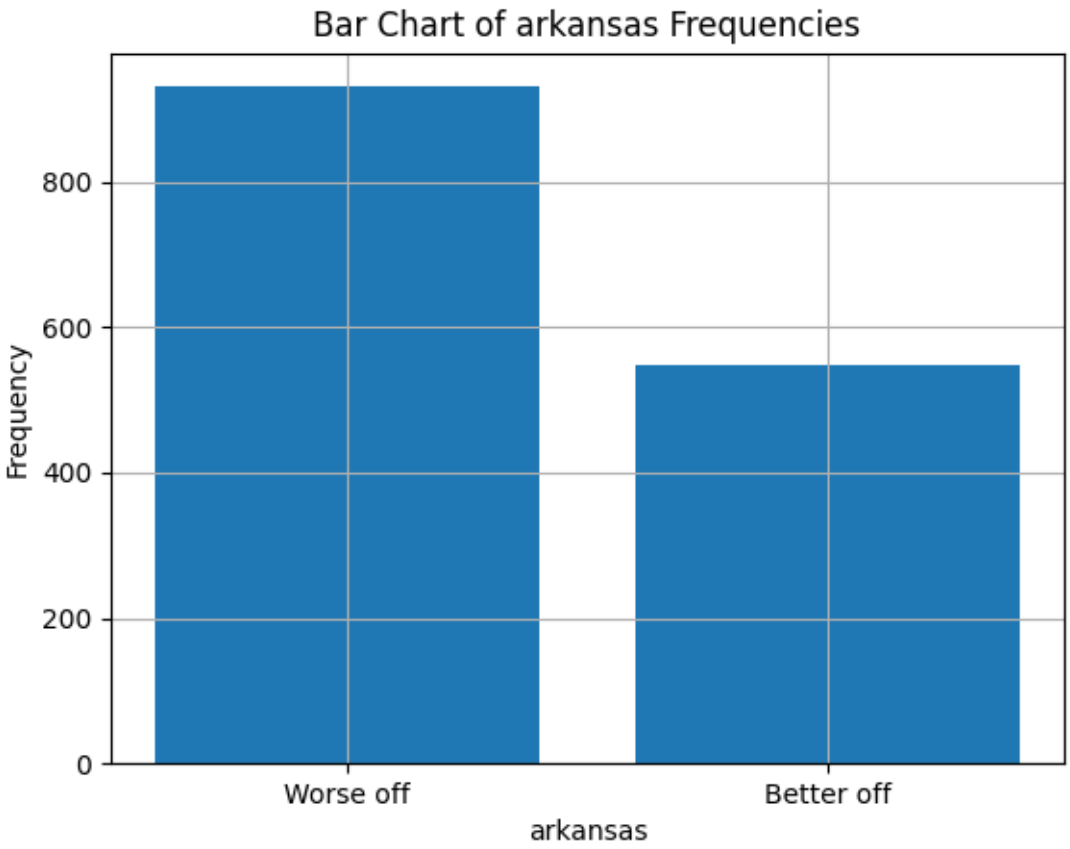


Figure A.3: Distribution of responses `Policy Question`

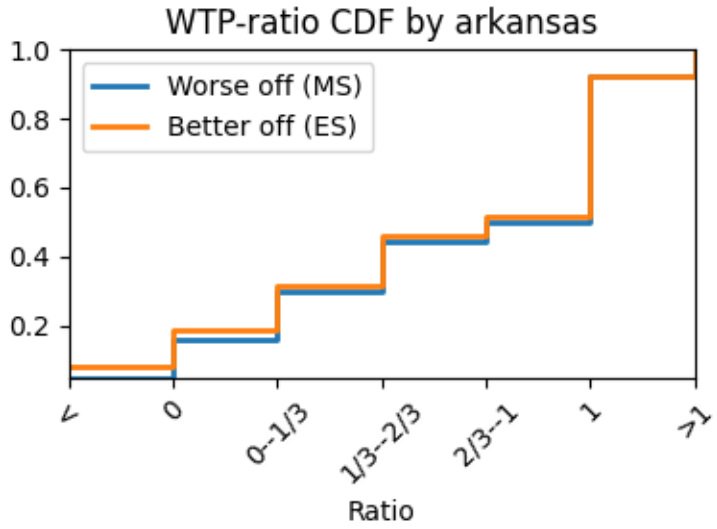


Figure A.4: Caption

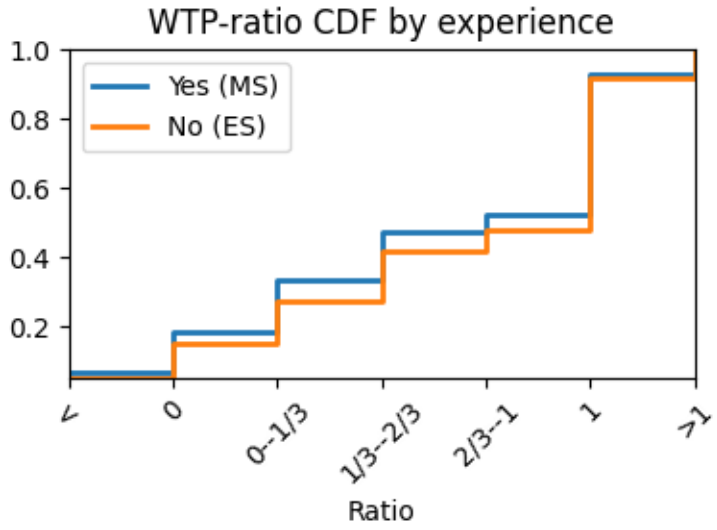


Figure A.5: Caption

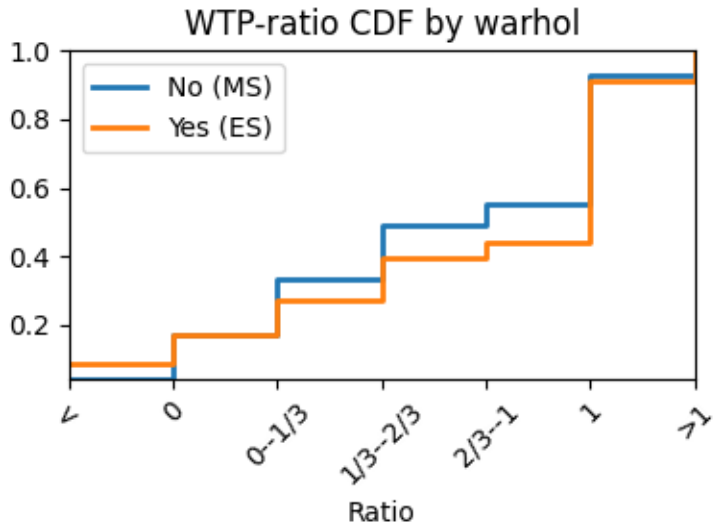


Figure A.6: Caption

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
warhol_Yes	0.16%	0.0%	0.62%	0.93%	17.13%	2.96%	6.54%	6.7%	21.18%	37.07%	6.7%
warhol_No	0.6%	0.12%	0.48%	0.12%	13.4%	2.87%	3.59%	10.29%	31.46%	31.34%	5.74%
arkansas_Better off	0.18%	0.0%	0.73%	0.73%	14.99%	3.11%	6.4%	8.59%	26.33%	32.72%	j6.22%
arkansas_Worse off	0.54%	0.11%	0.43%	0.32%	15.04%	2.79%	3.97%	8.81%	27.39%	34.48%	6.12%
experience_Yes	0.26%	0.0%	0.66%	0.53%	16.05%	3.03%	5.26%	9.61%	26.97%	31.84%	5.79%'
experience_No	0.56%	0.14%	0.42%	0.42%	13.93%	2.79%	4.46%	7.8%	27.02%	35.93%	6.55%
'all$_{questions_ES}$'	'0.88%	0.0%	1.75%	1.75%	14.91%	4.39%	5.26%	5.26%	25.44%	35.96%	4.39%
'all$_{questions_MS}$'	'0.72%	0.0%	0.36%	0.0%	13.04%	3.26%	2.9%	11.23%	32.97%	32.25%	3.26%

Table A.2: Data are percentages (fractions of the total. Letter values must sum up to 100%
Note: This table considers the whole sample. WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
'warhol_Yes'	0.0%	0.0%	0.0%	0.94%	15.49%	1.88%	3.76%	7.98%	26.29%	36.62%	7.04%
'warhol_No'	0.36%	0.36%	0.36%	0.0%	13.57%	3.93%	3.57%	11.79%	31.07%	31.43%	3.57%
'arkansas_Better off'	0.0%	0.0%	0.0%	0.54%	13.98%	2.15%	4.3%	9.14%	31.18%	33.87%	4.84%
'arkansas_Worse off'	0.33%	0.33%	0.33%	0.33%	14.66%	3.58%	3.26%	10.75%	27.69%	33.55%	5.21%
'experience_Yes'	0.42%	0.0%	0.42%	0.42%	13.33%	3.33%	4.17%	12.08%	27.5%	34.58%	3.75%
'experience_No'	0.0%	0.4%	0.0%	0.4%	15.42%	2.77%	3.16%	8.3%	30.43%	32.81%	6.32%
'all_questions_ES'	'0.0%	0.0%	0.0%	2.27%	15.91%	0.0%	2.27%	2.27%	31.82%	40.91%	4.55%
'all_questions_MS'	'1.08%	0.0%	1.08%	0.0%	10.75%	4.3%	3.23%	11.83%	27.96%	38.71%	1.08%

Table A.3: Data are percentages (fractions of the total. Letter values must sum up to 100%
Note: This table considers the sample in the baseline treatment only. WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
'warhol$_{Yes}$'	0.19%	0.0%	0.0%	0.39%	12.09%	2.14%	5.65%	7.6%	23.59%	42.5%	5.85%
'warhol$_{No}$'	0.0%	0.0%	0.15%	0.0%	10.36%	0.73%	2.92%	11.09%	35.91%	33.87%	4.96%
'arkansas$_{Betteroff}$'	0.23%	0.0%	0.23%	0.0%	10.59%	1.58%	5.63%	9.23%	30.18%	36.26%	6.08%
'arkansas$_{Worseoff}$'	0.0%	0.0%	0.0%	0.27%	11.41%	1.19%	3.18%	9.81%	30.9%	38.33%	4.91%
'experience$_{Yes}$'	0.0%	0.0%	0.17%	0.33%	11.3%	1.5%	4.32%	10.63%	30.56%	36.05%	5.15%
'experience$_{No}$'	0.17%	0.0%	0.0%	0.0%	10.91%	1.17%	3.86%	8.56%	30.7%	39.09%	5.54%
'all$_{questions_ES}$'	1.05%	0.0%	0.0%	0.0%	13.68%	2.11%	5.26%	5.26%	26.32%	42.11%	4.21%
'all$_{questions_MS}$'	0.0%	0.0%	0.0%	0.0%	10.09%	0.88%	2.19%	12.28%	35.96%	35.53%	3.07%

Table A.4: Data are percentages (fractions of the total. Letter values must sum up to 100%
Note: This table considers the high-quality sample for all treatments. WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

	$WTP<0$			$WTP=0$			$WTP>0$				
	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES>0$	$ES<0$	$ES=0$	$ES<WTP$	$ES=WTP$	$ES>WTP$
'warhol$_{Yes}$'	0.0%	0.0%	0.0%	0.58%	9.25%	0.58%	2.31%	8.67%	30.06%	41.62%	6.94%
'warhol$_{No}$'	0.0%	0.0%	0.0%	0.0%	10.73%	1.29%	2.58%	12.45%	36.05%	33.48%	3.43%
'arkansas$_{Betteroff}$'	0.0%	0.0%	0.0%	0.0%	9.49%	0.63%	3.8%	8.86%	36.08%	36.08%	5.06%
'arkansas$_{Worseoff}$'	0.0%	0.0%	0.0%	0.4%	10.48%	1.21%	1.61%	12.1%	31.85%	37.5%	4.84%
'experience$_{Yes}$'	0.0%	0.0%	0.0%	0.51%	8.72%	1.03%	3.08%	12.82%	31.28%	38.46%	4.1%
'experience$_{No}$'	0.0%	0.0%	0.0%	0.0%	11.37%	0.95%	1.9%	9.0%	35.55%	35.55%	5.69%
'all$_{questions_E}$'	0.0%	0.0%	0.0%	0.0%	12.82%	0.0%	2.56%	2.56%	33.33%	46.15%	2.56%
'all$_{questions_M}$'	0.0%	0.0%	0.0%	0.0%	9.88%	1.23%	2.47%	13.58%	29.63%	41.98%	1.23%

Table A.5: Data are percentages (fractions of the total. Letter values must sum up to 100%
Note: This table considers the high-quality sample in the baseline treatment only. WTP is the willingness-to-pay to give Receiver the books with original notes in *Learns* when he learns about it. ES is the willingness-to-pay to give Receiver the books with original notes in *NotLearns* when he does not learn about it.

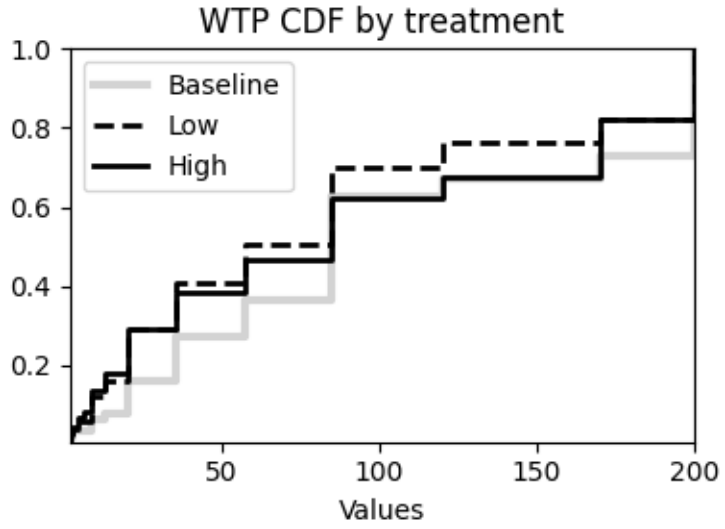


Figure A.7: Trad WTP for Mixed

We are giving a present to someone!

The questions we will ask you to answer involve another person. His name is Alex.

Alex loves economics, and we are going to give him a present!

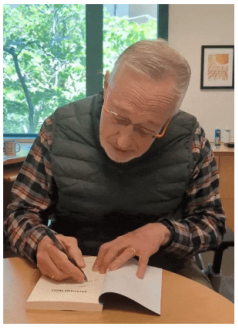
He is going to get two books by two Nobel laureates in economics—Professors Paul Milgrom and Alvin Roth (you can click each of their names to open their Wikipedia page). Professors Milgrom and Roth are professors at our university and have agreed to help with the study.

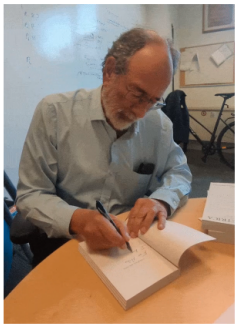
Alex has already read some of their work and told us he has great admiration for them.

The two books come with handwritten notes. But here is the twist! **We have two copies of each of these books.** One with **original notes** from the famous authors themselves (Profs. Milgrom and Roth), and one with **fake notes** written by someone excellent at copying their handwriting.

Here are the two books

Here are videos of the professors writing the notes





The fake versions of the handwritten notes are indistinguishable from the original ones. Professors Milgrom and Roth themselves could not tell which is which!

Alex will receive two books, either the two with the original handwritten notes or the two with the fake ones. We will return the two books that we do not give to Alex back to Professors Milgrom and Roth.

When you are ready, click "Next."

Next

Your task

In addition to the books, Alex may also receive a **surprise bonus**.

Your task in this study is to make choices that we will use to determine:

- which two books (original or fake) Alex receives,
- his surprise bonus.

In particular, we will randomly choose one participant like yourself and actually implement what they chose for Alex.

There is one more important detail: **We might or might not tell Alex whether the two books he gets are the ones with the original or fake notes.**

Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

We ask you to make your choices for each case. We will then choose a case randomly and implement that case.

Alex knows that if we don't tell him which books he got, he may have got the ones with the original or the fake notes.

When ready, click "Next."

Next

Baseline treatment

Your task

In addition to the books, Alex may also receive a **surprise bonus**.

Your task in this study is to make choices that we will use to determine:

- which two books (original or fake) Alex receives,
- his surprise bonus.

In particular, we will randomly choose one participant like yourself and actually implement what they chose for Alex.

There is one more important detail: **We might or might not tell Alex whether the two books he gets are the ones with the original or fake notes.**

Case 1: we WILL tell Alex whether the books he got are the ones with the original or fake notes

With a 75% chance, Alex will get the books with the fake notes, and with a 25% chance, **you will determine which books he gets.**

Case 2: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

With a 75% chance, Alex will get the books with the fake notes, and with a 25% chance, **you will determine which books he gets.**

We ask you to make your choices for each case. We will then choose a case randomly and implement that case.

Alex knows that if we don't tell him which books he got, there is at least a 75% chance that they are the ones with the fake notes.

When ready, click "Next."

Next

LowMS treatment

Your task

In addition to the books, Alex may also receive a **surprise bonus**.

Your task in this study is to make choices that we will use to determine:

- which two books (original or fake) Alex receives,
- his surprise bonus.

In particular, we will randomly choose one participant like yourself and actually implement what they chose for Alex.

There is one more important detail: **We might or might not tell Alex whether the two books he gets are the ones with the original or fake notes.**

Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

With a 75% chance, Alex will get the books with the original notes, and with a 25% chance, **you will determine which books he gets.**

Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

With a 75% chance, Alex will get the books with the original notes, and with a 25% chance, **you will determine which books he gets.**

We ask you to make your choices for each case. We will then choose a case randomly and implement that case.

Alex knows that if we don't tell him which books he got, there is at least a 75% chance that they are the ones with the original notes.

When ready, click "Next."

Next

HighMS treatment

Well done!

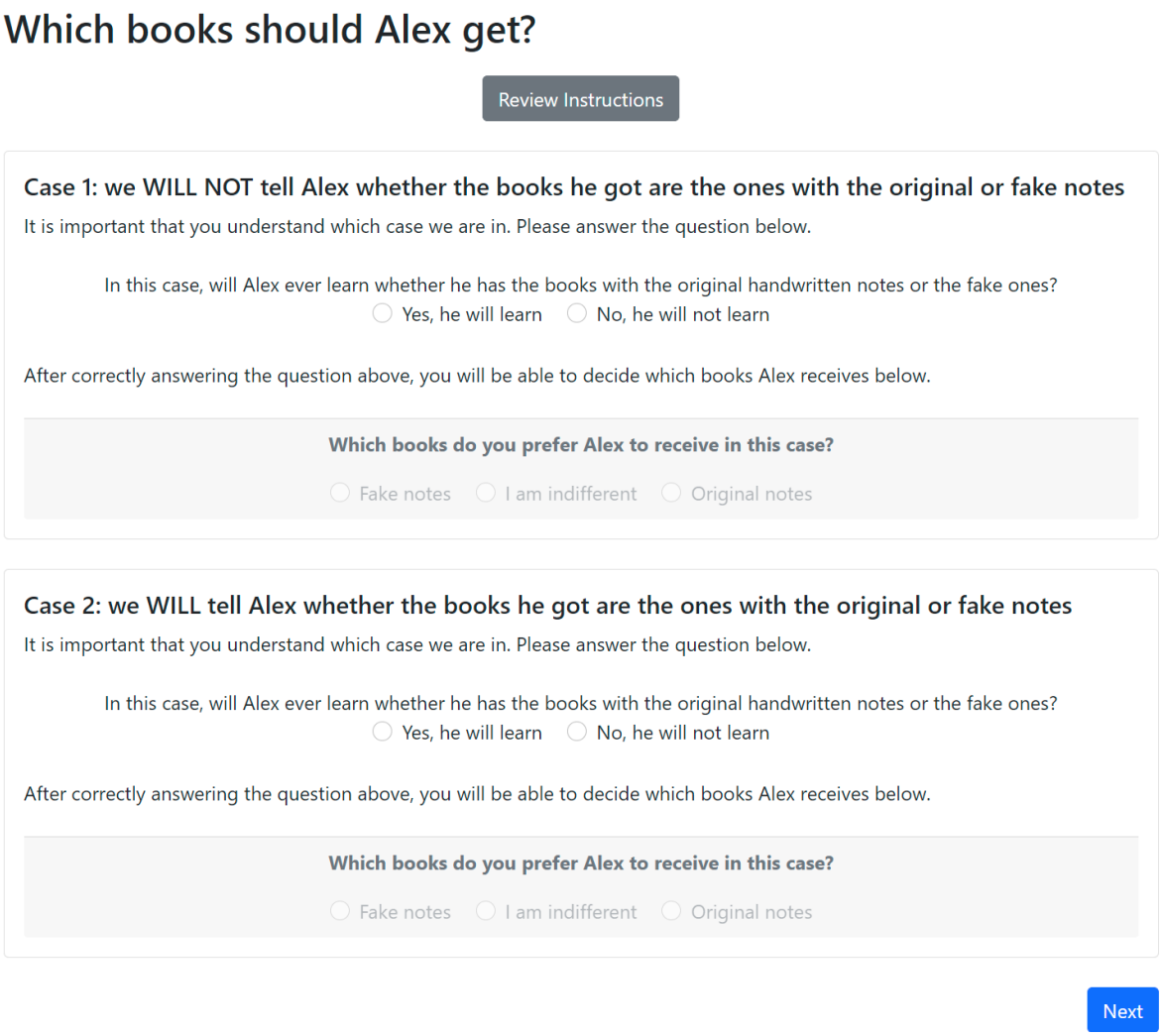
On the next pages, we will ask you 15 questions to determine which books Alex gets.

For example, one of them will be: *which books do you prefer Alex to receive? The ones with the original or fake notes?* There are other questions where we add a bonus for Alex to one of the options. You can click on this button to see all questions: Questions

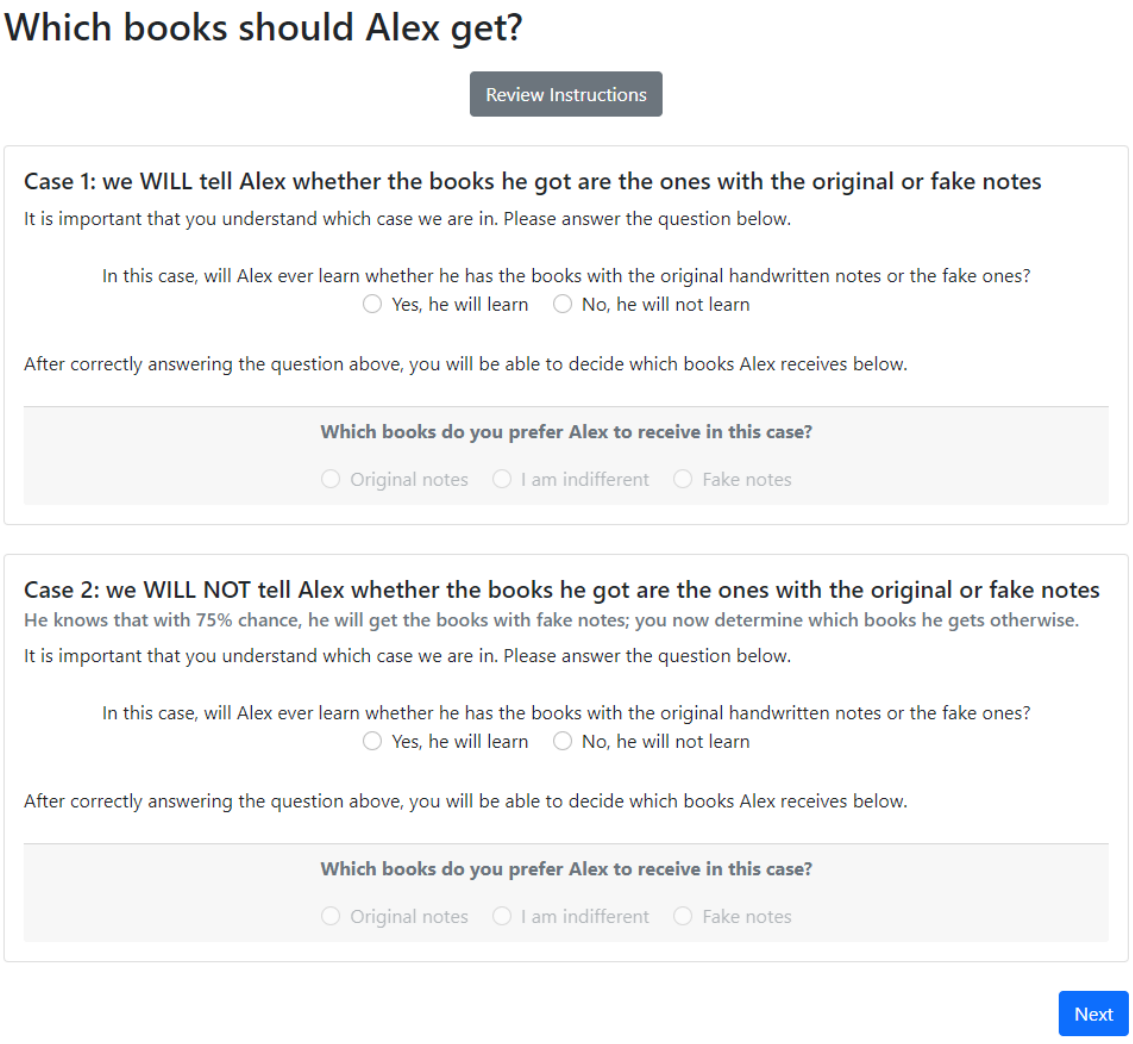
We will randomly pick one of the questions and implement whatever option you choose. This means that any question can be the one that determines what Alex gets, so please answer them carefully.

When you are ready, click "Next."

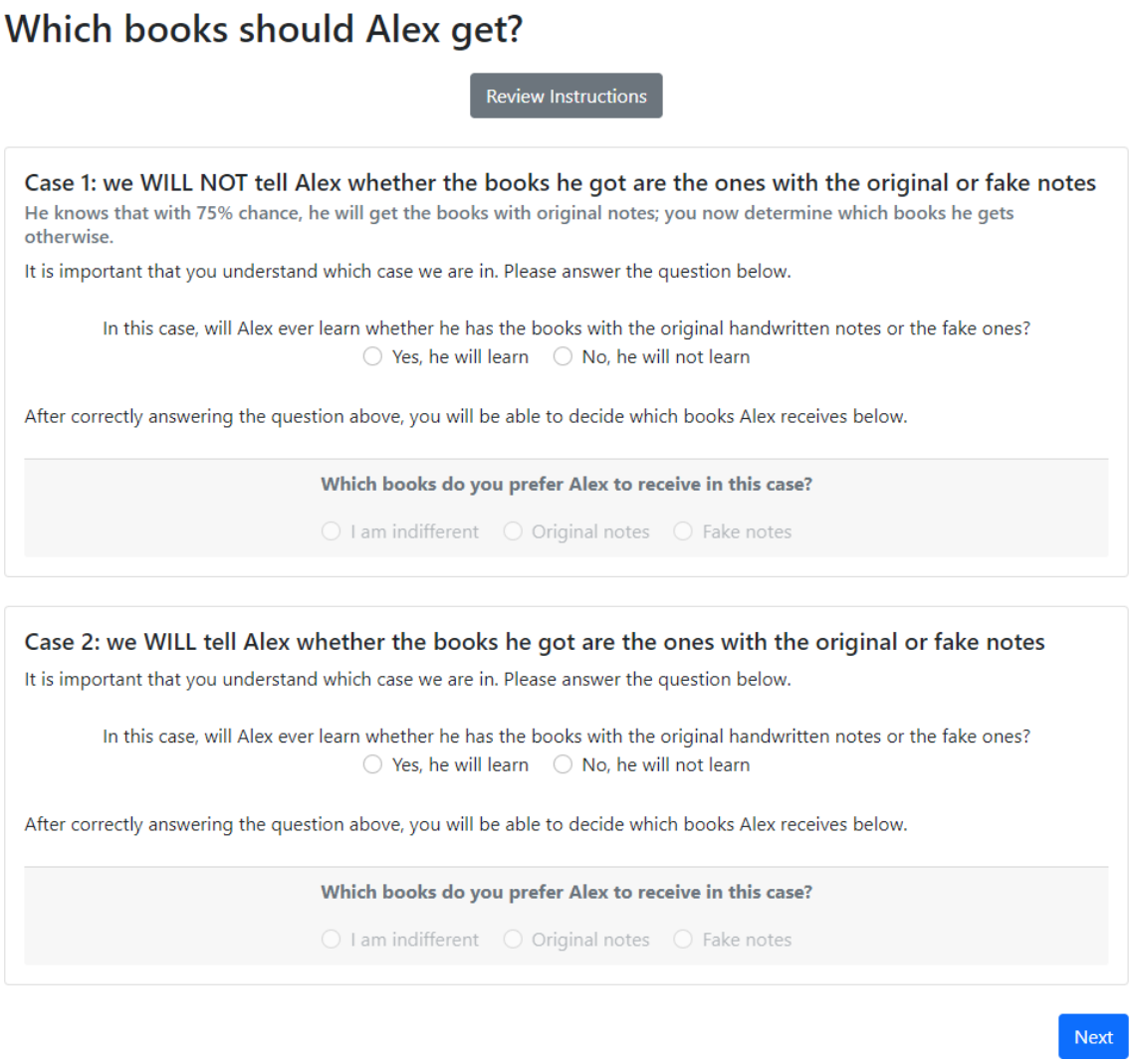
Next



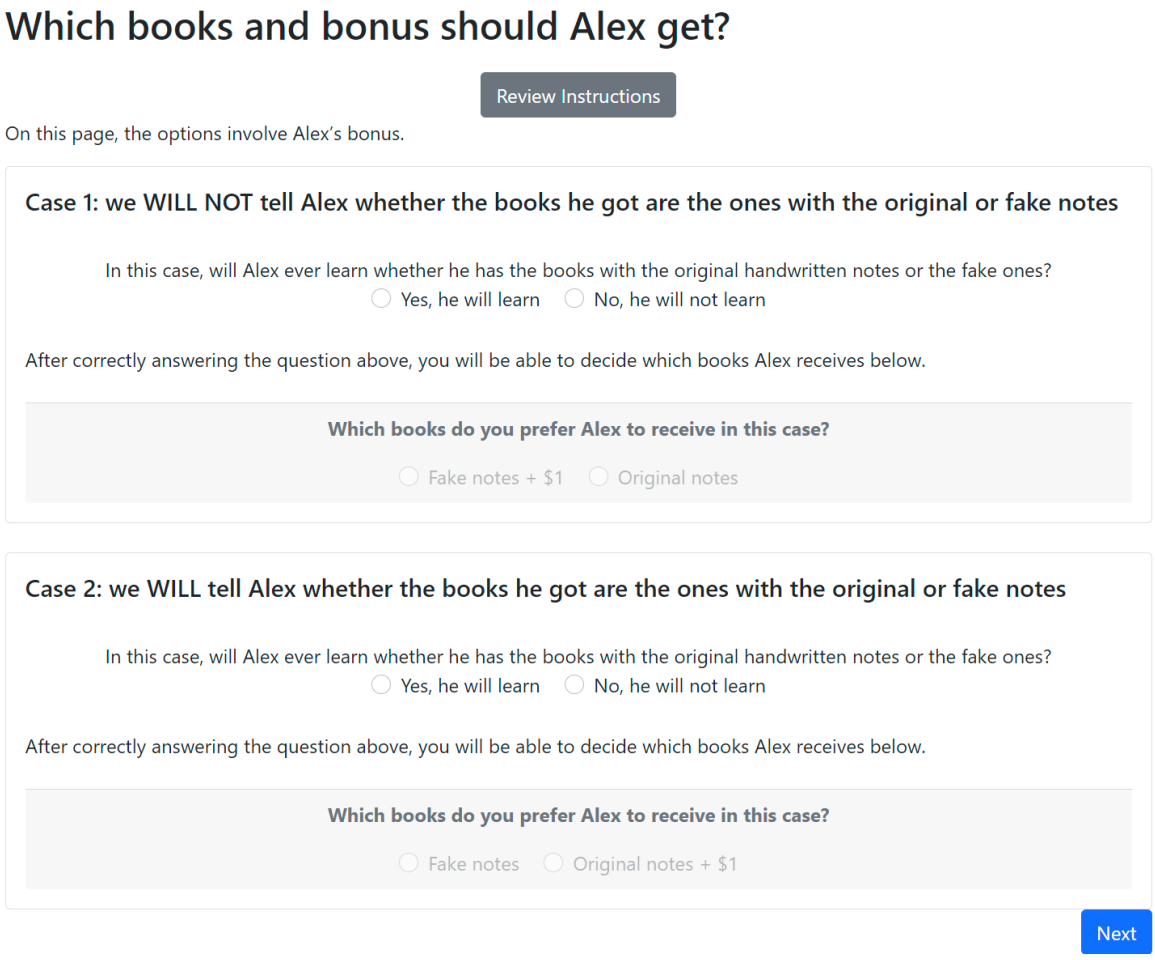
Baseline treatment



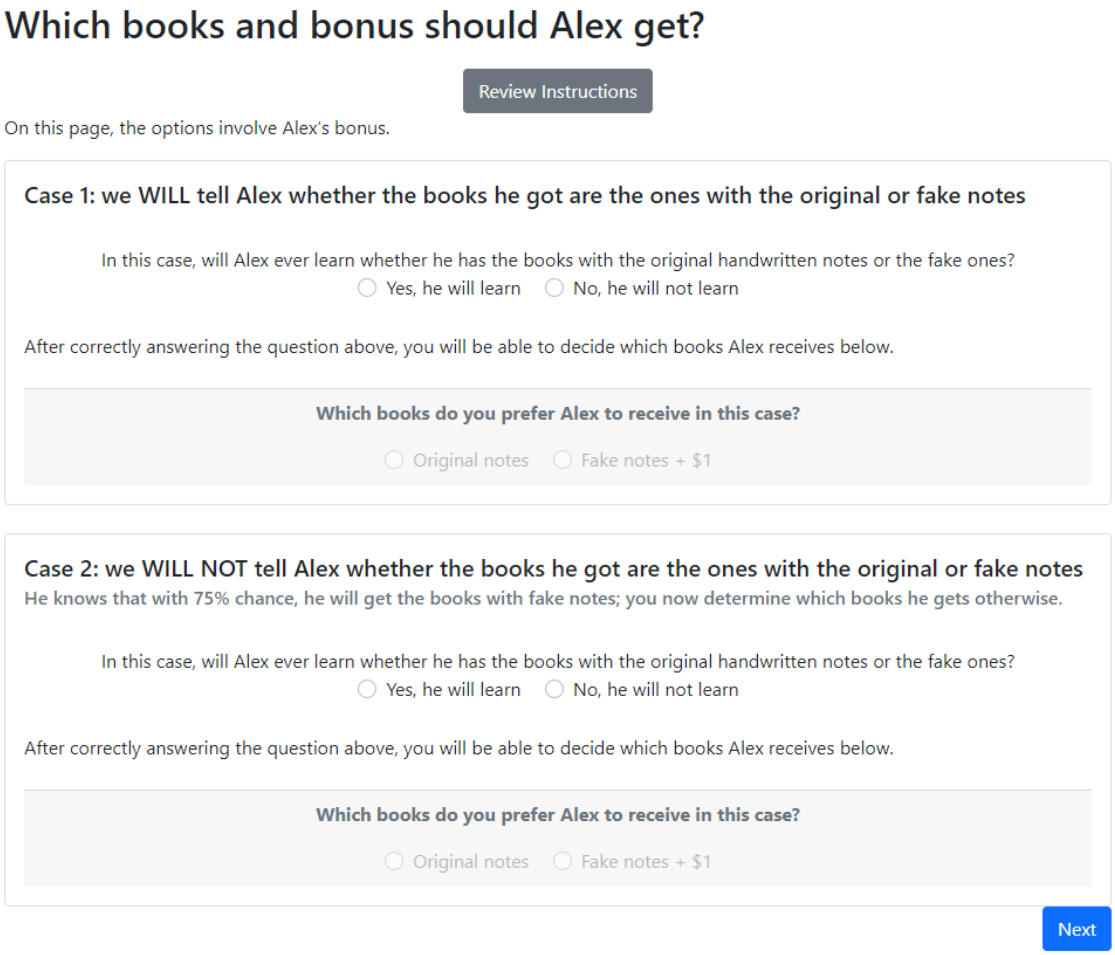
LowMS treatment



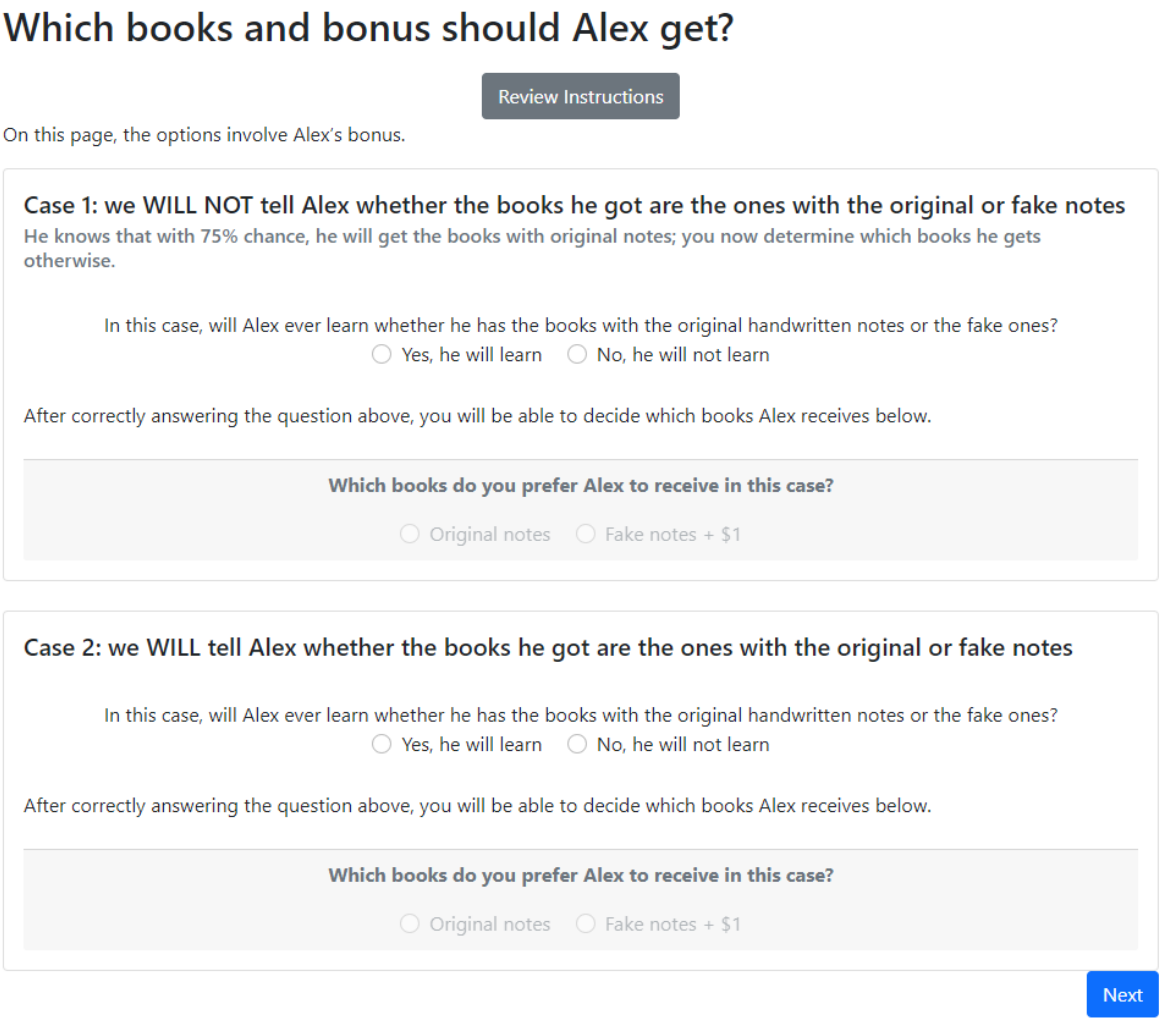
HighMS treatment



Baseline treatment



LowMS treatment



HighMS treatment

Which books and bonus should Alex get?

You make several choices again involving Alex's bonus, filling in a table like the one below. For each row, we ask you to choose between the original notes and fake notes with a bonus for Alex.

Fake notes and...		Original notes and...
...$2	OR	...$0
...$3	OR	...$0
...$5	OR	...$0
...$7	OR	...$0
...$10	OR	...$0
...$15	OR	...$0
...$25	OR	...$0
...$45	OR	...$0
...$70	OR	...$0
...$100	OR	...$0
...$140	OR	...$0
...$200	OR	...$0

We assume that once you choose the fake notes for one row, you will choose the fake notes for all rows below because the rows below simply make the fake notes better by increasing the bonus. You will only need to choose the row in which you switch from preferring the original notes to preferring the fake notes. You do that by clicking on the row.

Once you are confident that you understand how the table works, continue to give your answers.

Next

Which books and bonus should Alex get?

Review Instructions

Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Fake notes and...		Original notes and...
...$2	OR	...$0
...$3	OR	...$0
...$5	OR	...$0
...$7	OR	...$0
...$10	OR	...$0
...$15	OR	...$0
...$25	OR	...$0
...$45	OR	...$0
...$70	OR	...$0
...$100	OR	...$0
...$140	OR	...$0
...$200	OR	...$0

Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Fake notes and...		Original notes and...
...$2	OR	...$0
...$3	OR	...$0
...$5	OR	...$0
...$7	OR	...$0
...$10	OR	...$0
...$15	OR	...$0
...$25	OR	...$0
...$45	OR	...$0
...$70	OR	...$0
...$100	OR	...$0
...$140	OR	...$0
...$200	OR	...$0

Which books and bonus should Alex get?

Review Instructions

Case 1: we WILL tell Alex whether the books he got are the ones with the original or fake notes

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Original notes and...		Fake notes and...
...$0	OR	...$2
...$0	OR	...$3
...$0	OR	...$5
...$0	OR	...$7
...$0	OR	...$10
...$0	OR	...$15
...$0	OR	...$25
...$0	OR	...$45
...$0	OR	...$70
...$0	OR	...$100
...$0	OR	...$140
...$0	OR	...$200

Case 2: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

He knows that with 75% chance, he will get the books with fake notes; you now determine which books he gets otherwise.

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Original notes and...		Fake notes and...
...$0	OR	...$2
...$0	OR	...$3
...$0	OR	...$5
...$0	OR	...$7
...$0	OR	...$10
...$0	OR	...$15
...$0	OR	...$25
...$0	OR	...$45
...$0	OR	...$70
...$0	OR	...$100
...$0	OR	...$140
...$0	OR	...$200

Next

Which books and bonus should Alex get?

Review Instructions

Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

He knows that with 75% chance, he will get the books with original notes; you now determine which books he gets otherwise.

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Original notes and...		Fake notes and...
...$0	OR	...$2
...$0	OR	...$3
...$0	OR	...$5
...$0	OR	...$7
...$0	OR	...$10
...$0	OR	...$15
...$0	OR	...$25
...$0	OR	...$45
...$0	OR	...$70
...$0	OR	...$100
...$0	OR	...$140
...$0	OR	...$200

Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

In this case, will Alex ever learn whether he has the books with the original handwritten notes or the fake ones?

○ Yes, he will learn ○ No, he will not learn

After correctly answering the question above, please tell us which option you prefer by clicking the table below.

Original notes and...		Fake notes and...
...$0	OR	...$2
...$0	OR	...$3
...$0	OR	...$5
...$0	OR	...$7
...$0	OR	...$10
...$0	OR	...$15
...$0	OR	...$25
...$0	OR	...$45
...$0	OR	...$70
...$0	OR	...$100
...$0	OR	...$140
...$0	OR	...$200

Please review your responses for the two cases

Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes

1. I prefer Alex to receive the ones with the original notes and $1 over the ones with the fake notes
2. I prefer Alex to receive the ones with the original notes over the ones with the fake notes
3. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $1
4. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $2
5. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $3
6. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $5
7. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $7
8. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $10
9. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $15
10. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $25
11. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $45
12. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $70
13. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $100
14. I prefer Alex to receive the ones with the fake notes and $140 over the ones with the original notes
15. I prefer Alex to receive the ones with the fake notes and $200 over the ones with the original notes

Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

1. I prefer Alex to receive the ones with the original notes and $1 over the ones with the fake notes
2. I prefer Alex to receive the ones with the original notes over the ones with the fake notes
3. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $1
4. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $2
5. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $3
6. I prefer Alex to receive the ones with the original notes over the ones with the fake notes and $5
7. I prefer Alex to receive the ones with the fake notes and $7 over the ones with the original notes
8. I prefer Alex to receive the ones with the fake notes and $10 over the ones with the original notes
9. I prefer Alex to receive the ones with the fake notes and $15 over the ones with the original notes
10. I prefer Alex to receive the ones with the fake notes and $25 over the ones with the original notes
11. I prefer Alex to receive the ones with the fake notes and $45 over the ones with the original notes
12. I prefer Alex to receive the ones with the fake notes and $70 over the ones with the original notes
13. I prefer Alex to receive the ones with the fake notes and $100 over the ones with the original notes
14. I prefer Alex to receive the ones with the fake notes and $140 over the ones with the original notes
15. I prefer Alex to receive the ones with the fake notes and $200 over the ones with the original notes

Do the above answers reflect what you intended to answer, or do you want to give your answers again?

- ○ Yes, the answers above reflect what I intended to answer
- ○ No, I want to give my answers again

Please click "Next" to proceed to the next page.

Next

Thank you for your responses

Reminder:

- Case 1: we WILL NOT tell Alex whether the books he got are the ones with the original or fake notes
- Case 2: we WILL tell Alex whether the books he got are the ones with the original or fake notes

You gave different responses in Case 1 and Case 2. Why? Please tell us in approximately 1-3 sentences.
(There is nothing wrong with your answers! We are just interested in your reasoning)

In the remaining pages, we will present you with scenarios and ask you questions about them. Your responses are very important for our study; please think about the scenarios and answer carefully.

Next

The Experience Machine

If given the option, would you choose to plug into an experience machine that could provide you with an entirely immersive, simulated reality where you can experience any desirable scenario, despite not being real? Keep in mind that while plugged in, you would never be aware that you are in the experience machine and would believe that the simulated reality is real.

Suppose there was an experience machine that would give you *any* experience you desired (eating good food, having a successful career, making meaningful connections, etc.). While in the machine, you would not know that you are in it; you would think that what you are experiencing is actually happening.

Would you go into the machine?

- ○ Yes
- ○ No

Why? Answer in approximately 1-2 sentences.

Next

Flood in Arkansas

A small town in Arkansas experiences massive flooding, leaving many families homeless. To provide financial relief to the impacted families, the government temporarily increases taxes, including a $100 levy on John. John lives far away and *will never learn about the flooding or the relief effort*. However, he cares about helping others and would gladly contribute $100 to the relief effort if he knew about the flood.

Please tell us what you think using the information provided above. This is not a trick question; we want to understand what you think about the impact that the policy has on John.

Does the government raising taxes to provide financial relief make John better or worse off?

- ○ Better off
- ○ Worse off

Why? Answer in approximately 1-2 sentences.

Next

Andy Warhol

Hundreds of Andy Warhol fakes, and one original drawing worth $20k, sold for $250 each. An art collective purchased an original Warhol drawing and copied it 999 times. The copies are carefully created so that not even their creators can tell them apart from the original drawing. They then mixed the original together with the copies and sold the 1000 drawings.

Please watch the video (1min 1sec) the art collective made (audio is not needed). You can read more about this story here.

Someone got the original Andy Warhol drawing. Since the original drawing and the copies are indistinguishable, please assume that neither the person who got it nor anyone else will ever know which is the original drawing or who has it.

Is this person better off by getting the original one instead of a copy?

- ○ Yes
- ○ No

Why? Answer in approximately 1-2 sentences.

Next