Comparisons*

Thomas Graeber ⓡ Benjamin Enke

April 7, 2026

Abstract

Economic decisions across households, experts and markets often depend on comparison points such as reference points, anchors, goals, expectations and peers. A long-standing puzzle is that decisions sometimes *increase* in these comparators (assimilation) and sometimes *decrease* in them (contrast). We develop a simple taxonomy that predicts the sign. The idea is that when people are uncertain how to map decision inputs into outputs, comparison points are used as information. This logic predicts that decisions decrease in *input comparators*—those tied to exogenous decision inputs (e.g., a typical hourly wage)—and increase in *output comparators*—those tied to endogenous decisions or outcomes (e.g., typical daily earnings). We test and confirm these predictions in experiments on labor supply, investment and belief updating, involving both social and expectations-based comparison points. We further show that comparison effects weaken both when decision uncertainty is reduced and when comparators are less informative, consistent with a mechanism of comparison points as information. Finally, a systematic classification of the prior literature shows that the input-vs.-output taxonomy robustly predicts the sign of comparison effects across more than 100 experimental and field observational studies.

*We thank Dustin Fichmann and Felix Mao for outstanding research assistance. We thank Nick Netzer, Michael Woodford as well as seminar audiences at the Berlin Behavioral Seminar, Berkeley, Harvard, Wharton, Zurich and CESifo Munich for useful feedback and discussions. This research program and related results were made possible by the support of the NOMIS Foundation. Graeber: University of Zurich, Department of Economics, thomas.graeber@econ.uzh.ch; Enke: Harvard University, Department of Economics, and NBER, enke@fas.harvard.edu.

1 Introduction

A large body of empirical work in economics and other fields shows that decisions are often influenced by quantities such as reference points, goals, targets or anchors: quantities that we refer to as *comparison points*. What is the main comparative statics effect of comparison points for decisions? In practical applications, comparison effects can take either of two diametrically-opposed forms. Consider labor supply. First, people work less when the *piece rate* they expected or that a peer receives is higher (e.g., Gächter and Thöni, 2010; Cohn et al., 2014; Breza et al., 2018). Second, however, people work more when their expected or typical *earnings* are higher (e.g., Camerer et al., 1997; Abeler et al., 2011). Thus, a higher comparison piece rate decreases effort, but higher comparison earnings increase it. This is all the more puzzling because earnings and piece rates will often be positively correlated.

There are many more examples of such comparison effects of different signs in the literature, in settings ranging from finance to labor to consumption. For instance, asking prices for a house increase in its purchase price (Genesove and Mayer, 2001), movers arriving from more expensive cities rent pricier apartments than those arriving from cheaper cities (Simonsohn and Loewenstein, 2006), more ambitious goals in sports and at work are associated with better performance (Pope and Schweitzer, 2011; Clark et al., 2020), reported tax liabilities increase in the amount withheld by the IRS (Rees-Jones, 2018) and the probability of choosing an item increases in whether one was endowed with it (Loewenstein, 1988; Knetsch, 1989; Kahneman et al., 1990). On the other hand, candidate ratings decrease in the quality of the prior candidate (Radbruch and Schiprowski, 2025; Bhargava and Fisman, 2014), effort supply for a task of known ease decreases in ex-ante expectations about task ease (Bushong and Gagnon-Bartsch, 2023), and valuations for gambles decrease in the attractiveness of previously encountered gambles (Meyer and Hundtofte, 2023). These effects are sometimes referred to as "anchoring" or "assimilation" when the decision increases in the comparison point, and as "contrast" when the decision decreases in it.[1] In which situations, then, should we expect which effect, and what does this tell us about the mechanism through which comparison points affect behavior?

In this paper, we explore the hypothesis that the sign of comparison effects can be predicted based on the idea that people may be uncertain how to translate the features (inputs) of a problem into a decision (output), and partly resort to comparison points in gathering information about what a good decision may be. From an informational perspective, it is useful to distinguish between comparison points that relate to a decision input and those that relate to a decision output. Decision inputs are exogenous fundamentals that the decision maker takes as given for the decision at hand, such as an hourly wage in the labor supply example. Thus,

[1]This is when normalizing the effect of the main decision input to be positive; see Section 2.

a previous wage or another person's wage both represent *input comparators*. Decision outputs are endogenous decision variables controlled by the decision-maker, e.g., the number of hours worked or the resulting total earnings. Hence, the number of hours worked on a previous day or typical past earnings represent *output comparators*.

To illustrate why input and output comparators push in different directions, suppose I'm asked to decide how many hours to work at wage w. Further suppose for simplicity that I know that my utility-maximizing decision is given by $a^*(w) = \theta w$, yet I don't know θ, for example because I'm uncertain how much I dislike the task. Now suppose a friend with similar preferences decided to work for a_c hours at wage w_c. Then, a_c/w_c is a signal about my per-unit willingness-to-work θ. Thus, in the presence of uncertainty about the input-decision map, my labor supply *should* increase in my friend's labor supply or in his accumulated earnings (the output comparators) but decrease in their wage (the input comparator). In the terminology used in the literature, my labor supply should assimilate towards output comparators and should exhibit contrast to input comparators. Intuitively, output comparators provide information about a reasonable 'ballpark' of the decision (producing assimilation), while the input comparator provides information about the direction in which one needs to adjust (producing contrast).

Below, we show that this prediction does not depend on the 'ratio' formulation of input and output comparator but holds more generally. We thus view our empirical investigation not as a test of a specific model but of a broad class of models that feature learning from comparison points. Our hypothesis also does not require that comparison points are objectively informative or perceived as such; while in some contexts they may be, in others people may heuristically act as if they are informative, perhaps because people have become adapted to an environment in which comparison points are typically informative.

Our account has several implications that allow for sharp tests. First, that input and output comparators push in opposite directions. Second, this prediction applies to different decision domains, including both preferential choice contexts and problems with objectively correct solutions—whenever people don't perfectly know how to map circumstances into decisions. Third, it also applies to different types of comparisons, such as social comparisons and expectations-based comparison points. Fourth, the same comparison variable can produce both contrast and assimilation, purely depending on whether the variable is an input or output comparator for the decision at hand. Fifth, both types of comparison effects should weaken when there is less uncertainty about how to translate problem fundamentals into a decision. Sixth, comparison effects should weaken when the comparison point is transparently less informative.

Testing these hypotheses requires controlled experiments because, in field applications, the analyst often does not observe both types of comparison points. We thus implement a suite of controlled online experiments on choice, valuation and objective inference problems.

Baseline evidence. Our experimental strategy is to separately induce and vary input and output comparators. In our *Effort Valuation* paradigm, subjects are asked to state their minimum required compensation (WTA) to complete a specified workload of a standard real effort task. In this paradigm, the offered workload (i.e., number of tasks) is the input and the minimum compensation the output. In the *Social CP* condition, participants make their decisions while shown the workload offered to another participant along with that participant's decision. These correspond to social variants of input and output comparators, respectively. In the *Expectations CP* condition, the comparison point is instead induced via expectations using standard techniques (e.g., Abeler et al., 2011): with a 50% chance, the subject's decision does not count, and their workload and compensation are instead determined by an exogenously-imposed workload and compensation.

In both treatments, we find strong effects of the comparison points in the predicted directions. Holding fixed the objective workload in a problem, subjects' WTA increases in the output comparator (either other participants' WTA or the comparison earnings induced via expectations). On the other hand, subjects' WTA decreases in the input comparator (either the workload of other participants or the comparison workload induced via expectations).

To test whether the effects of input and output comparators also arise when only one of them is present, we run two additional treatments – *Input CP* and *Output CP* – and find that both effects persist also in the absence of the other comparison point.

To document that the logic of input and output comparators also holds in other domains, including decision problems with objectively correct solutions, we run two additional sets of experiments. In the *Investment* paradigm, subjects allocate a budget between a safe option and a risky gamble. The gamble's winning chance—which varies across rounds—is the decision input, and the amount wagered is the output. As in the *Effort Valuation* paradigm, we induce expectations-based comparison points: with 50% chance, the subject's own decision counts, and with 50% chance the experimenter instead invests an imposed amount into an imposed gamble with a different winning chance. The winning chance of the imposed gamble is the input comparator and the imposed investment amount the output comparator. We find that subjects invest a substantially lower amount when the comparison gamble has a higher winning chance (contrast to input comparator) but invest more when the comparison investment is higher (assimilation to output comparator).

In the *Beliefs* paradigm, we study a canonical Bayesian test-and-disease problem in which subjects guess the likelihood that a hypothetical patient has a medical condition, given a prior and a diagnostic test result. Here, we induce social comparison points: subjects observe the updating problem faced by another participant—including that participant's prior and test result (the input comparators)—along with that participant's stated posterior (the output comparator). We find that stated posteriors strongly decrease in both input comparators and strongly increase in the output comparator. In this design, preferences over comparisons (such as reference-dependent utility or social emotions like envy) cannot plausibly explain either direction of comparison effect. Taken together, the results across our effort valuation, risky investment, and belief updating paradigms suggest that the input-output comparator distinction applies to different types of decision problems and predicts the sign of comparison effects across preferential choice and objective inference.

Turning output comparator into input comparator. A key implication of our account is that the same comparison variable can predictably have opposite effects, purely depending on whether it constitutes an input or output comparator for the current decision.

In our *Effort Valuation* paradigm, subjects translated a total workload into a monetary valuation. In this setting, a comparison workload is an input comparator. We contrast this setting with the *Effort Choice* paradigm, in which subjects decide their total workload as a function of an exogenous piece rate. Here, a comparison workload is an output comparator. Our account thus predicts that the sign of the effect of the comparison workload flips across the two settings.

The evidence provides support for this idea. An exogenously higher comparison workload (induced via expectations) leads to considerably higher chosen workloads, in stark contrast to the *Effort Valuation* paradigm, in which it produces lower minimum required compensations.

Mechanism experiments: comparisons as information. If comparison points are indeed (implicitly) used as sources of information, then comparison effects should be more pronounced when there is less uncertainty about how to translate decision inputs into a decision.

To test this idea, we again return to the *Effort Valuation* paradigm and exogenously reduce participants' uncertainty about the input-decision map. To do so, we increase subjects' deliberation about the input-decision map by asking them to state a full 'policy function' prior to making their incentivized choices, i.e., to state their WTA for a large (and ordered) set of workloads. This treatment reduces the reliance on both input and output comparators by more than 40%, suggesting that a substantial share of comparison effects in this setting reflects decision uncertainty rather than stable preferences over comparisons.

Finally, we also implement a treatment in which we transparently reduce the informativeness of the comparison points. In this experiment, the comparison point is transparently and saliently randomly generated and meaningless. This substantially reduces the magnitude of both comparison effects (but does not eliminate them), again consistent with a mechanism of comparison points used as (implicit) sources of information.

Classifying results of prior studies. To investigate the broader validity of our input-vs.-output-comparator taxonomy, we revisit a large number of prior empirical studies on comparison points.

We assemble more than 100 experimental and observational studies on comparison effects from ten top journals in economics going back to 1970. We classify each paper according to whether it studies an input or output comparator, and whether the results are consistent with our taxonomy. This classification is subject to a number of caveats discussed in detail in Section 7. For instance, we typically observe only either the input or the output comparator, and our classification is inherently partly subjective.

We find that the results of the vast majority of papers are consistent with the predictions. For instance, returning to some of our opening examples, consider classic contrast effects. In deciding how many hours to work, expected wages or other people's wages are an input comparator for actual wages, and consistently have negative effects on effort supply. In a sequential interviewing context, the quality of the previous candidate is an input comparator for the quality of the current candidate, and has a negative effect on the evaluation of the current candidate. Similarly, in deciding whether to buy a concert ticket, the past price is an input comparator for the current price, and also produces a contrast effect.

Classic assimilation effects, on the other hand, are almost always found when the comparison point lives in the output domain, i.e., when it constitutes a decision or a resulting outcome. In setting an asking price for a house, the purchase price is an output comparator. In sports and at work, performance goals such as par in golf are output comparators for actual performance. Similarly, possessing a good at the beginning of an experiment constitutes an output comparator for owning it at the end. In all of these contexts, decisions increase in the comparison point, consistent with our taxonomy of output comparators producing assimilation effects. Of the 102 papers we are able to classify, 101 produce results consistent with our taxonomy.

Forecaster survey. To get a sense of whether the input-output taxonomy reflects knowledge already held by the profession, we implemented a forecaster survey on the *Social Science Prediction Platform* with 105 social science researchers. When first asked whether they have a general rule for predicting the sign of comparison effects, 24.8% of forecasters answered yes—but a manual review of these rules revealed that none described anything resembling the input-output distinction; the most common responses appealed to loss aversion, prospect theory, or domain-specific intuitions. We then asked forecasters to predict the signs of input and output comparator effects across four of our experimental paradigms. On average, forecasters predicted the correct sign of output comparator effects 51.4% of the time, but only 19.8% correctly predicted the sign of input comparator effects—below the 33% expected from uniform random guessing. These results suggest that our taxonomy captures a regularity that was not previously recognized.

Limit(ation)s. We believe that readers will be able to identify instances of comparison effects that go against our taxonomy. One reason is that, in field settings, input and output comparators are often correlated (e.g. hourly wage and accumulated earnings), so that the estimated effect of the observed comparator can be spuriously positive, negative, or zero depending on the relative strength of the two channels. We illustrate such omitted variable bias using misspecified regressions in our own data. A second reason is that our taxonomy is orthogonal to the well-known pattern that people assimilate to comparison points when the difference between comparison point and target objects is very small, as in work on minimum detectable differences and salience (e.g., Bordalo et al., 2020, 2025).

Third, our account naturally struggles to explain the effects of entirely uninformative comparison points, as in anchoring studies or when endowments are transparently randomized. As noted, one possibility is that people are accustomed to somewhat informative comparison points, and may thus heuristically use even entirely uninformative comparison points in an unfamiliar experimental environment.

Fourth, our hypothesis about the role of comparisons is entirely compatible with people *also* exhibiting preferences over comparisons, such as in models of reference-dependent preferences or of social emotions such as envy. Moreover, in its basic form, our hypothesis does not entail asymmetries between positive or negative comparisons (e.g., loss aversion). Below we discuss under which assumptions some of our results can potentially be explained by (unusual variants of) classic reference-dependence models, and why many others cannot.

Contribution and related Literature. Our paper builds on the large literature that documents the importance of reference points and other comparison points (e.g., Barberis, 2013). Our paper makes three contributions to this literature. First, we propose and test the input-vs.-output comparator taxonomy as a tool to predict the direction of comparison effects in applications, and to reconcile some of the contradictory patterns observed in the literature.

Second, our paper connects different literatures on comparison points. Traditionally, the literature has sometimes viewed reference points, sequential evaluation effects, anchors and other comparison effects as distinct. We propose that they are connected through the lens of decision uncertainty and information acquisition, and that this lens can be used to classify a broad range of comparison effects in the literature. For instance, while default effects are sometimes viewed as reflecting information effects (e.g., Madrian and Shea, 2001; Choi et al., 2002; Altmann et al., 2025), this perspective is less commonly invoked for reference points.

Our third contribution is to empirically document an important role for uncertainty in driving the differential effects of input- and output-based comparison points. In doing so, we link to a small recent theoretical literature that has modeled reference points as sources of comparative information for decision makers who are uncertain how to value options (Villas-Boas, 2024). For instance, Dean et al. (2026) present a model in which decision makers receive ordinal comparative information about how an option compares to a reference point; this model predicts contrast effects to comparison points and assimilation to beliefs about comparison points, broadly consistent with our results on input and output comparators.[2] Related, various experimental papers have documented cognitive or framing effects in reference dependence (e.g. Gneezy and Potters, 1997; Ariely et al., 2003; Imas, 2016; Webb et al., 2021). For example, comparison effects sometimes weaken with expertise (List, 2003, 2004) or financial literacy (Dhar and Zhu, 2006), and increase in decision uncertainty (Hsee, 1996; Jin et al., 2024) or complexity (Arieli et al., 2011).

More generally, the recent literature on decision and value uncertainty is largely silent on comparison effects (e.g., Agranov and Ortoleva, 2017; Woodford, 2020; Khaw et al., 2021; Gabaix, 2019; Frydman and Jin, 2021; Charles et al., 2024; Enke and Graeber, 2023; Enke et al., 2024; Bastianello and Imas, 2025; Augenblick et al., 2025; Yang, 2023).

Our paper also connects to work in perceptual psychology. In sequential magnitude estimation experiments, a recent result is that subjects' estimates decrease in the comparison stimulus but increase in the response to the comparison stimulus (e.g. Moon and Kwon, 2022; Imhoff and Barker, 2023; Sadil et al., 2024). Gallagher and Benton (2022) show that sequential dependence effects in perceptual experiments are more pronounced when the degree of uncertainty in the stimulus is higher. Our paper documents analogous effects in economic choice.

Section 2 lays out a simple stylized model. Sections 3-4 present the experimental design and results. Sections 5 and 6 present evidence on mechanisms. Section 7 classifies the results of prior papers, Section 8 reports on a forecaster survey and Section 9 concludes.

2 Framework

2.1 Defining Contrast and Assimilation

We first define contrast and assimilation effects without relying on a model of what generates these effects. Our exposition follows textbook definitions from social and perceptual psychology (e.g. Wedell et al., 2007).

Consider a problem of the form $\underset{a\in\mathbb{A}}{max}\ u(a,x)$, where a is an action and x an exogenous scalar problem input. For instance, in a psychology experiment, the decision maker (DM, he) may be asked to provide an estimate, a, of the number of dots on a screen (x). In an economic context, x may be a piece rate and a the chosen effort level. The optimal action is $a^*(x,\theta) = \arg\underset{a\in\mathbb{A}}{max}\ u(a,x)$, where θ captures all factors that influence how the decision-maker translates a

[2]Wu (2024) documents the importance of the decision maker's perceived uncertainty about the target and reference objects for contrast and assimilation. Bondi et al. (2025) show that range effects—a form of comparison effects—increasing in choice set complexity. Broadly related are also imperfect-information models in which decision makers learn from other options (e.g., Kamenica, 2008; Bhui and Xiang, 2021; Shubatt and Yang, 2023).

decision input into an optimal action. This includes preferences but more broadly encompasses the "technology" that determines the DM's optimal decision, such as constraints, scales, and other factors affecting the optimization problem. a^* is the decision the agent would take in the absence of any processing constraints, biases, preference uncertainty and the like. We normalize the input variable x such that the optimal action $a^*(x, \theta)$ is increasing in x.

The DM takes his decision in the presence of a comparison point, c. Denote the observed decision $a(x, \theta, c)$. Contrast and assimilation are defined as

$$\underbrace{a(x,\theta,c)}_{\text{Decision w/ CP}} = \underbrace{a^\star(x,\theta)}_{\text{Opt. decision w/o CP}} + \underbrace{w \times [c - a^\star(x,\theta)]}_{\text{Comparison effect}} + \underbrace{u}_{\text{Other biases}} = (1-w)a^\star(x,\theta) + wc + u\,. \quad (1)$$

Then, under our normalization that a^* increases in x, assimilation means $w > 0$. This implies that the decision is closer to the comparison point than it should be, and that the decision increases in the comparison point. For instance, if $w = 1$ (and $u = 0$), the DM plays exactly the comparison point. Contrast ($w < 0$) implies that the decision is further away from the comparison point than it should be, and that the decision decreases in the comparison point.

To illustrate, in a psychology experiment in which subjects estimate the number of dots on a screen, the comparison point may be given by the number of dots on the preceding trial. Then, assimilation means that the subject's estimate in the target round increases in the preceding number of dots, while contrast means that the current estimate decreases in it.

2.2 Decision Uncertainty and Comparison Points as Information

We consider the possibility that the DM is uncertain about how to map objective problem inputs into optimal decisions (outputs). We are agnostic about the sources of this uncertainty: the DM may be uncertain about his preferences, he may be unsure how to maximize, he may have imperfect perception, he may not have access to formal rules such as Bayes' rule, and more.

To reduce his uncertainty, he relies on a comparison problem that may be informative about how to map decision inputs into outputs, for example because he or someone else made a similar decision before. Below is a non-exhaustive list of examples.

1. When deciding how many hours to work at a given hourly wage, my own past decision about how much to work at a different wage may be informative about how I trade off money and leisure.

2. When deciding how many hours to work at a given hourly wage, the decision of someone else in the same context may be informative about how painful the job is, or—if our preferences are correlated—about how to trade off money and leisure.

3. In an experiment, the parameters set by the experimenter may be (perceived as) informative because experimenters set parameters that are ‘reasonable’.

Our main objective is to present a simple model that (i) highlights the differential effects of different types of comparison points; (ii) through the lens of comparison points acting—at least implicitly—as sources of information. The model should be viewed merely as an illustration of a general class of models to fix ideas. Our experiments are not designed to test a specific model variant. Rather, we view them as tests of the general idea that, when comparison points are used as information, different types of comparison points push in different directions.

2.3 Stylized Model

We continue to consider a problem of the form $\max_{a \in \mathbb{A}} u(a, x)$, with solution $a^*(x, \theta)$. We begin with a simple linear case for illustration. Suppose the optimal action is given by

$$a^\star(x, \theta) = \theta x, \qquad \theta > 0.$$

Here, θ again captures all factors that matter for the mapping of inputs into optimal decisions but does not depend on comparison points.

The DM wishes to choose an action a that tracks $a^\star(x, \theta)$. However, he does not know θ. The DM chooses an action equal to his mean belief about θ times x:

$$a(x, \text{info}) = \mathbb{E}[\theta \mid \text{info}]\, x.$$

We model the DM as Bayesian for simplicity. However, what matters for our predictions is not so much that the DM is literally Bayesian but rather that he treats comparison points as informative. The DM holds a prior about θ with mean μ_θ and variance σ_θ^2. Moreover, he has access to comparison data, generated according to

$$a_c = \theta_c x_c.$$

The parameter θ_c captures the mapping between inputs and optimal actions in the comparison context. It need not coincide with the DM's own mapping parameter θ. For example, the comparison problem may involve a different individual with correlated but not identical preferences, or a similar but not identical decision environment. The DM observes either one of or both a_c and x_c, which we refer to as *comparators*: variables that allow a comparison to a component of his own decision problem. x_c is an *input comparator* and a_c an *output comparator*.

This already highlights a recurring point of our paper: that the reduced-form definition of comparison effects in eq. (1) is ambiguous because it does not specify whether the comparison point c lives in input space ($c = x_c$) or output space ($c = a_c$).

Comparators have two characteristics. First, conditional on (x, θ), they do not affect the optimal action. In this model, comparison points are thus "supposedly irrelevant factors" (Thaler, 2016). Of course, we do not intend to rule out that true reference utility or social comparison utility exist—we're just abstracting from them here to illustrate our effect of interest.

Second, the DM believes that θ_c is informative in the sense of posterior monotonicity:

$$\frac{\partial\, \mathbb{E}[\theta \mid \theta_c]}{\partial \theta_c} > 0.$$

A simple case satisfying this condition is $\theta_c = \theta + \nu$, where ν is mean-zero noise. While there are many reasons to expect θ and θ_c to be different, we only require that θ_c is treated as informative about θ. The model, and this condition in particular, has an *as-if* nature: it neither requires that the comparison context is objectively informative, nor that the DM *explicitly* believes that it is. All we require is that the DM implicitly treats it as informative in his decisions.

Baseline case: Both comparators are observed. Suppose the DM observes (a_c, x_c) and infers $\theta_c = \frac{a_c}{x_c}$ exactly. He then chooses

$$a(x; a_c, x_c) = \mathbb{E}[\theta \mid \theta_c]\, x, \tag{2}$$

with $\frac{\partial\, \mathbb{E}[\theta|\theta_c]}{\partial \theta_c} > 0$ by informativeness. This baseline case yields our first prediction.

Prediction 1. *The DM's action $a(x; a_c, x_c)$ exhibits assimilation in the output comparator a_c and contrast in the input comparator x_c:*

$$\frac{\partial a}{\partial a_c} > 0 \qquad \textit{and} \qquad \frac{\partial a}{\partial x_c} < 0.$$

All proofs are in Appendix A. Intuitively, a higher output comparator a_c (holding x_c fixed) implies a higher $\theta_c = \frac{a_c}{x_c}$ and thus a higher posterior mapping $\mathbb{E}[\theta \mid \theta_c]$, which increases a. On the other hand, a higher comparison input x_c (holding a_c fixed) implies a lower $\theta_c = \frac{a_c}{x_c}$ and thus a lower $\mathbb{E}[\theta \mid \theta_c]$, decreasing the chosen action.

To make the mechanics concrete, consider a Gaussian case for illustration,

$$\theta \sim N(\mu_\theta, \sigma_\theta^2), \qquad \theta_c = \theta + \epsilon, \qquad \epsilon \sim N(0, \sigma_\epsilon^2),$$

with θ and ϵ independent. Standard Bayesian updating yields

$$a(x; a_c, x_c) = (1-\lambda) \cdot \mu_\theta \cdot x + \lambda \cdot a_c \cdot \frac{x}{x_c}, \qquad \lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}. \tag{3}$$

The formula illustrates two features. First, the DM's action is a precision-weighted average of the prior mean μ_θ and the inferred mapping a_c/x_c from the comparison problem. Second, the comparison-based term involves comparative thinking: the ratio x/x_c scales the comparison action a_c to the DM's own context. If the input in the target problem is greater than in the comparison problem ($x > x_c$), he scales up; if smaller, he scales down. Equation (3) implies assimilation in the output comparator a_c and contrast in the input comparator x_c.

We use this Gaussian setup to illustrate how comparator effects depend on the degree of mapping uncertainty and informativeness, but the predictions are general (see Appendix A).

Prediction 2. *In the linear Gaussian benchmark:*

*(2a) (*Mapping uncertainty*) The magnitude of comparator effects increases in prior mapping uncertainty:* $|\partial a/\partial a_c|$ *and* $|\partial a/\partial x_c|$ *increase in* σ_θ^2.

*(2b) (*Informativeness*) The magnitude of comparator effects increases in the precision of the comparison signal. Equivalently,* $|\partial a/\partial a_c|$ *and* $|\partial a/\partial x_c|$ *decrease in the noise variance* σ_ϵ^2.

Only one comparator observed. What happens if the DM observes only one comparator? In practice, some comparison points are unobservable and some are more salient than others, for example because they are highlighted by experimenters. In such cases, the same objective decision problem can produce very different behavior purely depending on whether the input or output comparator is salient. In such cases, the comparison context (a_c, x_c, θ_c) is still generated by $a_c = \theta_c x_c$, but the DM either sees only x_c or a_c. Appendix A.3 shows that, under mild assumptions, the predictions stated above still hold.[3]

Actions, outcomes and bunching. In some applications, the output comparator of interest is a_c itself (e.g., previous number of hours worked). In others, the relevant comparison quantity is an associated *outcome*, such as total accumulated earnings. Suppose that outcome, y, and action are linked through $y = g(a, x, \eta)$, where η captures factors outside the DM's control.

[3]Matters are different when the DM observes both comparators but the analyst observes only one of them. In these situations, omitted variable bias can distort inferences about contrast and assimilation because a_c and x_c will often be correlated. For instance, if an analyst runs a regression of decisions on the observed input comparator x_c, he may incorrectly estimate an assimilation effect with respect to the input comparator because a_c and x_c are correlated. See Appendix A.6. This omitted variable bias highlights the need to implement carefully controlled experiments, in which we can observe both comparators. See Appendix C for evidence from our experiments.

We normalize $g(\cdot)$ to increase in a without loss of generality. Based on the distinction between actions and outcomes, we also distinguish between two types of output comparators: *action comparators*, a_c, and *outcome comparators*, y_c. Because outcomes increase in actions, all comparative statics predictions for a_c derived above apply equally to y_c, and for simplicity we call both output comparators.

Some empirical work identifies comparison effects through *bunching* at *outcome comparators*, such as earnings in labor supply, marathon finish times, or par in golf. From the perspective of our framework, such bunching can be viewed as an extreme form of assimilation that arises when the DM places nearly all weight on comparison-based information and little weight on their own prior. Suppose that the DM only observes y_c but not the associated action, a_c. For instance, the DM may observe his goal for a finishing time in marathon, but not how much physical effort is required to achieve it. Further suppose that prior mapping uncertainty is large ($\sigma_\theta^2 \to \infty$), such that the DM almost entirely relies on the comparison point in forming beliefs about θ. Then, if $x \approx x_c$ (or if the DM is inattentive to the precise value of x and thus relies on x_c), the DM chooses an action that replicates the comparison outcome, $g^{-1}(y_c; x_c)$, which produces outcome $y = y_c$. Intuitively, when the DM has little knowledge about how to solve the problem, he treats the observed outcome in the comparison problem as a guide for what outcome to generate in his own decision. Appendix A.4 provides the formal details.

This said, we emphasize that our stylized setup will typically not predict bunching at the outcome comparator. At the same time, models of comparison points as information that leverage the idea of partially ordinal information ('I know that $x > x_c$ and thus that $a > a_c$') naturally predict bunching at comparison points (Dean et al., 2026). We are agnostic about which precise mechanism gives rise to bunching, except that we emphasize that both of these accounts predict this extreme form of assimilation with respect to output comparators, rather than input comparators. This is the prediction we test.

General (nonlinear) model. The key force behind our results is not the 'ratio structure' of the Gaussian-linear benchmark, but the fact that comparison data are used to infer an unobserved mapping from inputs to optimal actions. In Appendix A we show that—under monotonicity and invertibility conditions—the same sign predictions extend to a broader nonlinear class.

Interpretation of transparently uninformative comparison points. The stylized setup above assumes that—if the DM is uncertain about how to map circumstances into optimal decisions—the comparison point is objectively informative, or perhaps at least perceived as such. In practice, people may also respond to comparison points that are not (perceived as) informative, such as when an experimenter transparently randomizes the comparison point. In these situations, people may still heuristically act *as if* the comparison point contains information, perhaps because they have become adapted to somewhat-informative comparison points. The key implication for us (in particular for our classification of the prior literature) is that even in those contexts in which comparison points are transparently informative, we expect to see assimilation to output comparators and contrast to input comparators.

3 Experiment Design

Table 1 presents an overview of all experiments that we conducted for this paper. There are four main paradigms (*Effort Valuation*, *Investment*, *Beliefs* and *Effort Choice*). The treatment labels indicate the type of comparison points being presented in the experiment (input and / or output comparator), and the source of the comparison points (expectations or social comparisons).

The *Effort Valuation* paradigm is our workhorse application, in which we study (i) both social and expectations-based comparison points and (ii) situations in which both input and output comparators are shown, and situations in which only one of the two is displayed. Moreover, we also use this workhorse paradigm (iii) to study whether the same comparison variable can have positive and negative effects, purely depending on whether it constitutes an input or output comparator (by comparing with paradigm *Effort Choice*); and (iv) to conduct mechanism treatments that study whether comparison points affect behavior (at least in part) because they serve as sources of information. The *Investment* and *Beliefs* paradigms serve to study whether the logic of input and output comparators also holds outside of the context of effort supply and, in particular, in a setting with an objectively correct solution.

Appendix H provides all experimental instructions, comprehension quizzes used, the flow of each experiment and example decision screens.

3.1 *Effort Valuation* Paradigm

In the *Effort Valuation* paradigm, subjects are asked to state their minimum acceptable compensation (willingness-to-accept, WTA) for completing a specified workload of a standard real effort task. In this paradigm, the offered workload (i.e., number of tasks) is the input and the minimum compensation is the decision (or output). There are 15 independent rounds. Each round specifies a workload for one of three canonical real-effort tasks: transcribing blurry Greek letters, counting how often a specific number occurs in a table, and setting sliders to a specific value. An example is that subjects are asked to state their WTA for transcribing 1,200 Greek letters. The variation in tasks serves to minimize direct between-round comparisons. See Figures 11–13 in Appendix H.1.1 for details on each task.

Table 1: Experimental paradigms and treatment conditions

Paradigm	Treatment condition	Objective	# subjects
Effort Valuation Decision: WTA Input: workload	No CP	Baseline; no comparison points	150
	Both CP (social)	Social input and output comparators	250
	Input CP (social)	Only social input comparator	250
	Output CP (social)	Only social output comparator	250
	Both CP (expectations)	Expectations-based input and output comparators	250
	Deliberation (social)	Mechanism: reduce uncertainty	450
	Time (social)	Mechanism: reduce uncertainty	350
	Informativeness (expectations)	Mechanism: reduce informativeness	300
Investment Decision: amount wagered Input: winning chance	No CP	Baseline; no comparison points	150
	Both CP (expectations)	Expectations-based input and output comparators; document same effects outside of effort context	250
Beliefs Decision: posterior belief Inputs: prior, accuracy, test result	No CP	Baseline; no comparison points	150
	Both CP (social)	Social input and output comparators; document same effects in obj. task	250
Effort Choice Decision: workload Input: piece rate	No CP	Baseline; no comparison points	150
	Both CP (expectations)	Comparing with *Effort Valuation*, flip workload from output to input	250

In each round, we independently draw a workload for the participant. To make the workloads somewhat similar across task types, we first draw a time estimate (from 10 to 70 minutes) and then multiply it by a conversion factor for each task type that is based on pilot data.

Participants were incentivized. A randomly-selected round was paid out using a standard Becker-deGroot-Marschak mechanism (see Appendix Figure 9). The participant had to complete the full workload of the randomly-selected round at the end of the study and received the corresponding compensation as a bonus. We study four between-subject conditions.

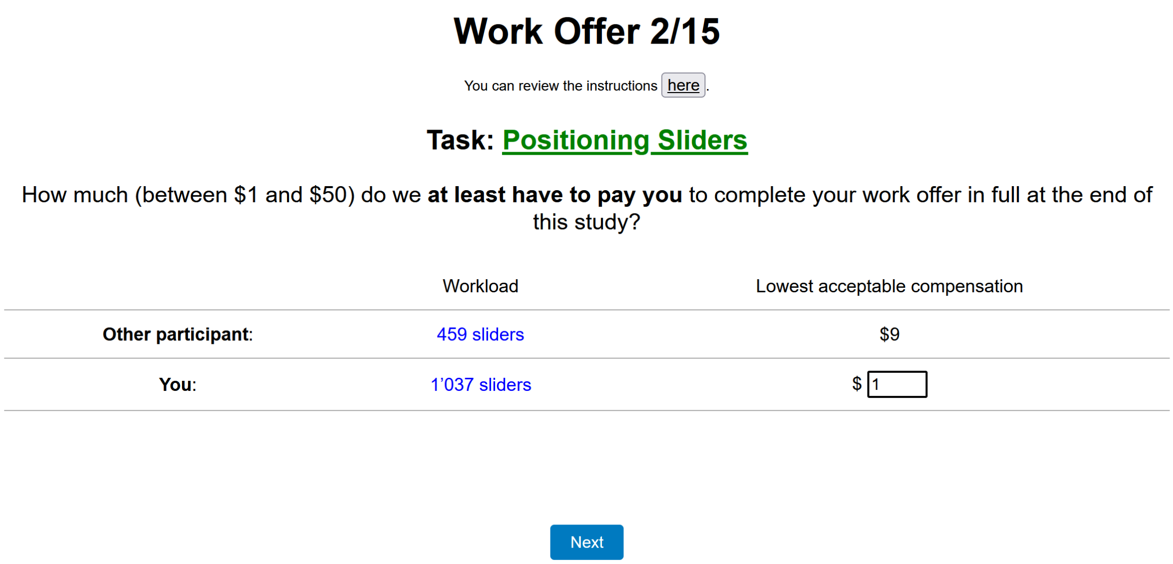


Figure 1: Decision screen in the *Social Both CP* treatment in the *Effort Valuation* paradigm.

No CP. In the *No CP* condition, participants make their decisions without an externally provided comparison point. This serves as a benchmark.

Both CP (social). Participants make the same decisions as in *No CP*, but they are also shown the workload offered to a random *No CP* participant in a randomly-selected round, along with that participant's decision. These correspond to social input and output comparators, respectively. Appendix Figure 32 shows how the comparison points were introduced. After subjects had completed the instructions and a comprehension quiz, they were told: "On the next screen, you will evaluate your first work offer. Please note: We will also show you the workload another participant was offered as well as the lowest acceptable compensation they chose for their work offer. The other participant and their workload were selected at random." Figure 1 shows a screenshot of a decision screen.

Both CP (expectations). In this treatment, the comparison point is instead induced via expectations, as in, for example, Abeler et al. (2011). As shown in Appendix Figure 30, subjects were told that (if a round was randomly selected to count), their actual workload and compensation would be determined by a coin flip. With 50% chance, the participant's own decision counts. With 50% chance, the subject's workload and compensation are determined by a "default workload and default compensation". This default workload and compensation (the expectations-based comparison points) are transparently independent of the subject's decisions and displayed on the decision screen. For instance, subjects may be told on their decision screen that, if a coin flip comes up Heads and their own decision does not count, the (default) workload is to transcribe 900 Greek letters and the (default) compensation is $50, see Appendix Figure 31 for a screenshot.

The instructions emphasize to subjects that the default workload and compensation are exogenously-determined and don't vary with the subject's decisions. At the same time, the instructions never suggest that these quantities are randomly-determined. This may be relevant from the perspective of our model, in which the comparison point is treated as informative about the DM's best decision. In our experiment, the exogenous default workload and compensation may still be perceived as informative by subjects: most likely, the experimenter does not select workloads and compensations at random from the set of all positive numbers. Rather, arguably, the experimenter sets default compensations that are "reasonable."

Input CP only / Output CP only (social). To study whether the dual pattern of contrast to the input comparator and assimilation to the output comparator only emerges when both comparison points are supplied jointly, we also study treatments in which only one of them is displayed. In *Input CP (social)*, we provide only another participant's offered workload, while the other participant's WTA is displayed as "??" on the screen. Treatment *Output CP (social)* is analogous, except that only the other participant's decision is displayed.

3.2 *Investment* Paradigm

Subjects decide how much of a $1,000 budget to wager on a risky gamble that doubles the investment with a certain winning chance, and halves the investment with the remaining probability. The money not wagered is kept with certainty. Subjects make decisions across 15 rounds with varying winning chances. One round was randomly selected for payment and participants were paid the final value of their investment, divided by 100. In this setting, the gamble's winning chance is the input and the amount wagered the output. We run two treatments.

No CP. This treatment serves as a benchmark without comparison point.

Both CP (expectations). In this treatment, we provide an expectations-based comparison point akin to the one in the *Effort Valuation* paradigm: with 50% chance, the subject's own decision counts, and with 50% chance the experimenter invests an "imposed" amount on behalf of the subject into an "imposed" gamble that has a different winning chance than the gamble the subject can invest in. Here, the winning chance of the imposed gamble is the input comparator and the imposed amount wagered the output comparator. The instructions again emphasize that the imposed gamble doesn't vary with the subject's decisions, but we also take care to make sure that we never suggest that the imposed amount wagered is randomly-determined.

3.3 *Beliefs* Paradigm

We also study a structured belief updating problem in which direct preferences over comparisons cannot plausibly play a role. We implement a canonical test-and-disease problem (e.g., Kahneman and Tversky, 1973) in which subjects are asked to guess whether a hypothetical patient has a certain medical condition, given the prior likelihood of the condition and a test result with specified accuracy. In this paradigm, there are three inputs: the prior, the test accuracy and the test result (positive or negative). The output is the subject's stated posterior belief. Each participant completes 15 rounds. A subject won $20 if their guess in a randomly-selected round was within ±3 percentage points of the statistically-correct guess. We run two treatments.

No CP. This condition provides a benchmark without comparison point.

Both CP (social). In each round, subjects additionally see the updating problem (prior, test accuracy and test result) faced by a random *No CP* participant, along with that participant's guess. Here, the primitives of the updating problem constitute the input comparators and the other participant's guess the output comparator. We design the updating problems such that either the prior or the signal realization differ between the comparison task and the subject's own target task. If subjects had no uncertainty about how to translate the primitives of the updating task into a posterior belief, there would be no reason to rely on the comparison points. While intrinsic preferences over the social comparison may play a role in, for example, our effort supply experiment ("I want to earn at least as much as the other subject"), it seems difficult to imagine that a subject would have intrinsic preferences over whether their own stated posterior is higher or lower than that of another subject.

3.4 Procedures

Participants in all studies received a base reward that corresponded to an hourly wage of $9 for the estimated completion time of the experiment. Median completion times were 16 minutes in *Effort Valuation*, 14 minutes in *Effort Choice* (excluding the time it took to complete the actual work task), 10 minutes in the *Investment* paradigm and 12 minutes in the *Beliefs* paradigm. Across all experiments, the average payment to subjects was $4.5, which includes an average show-up fee of $2.98.

All experiments were conducted on the online platform Prolific. All experiments were pre-registered; Appendix Table B1 provides links to all pre-registrations. The pre-registrations detail the exclusion restrictions for our main analyses: (i) we exclude participants who fail a comprehension test on the baseline instructions (which are identical across treatments within a given paradigm); (ii) to validate that each respondent is human, we require that each participant record a video of themselves in the beginning of the study; we exclude any participant who we cannot identify as human; (iii) we use a standard reCaptcha bot test on the first page of the survey; (iv) we additionally collect Qualtrics's Captchav3 bot scores and exclude every participant with a score of 0.5 or below (meaning that, according to Qualtrics, the participant is more likely to be a bot than a human).

Our main data collection was completed prior to the release of the first widely available LLM agent (ChatGPT Agent on July 17, 2025). Only the *Deliberation* and *Informativeness* treatments were collected later. However, no commercially available LLM agent could pass our video recording test at the time of data collection.

4 Results

We here present the results from our pre-registered main analyses. Pre-registered robustness checks and secondary analyses are reported in Appendix G.

4.1 *Effort Valuation*

As pre-registered, we normalize the workloads of the three different real effort tasks by expressing them in terms of their average completion time. This allows us to analyze all decisions in a single set of (pre-registered) regressions.

For completeness, we first note that, in all treatments, subjects' minimum acceptable compensations (WTAs) unsurprisingly strongly increase in the required workloads. This suggests that subjects understood the general experimental setup. Our main object of interest, however, is how subjects' WTAs vary as a function of the input comparator (the comparison workload) and the output comparator (other subjects' WTA or exogenously-imposed payments).

Figure 2 summarizes the results. The panel titles show treatment labels. The panels on the left-hand side show how average WTA varies with the input comparator. These figures are partial effect plots (or conditional binscatters) that control for the actual input (required workload), the output comparator, as well as task type and participant fixed effects.

The panels on the right-hand side show how average WTA depends on the output comparator, which is given either by other subjects' WTA (in the Social CP treatments) or by the exogenously-imposed compensations (in the expectations treatment). The right-hand side figures are also partial effect plots, constructed controlling for the required workload, the input comparator, as well as task type and participant fixed effects.

We see that subjects' WTA strongly decreases in the input comparator, which is the comparison workload. At the same time, WTAs always strongly increase in the output comparators. These results hold both when the comparison point is given by expectations and when it reflects social comparisons. Moreover, the results also hold when only one type of comparison point (input or output) is supplied.

To investigate the statistical significance of these results, we turn to regression analysis, summarized in columns (1)–(5) of Table 3. In these analyses, the dependent variable is the stated WTA in each subject-round, for a total of 15 observations per subject. The independent variables are the offered workload, the input comparator and the output comparator. As we pre-registered, we include fixed effects for the task type (i.e., the type of real effort task) and participant fixed effects. The standard errors are clustered at the participant level. Again, the workload variables are standardized across real effort tasks.

The regression results confirm the picture that emerges from Figure 2. We always see a pronounced negative effect of the input comparator and a positive effect of the output comparator.

4.2 *Investment*

Figure 3 visualizes the results for the *Investment* experiment. The left-hand panel shows average investments (in % of the budget) as a function of the input comparator, i.e. the default asset's winning chance. The right-hand panel shows average investments as a function of the output comparator, i.e., the default investment. Again, these figures are constructed as conditional binscatters that control for the current winning chance (the input), participant fixed effects as well as the output comparator (in the left-hand side panel) or the input comparator (in the right-hand side panel). We find a pronounced negative effect of the input comparator and a positive effect of the output comparator. Columns (6)–(7) of Table 3 confirm the statistical significance of these patterns.

4.3 *Beliefs*

Unlike in the *Effort Valuation* and *Investment* experiments discussed above, in the *Beliefs* experiment there are multiple input parameters that determine a subject's decision. In analyzing the data, we resort to standard Grether (1980) techniques, meaning we relate a subject's stated log posterior odds to the log prior odds and the log likelihood ratio (which incorporates both the test accuracy and the test result). Thus, the input comparators are the log prior odds and log likelihood ratio encountered by another subject, and the output comparator is the other subject's stated log posterior odds.

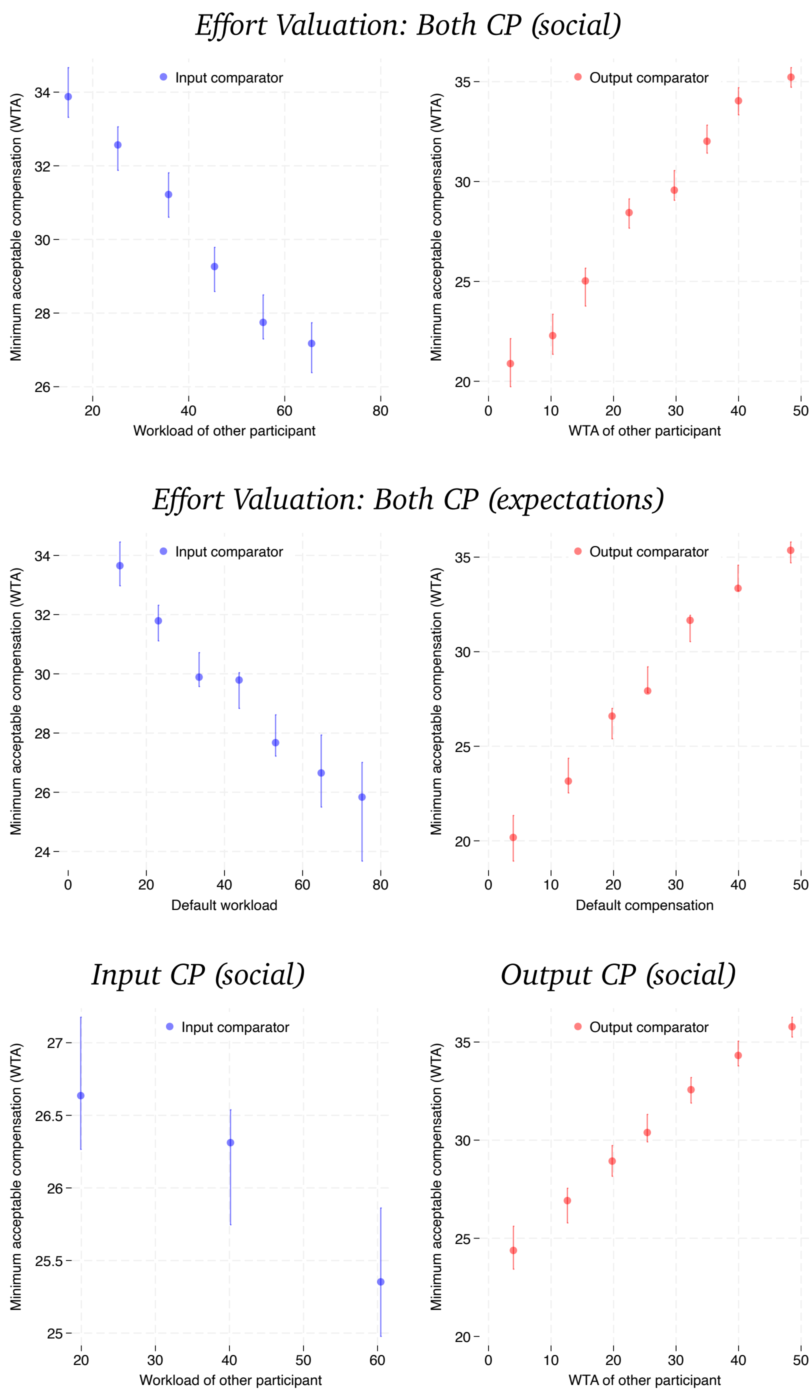


Figure 2: Results of treatments in *Effort Valuation* paradigm. All left-hand side panels show average WTA as a function of the input comparator, which is given by other subjects' workload in the social treatments and by the default workload in the expectations treatment. All right-hand side panels show average WTA as a function of the output comparator, which is given by other subjects' WTA in the social treatments and by the default compensation in the expectations treatment. All panels are constructed from multiple regressions as in columns (2)–(4) of Table 3. Points show binned mean estimates, with 95% confidence intervals based on robust standard errors, computed using Stata's *binsreg* command, which approximates the conditional mean function with a piecewise polynomial. Asymmetric CI's can result from nonparametric local fits.

Table 3: Regression results for *Effort Valuation*, *Investment* and *Beliefs* experiments

Paradigm:	**Effort valuation**					**Investment**		**Beliefs**	
Dependent variable:	Minimum acceptable compensation (WTA)					Amount wagered		Log posterior odds	
Treatment:	*No CP*	*Both CP (social)*	*Both CP (expectations)*	*Output CP only (social)*	*Input CP only (social)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (social)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input (workload)	0.146***	0.243***	0.190***	0.183***	0.150***				
	(0.0139)	(0.0113)	(0.0130)	(0.0113)	(0.0110)				
Input comparator (workload)		-0.142***	-0.133***	-0.0302***					
		(0.0120)	(0.0138)	(0.00976)					
Output comparator (WTA)		0.321***	0.338***		0.250***				
		(0.0176)	(0.0195)		(0.0169)				
Input (winning chance)						8.501***	8.230***		
						(0.280)	(0.290)		
Input comparator (winning chance)							-0.371**		
							(0.176)		
Output comparator (amount)							0.0642***		
							(0.0147)		
Input (log prior odds)								0.440***	0.403***
								(0.0320)	(0.0297)
Input (log likelihood ratio)								0.845***	0.773***
								(0.0442)	(0.0319)
Input comparator (log prior)									-0.0678**
									(0.0272)
Input comparator (log likelihood ratio)									-0.0439*
									(0.0232)
Output comparator (log posterior)									0.155***
									(0.0242)
Observations	2175	3660	3690	3735	3690	2175	3705	2180	3610
Subjects	145	244	246	249	246	145	247	148	246
R-squared	0.754	0.693	0.654	0.735	0.665	0.791	0.682	0.635	0.550

Notes. Main pre-registered specification. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for paradigms *Effort Valuation* and *Effort Choice* also include task type fixed effects.

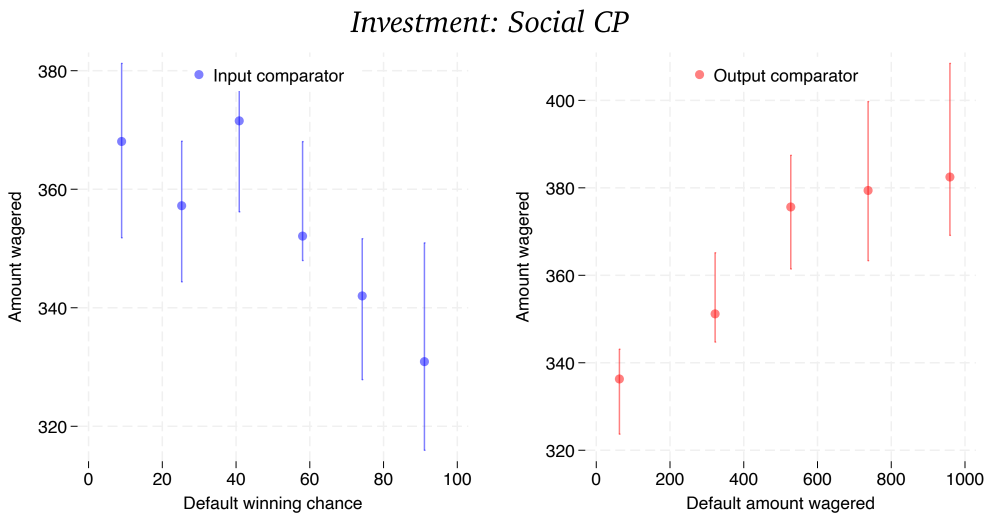


Figure 3: Results of *Investment* experiment. The left panel shows average investments as a function of the comparison winning chance. The right panel shows average investments as a function of the comparison amount wagered. All comparators are based on expectations. Both panels are constructed from multiple regressions as in column (7) of Table 3. Points show binned mean estimates, with 95% confidence intervals based on robust standard errors.

Figure 4 visualizes the results. We again find negative effects of both input comparators (the comparison log prior odds and the comparison log likelihood ratio) and a positive effect of the output comparator. Columns (8)–(9) in Table 3 report the associated Grether regressions.

This final application to beliefs and inference suggests that the empirical regularity of a negative effect of the input comparators and of a positive effect of the output comparators is not limited to preference-based choice but similarly emerges in objective problems in which, for example, reference-dependent preferences cannot play a role.

4.4 Robustness and Secondary Analyses

Appendix G reports the full set of pre-registered robustness checks and secondary analyses across all paradigms and treatments (Tables G1–G14). These include alternative sample restrictions (strict human verification, response-time validity, protocol compliance), median regressions, round interactions, and sample splits by whether the comparison is favorable or unfavorable. The results are highly stable: across all specifications, the core pattern of input contrast and output assimilation effects is unchanged in sign, significance, and approximate magnitude.

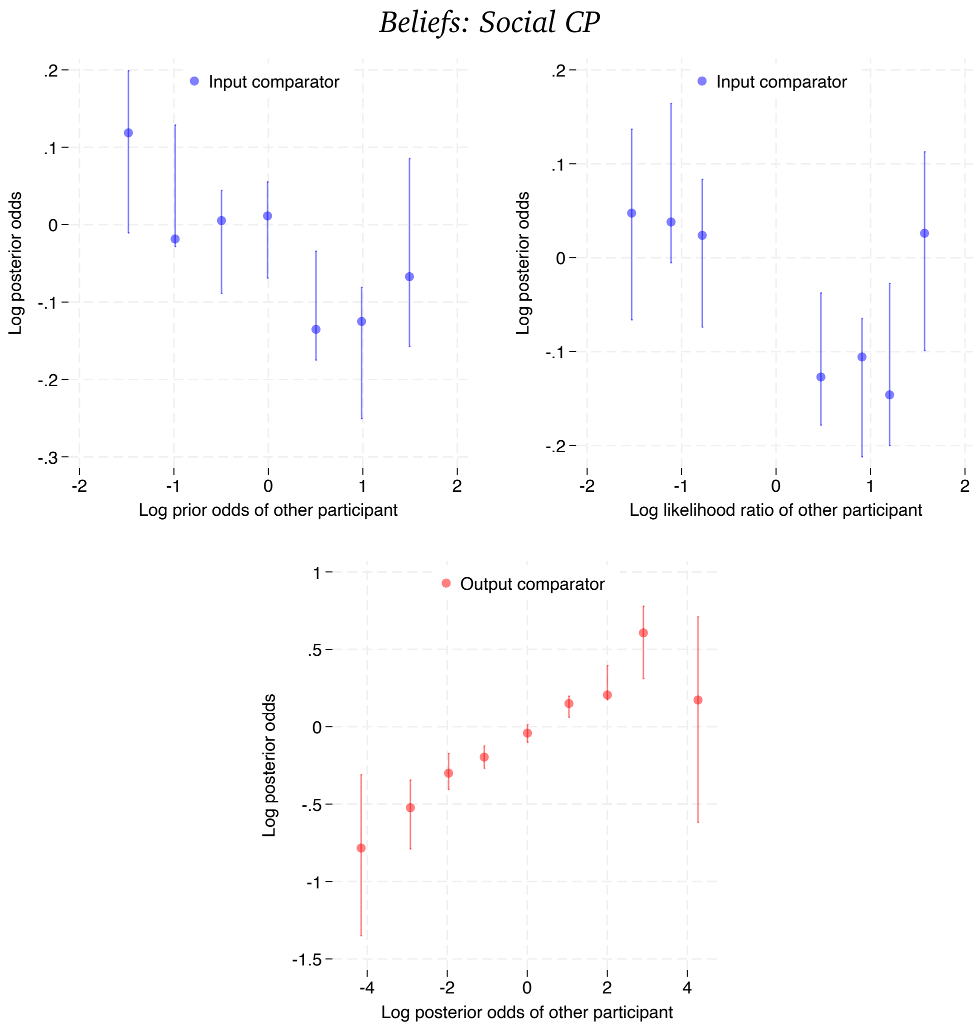


Figure 4: Results of *Beliefs* experiment. The top panels show average log posterior odds as a function of the comparison log prior odds and the comparison log likelihood ratio. The bottom panel shows the results for the comparison log posterior odds. All comparators are from a social comparison. The panels are constructed from multiple regressions as in column (9) of Table 3. Points show binned mean estimates, with 95% confidence intervals based on robust standard errors.

5 Manipulating the Type of Comparison Point

A key implication of our account of the effects of comparison points is that the same comparison variable should predictably push decisions up or down, purely depending on whether it constitutes an input or output comparator for the decision at hand.

In the *Effort Valuation* paradigm, the problem input is the specified workload of a given task, which subjects are required to map into a WTA. We complement this paradigm with the *Effort Choice* paradigm, in which subjects are asked to decide how many work tasks they would like to complete at a known piece rate. In this design, the total workload is an output (decision) rather than an input. Thus, across these two paradigms, a comparison workload (the workload of another subject) is either an input comparator or an output comparator.

The *Effort Choice* paradigm is otherwise very similar to the *Effort Valuation* paradigm. In each of 15 rounds, subjects are offered a piece rate and choose a total workload for themselves. Again, there are two treatments.

No CP. We collect decisions in the absence of a comparison point.

Expectations Both CP. Akin to the expectations-based comparison points in the *Effort Valuation* paradigm, we induce comparison points via expectations. With 50% chance, the subject's chosen workload (and associated piece rate) count, while with 50% chance the subject receives a "default" piece rate and completes a "default" workload, both of which the subject has no influence on. For instance, a subject may be told that they need to decide how many tasks they would like to complete at a piece rate of \$0.20, knowing that with probability 50% their decision will not count and, instead, they will need to complete 40 tasks for a piece rate of \$0.10.

See Appendix H for screenshots of instructions, comprehension quizzes and decision screens.

Results. Figure 5 shows the main result by comparing how average decisions in *Effort Valuation* and *Effort Choice* vary as a function of the comparison workload. The y-axes show average decisions, where the left axis indicates decisions in *Effort Valuation* (where subjects stated a WTA) and the right axis shows decisions in *Effort Choice* (where subjects chose a total workload). As seen earlier, in the *Effort Valuation* paradigm, decisions strongly decrease in the comparison workload. In *Effort Choice*, on the other hand, where the comparison workload is turned into an output comparator, decisions strongly increase in the comparison workload. Appendix Table B3 provides an econometric analysis.

We interpret this result as showing that what matters for the sign of a comparison effect is not so much what the specific comparison variable is but whether it represents an input or output comparator for the decision at hand.

6 Mechanism Evidence: Comparisons as Information

Our hypothesis is that the distinction between input and output comparators matters because comparison points may (at least implicitly) be used as sources of information for decision makers who are uncertain how to map the inputs of a decision problem into an output (decision). We are agnostic about the precise sources of this uncertainty. For instance, people might not know their utility exchange rate between consumption and leisure, they might not have perfect access to their own cognitive production function of solving real-effort tasks (how long will it take me

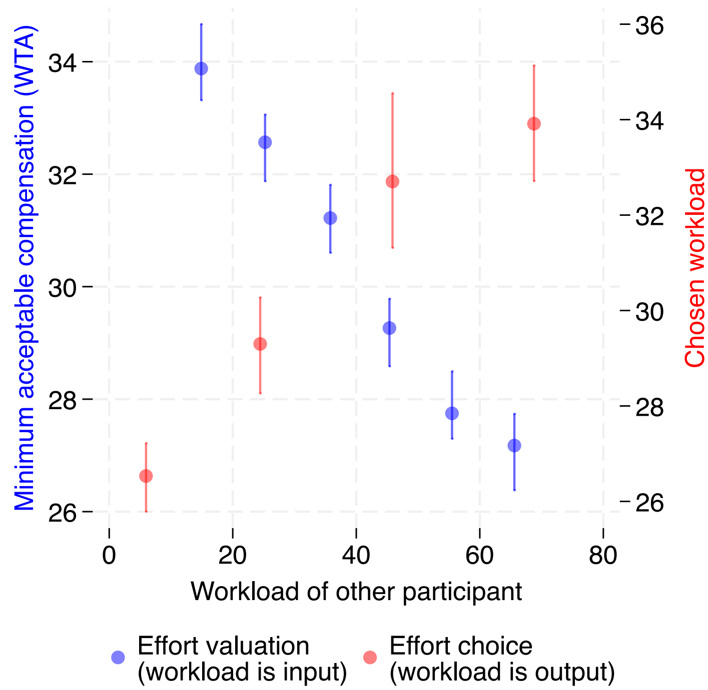


Figure 5: Results in treatments *Effort Valuation* and *Effort Choice*. The x-axis shows the comparison workload supplied to subjects and the y-axis average decisions. In *Effort Valuation*, shown on the left axis, subjects decide their minimum required compensation. In *Effort Choice*, shown on the right axis, subjects decide their total workload.

to count 23 tables?), they might not know Bayes rule, they might have uncertainty about their true degree of risk aversion, or they might find it difficult to optimize. Indeed, given the variety of experiments we run, our proposed mechanism of comparison points as information likely requires that different sources of uncertainty are at play.

To provide a test of the hypothesis that comparison points partly serve as information about what a good decision may be, we designed three experiments, two that manipulate the degree of uncertainty about how to translate the inputs of a problem into a decision (Prediction 2 (a)), and one that manipulates the informativeness of the comparison data (Prediction 2 (b)).

6.1 Reducing Input-Output Mapping Uncertainty

We return to our main application, *Effort Valuation*. In that paradigm, subjects are effectively asked to translate a workload into a monetary valuation. As formalized in Section 2, subjects may not know their optimal policy function: the function that maps inputs into decisions. Our first treatment manipulation consists of encouraging subjects to deliberate more (and in a more structured fashion) about their policy function.

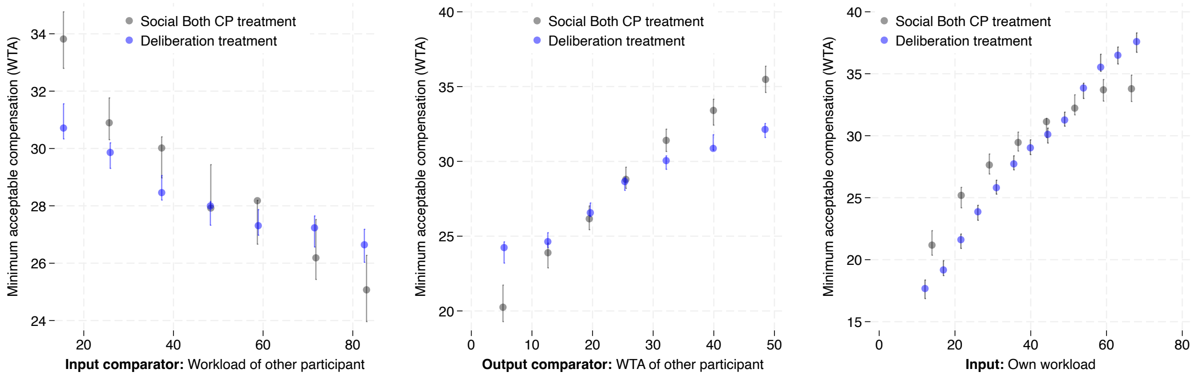


Figure 6: Effect of deliberation on sensitivity to own input and comparators. Results of conditions *Deliberation* and *Social Both CP* in the effort valuation paradigm. The y-axis shows minimum required compensations. From left to right, the x-axis shows the input comparator, the output comparator, and the participant's own input.

Design. The *Deliberation* condition is a variant of the *Social Both CP* condition. Before participants make their incentivized decisions, they encounter a screen that asks them to state—in purely hypothetical terms—their WTA for each of up to 13 different possible workloads of one of the three tasks, in the form of a list (see Appendix Figure 16). Participants only filled out this list for one randomly selected real effort tasks, and they subsequently made their main decisions also for this real effort task. The idea behind this treatment is that stating one's decision for each of many different values of the input helps subjects to familiarize themselves with the decision and better learn (or reduce their uncertainty about) their policy function. Our pre-registered prediction is that this reduces the reliance on the externally-provided social comparison points as sources of information.

Results. We pre-registered a collection of 300 subjects for the *Deliberation* treatment alongside a recollection of 150 subjects for the *Social Both CP* condition, which permits a clean treatment comparison on data from the same collection. Figure 6 summarizes the results. The left-hand panel shows that the *Deliberation* treatment substantially reduced participants' sensitivity to the input comparator, by about 46% (also see Table 4, column (1)). The middle panel documents that deliberation also reduced the effect of the output comparator, by 41%. The right-hand panel suggests that, at the same time, deliberation significantly increased sensitivity to the actual decision input (the subject's own offered workload), by 55%. This set of results is consistent with our hypothesis: if deliberation reduces mapping uncertainty, the strength of comparison effects decreases (and behavioral attenuation, as in Enke et al. (2024), decreases as well).

Reducing uncertainty through a familiar currency. We conducted a second, alternative intervention to reduce mapping uncertainty. We hypothesized that while translating the abstract features of an unfamiliar real effort task (such as transcribing 30 Greek letters) into money is unfamiliar, translating a more common "currency" such as the expected completion time into a monetary valuation may be more familiar and thus reduce uncertainty. In the *Time* Treatment, we thus additionally displayed the expected total duration of a workload. The results of this treatment, summarized in column (2) of Table 4, mirror those of the *Deliberation* treatment: expressing workloads in a second, more familiar currency both decreases sensitivity to the input and output comparators (by 44% and 42%, respectively), and it increases the sensitivity to the workload itself (by 68%). Details are provided in Appendix D.

6.2 Reducing Comparator Informativeness

As discussed in Section 2, we embrace the idea that comparison points may be used as information either because they are actually perceived to be informative or—more heuristically—because people may have become adapted to somewhat-informative comparison points. Still, on the margin, the combination of these two considerations predicts that comparison points should matter less when they are known to be uninformative for the decision at hand. This is not the case in our treatments involving expectations-based comparison points reported above. The reason is that in those treatments the comparison points are not randomly drawn from the universe of possible values but, rather, from a set of values the experimenter deems reasonable. For example, the default workload and the default payment in the *Expectations Both CP* treatment could be perceived by subjects as providing information about what the experimenter thinks a 'reasonable' reward and workload in our experiment is.

Design. We design an experiment in which we vary, across subjects, whether the expectations-based comparison point is saliently uninformative. In one treatment, we replicate the *Expectations Both CP* treatment, in which we provide no information to subjects about how the default workload and the default payment (that determine their outcomes with 50% chance) are determined. In practice, the comparison points are drawn from other subjects' decisions.

In the *Informativeness* treatment, we instead transparently inform subjects that the default workload and default payment are determined at random. Specifically, the instructions clarify that "you will get a randomly drawn, arbitrary compensation in return for completing a randomly drawn, arbitrary workload. These two random numbers were determined arbitrarily and have no meaning for your decision, and you have no influence on them."

Results. We pre-registered a collection of 150 subjects for the *Informativeness* treatment alongside a recollection of 150 subjects for the *Expectations Both CP* condition. Column (3) of Table 4

Table 4: Regression results for manipulations of mapping uncertainty and informativeness

Paradigm:	**Effort valuation**		
Dependent variable:	Minimum acceptable compensation (WTA)		
Treatment:	*Deliberation & Both CP (social)*	*Time & Both CP (social)*	*Random & Both CP (expectations)*
	(1)	(2)	(3)
Input (workload)	0.234***	0.244***	0.192***
	(0.0167)	(0.0112)	(0.0171)
Input comparator (workload)	-0.118***	-0.142***	-0.127***
	(0.0119)	(0.0121)	(0.0149)
Output comparator (WTA)	0.340***	0.326***	0.325***
	(0.0250)	(0.0170)	(0.0253)
Deliberation=1 × Input (workload)	0.129***		
	(0.0210)		
Deliberation=1 × Input comparator (workload)	0.0543***		
	(0.0140)		
Deliberation=1 × Output comparator (WTA)	-0.139***		
	(0.0289)		
Time displayed=1 × Input (workload)		0.167***	
		(0.0158)	
Time displayed=1 × Input comparator (workload)		0.0624***	
		(0.0145)	
Time displayed=1 × Output comparator (WTA)		-0.136***	
		(0.0204)	
1 if random treatment=1 × Input (workload)			-0.0378
			(0.0236)
1 if random treatment=1 × Input comparator (workload)			0.0877***
			(0.0214)
1 if random treatment=1 × Output comparator (WTA)			-0.106***
			(0.0340)
Observations	6744	8850	4456
Subjects	450	590	298
R-squared	0.774	0.729	0.685

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects and task-type fixed effects.

summarizes the results. In line with our predictions, we find that the informativeness manipulation reduced sensitivity to input comparators by 69% and to output comparators by 33%, in line with an account of comparisons as information. At the same time, we still find sizable and significant comparison effects in the *Informativeness* treatment. As noted above, one potential explanation for this is that, in everyday environments, it is uncommon to be directly presented with entirely uninformative comparison points, which may lead people to heuristically still make use of them. Again, what matters here from the perspective of our hypothesis is the input-vs.-output comparator taxonomy: when the comparison points are transparently uninformative, subjects still exhibit contrast to the input comparator and assimilation to the output comparator, thus acting *as if* they used them as information.

7 Classifying Prior Studies

To investigate the broader explanatory power of the input-output taxonomy, we revisit prior field observational and experimental studies on comparison effects.

7.1 Classification Approach

Sample. We assembled a sample of papers via a keyword search on *Web of Science*, looking at ten leading economics journals between 1970 and 2025.[4] We search for the keywords 'reference point', 'reference dependence', 'reference-dependent', 'comparison effect', 'aspiration level', 'framing', 'loss aversion', 'frame dependence', 'context dependence', 'context-dependent', 'contrast effect', 'assimilation effect', 'endowment effect', 'status quo', 'anchor' and 'anchoring' in the paper's title and abstract.[5] This search returned 596 papers. Our reliance on keyword-based search means that whether a paper enters the sample depends partly on whether the paper uses language such as "reference point" or "anchor." We then manually screened each paper to determine whether it contains classifiable empirical evidence on a comparator effect.

Inclusion criteria. Our main inclusion criterion is that we only consider papers in which the proposed comparison point exhibits some variation. This variation need not be randomly assigned. However, this criterion excludes several papers in the literature that, for example, estimate reference-dependent preferences relative to a reference point of $0 based on binary lottery choice data.[6]

We take papers at face value with respect to their own framing of what constitutes the comparison point and how it affects behavior; we do not screen for potential confounds such as omitted variables.

We exclude a paper if any of the following holds: *(i)* it is a pure theory paper; *(ii)* the comparison point is not measured, observed or hypothesized to be a specific quantity but is, instead, estimated from behavior through a structural model; *(iii)* the paper tests a comparator effect but finds no significant result or a non-monotonic effect; *(iv)* the manipulation varies the choice options themselves—i.e. what we would characterize as decision inputs—rather than a comparison point; and *(v)* the variation is not in the magnitude of the comparison point but in its existence (e.g., a reference point is induced vs. not induced) or in its type (e.g. more or less frequently observed comparisons as in myopic loss aversion studies).

This leaves us with a final sample of 102 studies. This sample exhibits large variation across contexts. Some are lab experiments, some field experiments and some rely on observational data. Some use fairly abstract designs, others more applied ones. Some are focused on identi-

[4]The journals include *American Economic Review, Quarterly Journal of Economics, Econometrica, Journal of Political Economy, Review of Economic Studies, Review of Economics and Statistics, Management Science, American Economic Journal: Microeconomics, Games and Economic Behavior and Journal of the European Economic Association*.

[5]We do not include information provision experiments that study belief updating. Our paper is thus complementary to Coffman et al. (2024) who study assimilation and contrast effects in a model of information nudges.

[6]We *do* include papers that implicitly amount to a fixed reference point of zero when the comparison quantity itself exhibits variation (for example, when decision makers are hypothesized to have a reference point of final earnings of zero but exhibit different starting balances).

fying a specific behavioral effect, while others focus on applied or policy questions in finance, public economics or development economics. As a result, the comparison points in our final sample also exhibit large variation across papers. They include, for example, earnings reference points, purchase prices of owned assets, personal best scores, ownership expectations, statutory retirement ages, peer wages, the quality of a previously-evaluated candidate, and par in golf. While many of the results we classify are often explained using prospect-theoretic explanations, others do not admit an interpretation in terms of reference-dependent utility functions (or are at least not framed as such by the authors).

Classification of comparator type and effect direction. For each classifiable paper, we identify four objects: the relevant decision input (e.g. hourly wage), the main decision (e.g. labor supply), associated outcomes (e.g. total earnings), and the comparison point(s) analyzed by the authors (e.g. expected earnings). We classify whether the comparison point relates to a decision input or output (decision or outcome), and in which direction it moves the decision.

In classifying whether a variable is an input or output, it is important to keep in mind that the same variable can be either an input or an output depending on the decision at hand. The key feature is whether, for the decision the analyst is interested in, the variable is taken as given by the decision maker (in which case it is an input), or whether he implicitly or explicitly chooses it (in which case it is an output).

For input comparators, we record contrast when the comparator moves the decision in the opposite direction as the main input; assimilation when they move decisions in the same direction. For output comparators, assimilation is recorded when the decision or outcome moves toward the comparator; contrast when it moves away. We then record whether the finding is consistent with the taxonomy, i.e. input comparators producing contrast and output comparators producing assimilation.

7.2 Caveats and Limitations

First, our classification is inherently partly subjective in nature. Papers are rarely designed with our taxonomy in mind, and the boundary between a comparator and a decision input is sometimes ambiguous. When multiple readings are defensible, we follow the paper's own framing, but reasonable researchers could disagree.

Second, the inclusion criteria spelled out above imply that several papers that concern comparison points (or are framed as such) are excluded from our analysis, for example because there is no variation in a posited comparison point.

Third, excluding null results is a clear limitation because it may bias the share of findings that are consistent with our taxonomy.

Fourth, unlike in our experiments in which we observe both the relevant input comparator and output comparator, the vast majority of papers in the literature studies (or posits) only one comparison point. This is potentially problematic because to the extent that both input and output comparators are present, they will often be correlated. As discussed in Appendix C, this introduces a potential omitted variables problem because input and output comparators will often be correlated. In our classification, we ignore such potential omitted variables and only classify the comparison point posited by the paper's authors. This effectively (and we believe sensibly) assumes that the researchers have information about which comparison point is most salient in the decision context of interest. At the same time, the existence of this potential omitted variables problem also serves to illustrate the usefulness of controlled experiments that separately measure and vary both input and output comparators.

Finally, we emphasize that the literature classification only tests the input-vs.-output-comparator taxonomy as such, rather than the idea of comparison points as sources of information for uncertain decision makers.

7.3 Results

Appendix Table F1 reports the full annotated classification of all 102 classifiable papers. According to our classification, 26 papers study an input comparator and all of them report a contrast effect. 77 papers study an output comparator, 76 of which report an assimilation effect.[7] The results thus closely align with our taxonomy also outside the context of our own experiments.

Table 5 illustrates by placing several prominent papers in the literature into what we believe are instructive—but certainly not exhaustive—categories of comparison effects. First are papers on targets such as income targets or goals in sports. These papers generally feature a comparison point in outcome space (e.g. total accumulated earnings), and the typical finding is that higher comparison outcomes produce higher outcomes—an assimilation effect.

Second are papers that study comparison points given by expected or past attributes, such as the piece rate of a peer (compared with one's own piece rate) or the expected ease of a task (compared with its actual ease). These are input comparators—the decision is how much to work, which depends on one's piece rate and the ease of the task. Consistent with our taxonomy, these papers find that a higher comparison point in input space produces a contrast effect (e.g. lower labor supply). This resolves one of the 'puzzles' raised in the Introduction—why labor supply increases in reference earnings but decreases in reference piece rates.

[7] Note that one of the 102 papers separately reports both an input comparator and output comparator effect, so that the total number of documented effects here sums to 103.

Third, several papers study sequential evaluation contexts, for example in sequential candidate interviewing, sequential decisions by judges or sequential ER admission decisions by physicians. According to our taxonomy, what matters here is whether the paper studies the effect of the previous decision (an output comparator) or of the attributes of the previous case (an input comparator). For instance, Jin et al. (2024) find that physicians' ER admission decision increases in whether they admitted the previous case—an assimilation to an output comparator (see Bindler and Hjalmarsson, 2019, for very similar evidence in the context of jury verdicts). In contrast, Radbruch and Schiprowski (2025) look at whether the evaluations of interviewers increase or decrease in the noisily measured quality of the previous candidate (rather than in their own past decision for the previous candidate). They find a contrast effect with respect to this input comparator, as predicted by our taxonomy.

The only paper in our sample that yields findings inconsistent with our predictions is a part of this category: the negative autocorrelation in sequential decisions by judges and loan officers documented by Chen et al. (2016). This result implies a contrast effect with respect to an output comparator (previous decisions). We speculate that this result may be driven by the omitted variables problem discussed above, i.e., that the researchers perfectly observe past decisions but do not perfectly observe the attributes of the previous case. Because past decisions and past case attributes will be strongly correlated, we conjecture (without evidence) that the true effect is driven by the relevant input comparator ('This file looks worse than the previous file') rather than the output comparator ('I said yes on the previous file, and things need to revert, so I will say no on this one').

A fourth category of papers likewise illustrates how the same context can give rise to either assimilation or contrast effects, purely depending on whether the comparison point relates to an input or output of the decision problem. The probability with which the owner of an asset (e.g. a house) decides to sell it decreases in the asset's purchase price. This contrast effect pertains to an input-based comparator because the input for the selling decision is the asset's current value. Compare this setting with the associated decision on the intensive margin: the listing price for the asset increases in the past purchase price. This is an assimilation effect predicted by our taxonomy because the purchase price is now an output comparator for the listing price.

Fifth are papers on the endowment effect such as exchange asymmetries. In these settings, the decision is which of two options to choose, the inputs are the attributes of the two choice options, and the outcome of the decision is ownership at the end of the experiment. The comparison point is induced ownership at the beginning of the experiment, an output comparator (regarding the outcome). Consistent with our taxonomy, the endowment effect is an assimilation effect in output space.[8]

[8]The same holds true for papers that do not measure deterministic endowments but expected ownership. In

Table 5: Examples of classifications of prior studies

Phenomenon	Input	Output (decision)	Output (outcome)	Comparison point	Type of CP	Typical result	CP effect	Examples
Targets / goals:								
Income targeting	Wage / piece rate	Labor supply	Earnings	Typical / expected earnings	Output	Labor supply increases in typical / expected earnings	Assimilation	Thakral and Tô (2021), Abeler et al. (2011)
Tax targeting	Receipts / expenses	Effort to find receipts	Reported tax liability	Withholding amount	Output	Reported liability clusters at withholding amount	Assimilation	Rees-Jones (2018)
Sports goals	Game conditions	Effort supply	Score (golf score, marathon time)	Target score (golf par, marathon goal)	Output	More effort below CP / excess mass at CP	Assimilation	Pope and Schweitzer (2011), Allen et al. (2017)
Past / expected attributes:								
Comparison wage	Wage	Labor supply	Earnings	Peer's wage	Input	Labor supply decreases in peer's wage	Contrast	Gächter and Thöni (2010); Breza et al. (2018)
Expected attribute	Ease of task	Labor supply	Earnings	Expected task ease	Input	Labor supply decreases in expected task ease	Contrast	Bushong and Gagnon-Bartsch (2023)
Sequential evaluations:								
Prior attributes	Candidate quality	Candidate rating	n/a	Quality of prior candidate	Input	Rating decreases in prior quality	Contrast	Radbruch and Schiprowski (2025)
Prior decision	Patient condition	Y/N admit to ER	Patient outcome	Admission decision prior patient	Output	Prob. admit increases in previous admission	Assimilation	Jin et al. (2024)
Past prices paid:								
Extensive margin	Current asset value (stock, house)	Keep / sell	Ownership	Purchase price	Input	Prob. keep decreases in purchase price	Contrast	Andersen et al. (2022)
Intensive margin	Current asset value	Listing price	Sale / price	Purchase price	Output	Listing price increases in purchase price	Assimilation	Genesove and Mayer (2001), Andersen et al. (2022)
Endowment effect:								
Exchange asymmetry	Attributes of goods	Choose good	Ex-post ownership	Ex-ante (expected) ownership	Output	Prob choosing increases in (expected) ownership	Assimilation	Kahneman et al. (1990)
'Normal' decisions:								
Defaults	Pension, benefits	Retirement age	Payouts	Statutory pension thresholds	Output	Retirement age increases in threshold	Assimilation	Seibold (2021), Behaghel and Blau (2012)

8 Forecaster Survey: Measuring Priors About Our Findings

DellaVigna et al. (2019) advocate for measuring the profession's prior knowledge because "Once a paper is presented, priors are typically lost as hindsight bias [...] makes it all too easy to see a result as one that 'we knew already'." Because we view the (ex-post perhaps 'obvious') input-output-comparator taxonomy as one of the contributions of this paper, we obtain a measurement of priors about through a forecaster survey on the *Social Science Prediction Platform* (SSPP).

Design. The survey comprised three blocks. In the first block, we define and explain what comparison points are, casting a wide net for what constitutes a comparison point ("[...] such as reference points, goals, anchors, norms, expectations, social comparisons etc."). We further explain the direction of a comparison effect based on the sign of the coefficient in a corresponding regression. We then ask participants: "Think about all the different possible comparison points you can imagine. Do you have a general rule for predicting the comparison effect in a regression of decisions on the decision parameter and the comparison point? Put differently, do you have a general prediction for which types of comparison points create a positive, negative, or no effect?" Participants could answer "Yes" or "No". Conditional on answering "Yes", we asked the follow-up question "What is your rule for predicting whether a comparison effect creates a positive, negative or no effect?" with an open entry text box. The full screen is reproduced in Appendix Figure 51.

In the second block we complemented this abstract elicitation of a general taxonomy with predictions about our concrete experimental settings. We described one of our experimental paradigms and showed the corresponding decision screen participants saw in that experiment. We then showed the forecasters our basic regression specification, the sample size, and the sign of the effect of the main decision input. We asked the forecaster to predict the sign(s) of the effect of the comparator(s) by choosing from {significantly positive, significantly negative, not significantly different from zero, I don't know}. Participants made predictions for four different experimental paradigms: *Effort Valuation* (either the Input CP (social) or Output CP (social) condition, randomly assigned), *Effort Choice* (Both CP (expectations) condition), *Beliefs* (Both CP (social) condition), and *Investment* (Both CP (expectations) condition).

In the third and final block, we again showed the prediction task from the first block, asking for whether the forecaster now has a general rule, and if so, what it is. A full set of screenshots of the forecaster survey are provided in Appendix H.3.

these studies, researchers exogenously induce a probability with which a subject will own a good, finding that a higher probability of owning increases the probability of choosing the option. Again, this comparison point of expected ownership lives in output space.

Logistics and Sample. The survey was fielded between October 1 and November 15, 2025, on the SSPP with Study ID sspp-2025-0043-v1. We received a total of 105 complete responses. Our sample consists of social science researchers: 90% have at least a Master's degree, and 94% work in economics. The sample skews junior: 61% are PhD or graduate students, 19% are non-faculty researchers (including postdocs), and 14% are professors or faculty. Geographically, 48% are based in the United States, 28% in Europe, and the remaining 24% elsewhere.

Results. In their first prediction, 24.8% of the forecasters indicated that they have a general rule for predicting the sign of comparison effects. A manual analysis of these 26 rules showed that none corresponded to our taxonomy of input and output comparators. The most common rule types were appeals to loss aversion or prospect theory (e.g., "comparisons that frame matters in terms of giving something up will have a differently signed effect from comparisons that frame matters in terms of gains"), predictions of a single sign without a structural basis (e.g., "an anchor typically creates a positive effect"), and domain-specific intuitions (e.g., "people are averse to feeling as though they're getting an unusually bad deal").

In the second block eliciting predictions of comparison effects in concrete experimental settings, forecasters predicted the correct sign of the *output* comparator 51.4% of the time. For *input* comparators, only 19.8% of predictions were correct—below the 33% that would be expected from random guessing across the three signed options, and consistent with a systematic bias toward predicting a positive effect even when the correct answer is negative. Appendix Table E1 reports the full breakdown by paradigm and comparator type. The asymmetry is large and statistically significant ($\chi^2 = 91.3$, $p < 0.001$). The pattern holds across all four paradigms. The worst performance is on the beliefs paradigm, where the input comparators are the prior and likelihood ratio from another subject's updating problem—both correctly negative—yet only 6–12% of forecasters predicted a negative sign.

In the third block, 31.4% of forecasters indicated that they now had a general rule for predicting comparison effects, up from 24.8% before making the concrete predictions. Yet none of the 33 rules provided in block three corresponded to our taxonomy.

We interpret this set of results as suggesting that even though social science forecasters correctly predict the sign of output comparator effects about half the time, the input-vs.-output taxonomy was not part of the profession's conceptual toolkit prior to this paper.

9 Discussion

Reference-dependent preferences. Our account does not stand in contradiction to the existence of reference-dependent preferences.[9] To take a trivial example, adaptation effects such as that lukewarm water feels more pleasant after cold water represent clear cases in which sensations and utility depend on reference points. At the same time, our classification of prior studies suggests that not all of the evidence that is frequently interpreted through the lens of models of reference-dependent preferences represents utility-related effects rather than various form of decision uncertainty.

In this section, we discuss which of the results in this paper are potentially compatible with canonical models of reference-dependent preferences, and which ones can more clearly be attributed to an information channel. First, several of our results are not predicted by models of reference-dependent preferences. This includes the evidence on the objective belief formation task and the comparative statics results that the comparison effects weaken when the decision problem involves less uncertainty or when the comparisons are less informative.

On the other hand, we suspect that some of our other results on input and output comparators could be fit by unusual variants of models of reference-dependent preferences (Kahneman and Tversky, 1979; Kőszegi and Rabin, 2006). In their standard formulations, these models are defined over utility outcomes. As such, especially for the rational-expectations formulation in Kőszegi and Rabin (2006), it is not obvious that the model would make opposite predictions for contexts in which a comparison point is given by an input (e.g., a piece rate) or an output (e.g., total earnings).[10] Yet, while standard formulations of such models may not immediately generate the effects we document, we suspect that unusual variants of them might. For example, it may be that input comparators directly enter the utility function. For instance, it may be that people experience higher effort costs when the piece rate at which they work is below a comparison piece rate, or that the utility they derive from an investment return is lower if the comparison return is higher. We have not attempted to write down and solve such models, and again, they would not naturally produce comparative statics on decision uncertainty and comparator informativeness.

A cookbook for applications. Leveraging our taxonomy for making predictions in empirical applications requires researchers to take a stand on what is an input or an output in their context of interest. A main challenge is that the same variable can be either an input or an

[9]We are grateful to Botond Kőszegi for comments and discussions on this topic.

[10]Empirical analyses of reference dependence have sometimes even mechanically *equated* inputs and outputs. For instance, in the literature on the labor supply of cab drivers, the hourly wage of the driver gets backed out by the researcher by dividing total earnings by the number of hours worked (Camerer et al., 1997; Farber, 2008). Yet, as our paper shows, reference earnings and reference wages have psychologically very different effects.

output for different decisions, see Section 5. We reiterate that inputs are variables that are taken as given by the decision maker, for the decision the researcher studies. Outputs are variables that are implicitly or explicitly controlled by the decision the researcher studies. Accordingly, input comparators are variables that are related or similar to inputs, and output comparators are related or similar to outputs. We recognize that this isn't a formal definition, yet we also believe that—in both theorizing and empirical applications—economists routinely take a stance on which variables are taken as given by the decision maker and which ones are implicitly or explicitly chosen. For instance, whenever a formal model motivates the analysis of an empirical setting, it is usually straightforward to declare inputs and outputs and, as a result, input and output comparators.

Conclusion. Given the importance and pervasiveness of relative judgments, comparison points are one of the central objects of interest in behavioral economics. Yet in empirical applications it is often difficult to predict even the direction of comparison effects. The mixed results include multiple settings of economic interest, including labor supply, candidate interviewing, house purchases, criminal sentencing, and others. Sometimes, these observed opposite effects are framed as distinct or even mutually exclusive phenomena.

This paper has argued that, at least in part, contrast and assimilation effects are not "either-or" phenomena, but that both are simultaneously at play when people find it difficult to map problem primitives into a decision. In our account, comparison points provide information about which decision is sensible, and the informational content of input and output comparators is different.

We believe that our results may be helpful going forward in two interrelated ways. First, as illustrated by our classification of the prior literature, the distinction between input and output comparators may help in predicting the sign of comparison effects in empirical applications. Second, our mechanism evidence speaks to recent models of reference points as information, which predict that reference effects are stronger when the decision maker exhibits more pronounced uncertainty about how to value objects or actions (Villas-Boas, 2024; Dean et al., 2026).

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–492. 10.1257/aer.101.2.470.

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic choice and preferences for randomization." *Journal of Political Economy* 125 (1): 40–68.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science* 63 (6): 1657–1672. 10.1287/mnsc.2015.2417.

Altmann, Steffen, Armin Falk, and Andreas Grunewald. 2025. "Communicating through defaults." *Review of Economics and Statistics* 107 (1): 256–268.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai. 2022. "Reference Dependence in the Housing Markets." *American Economic Review* 112 (10): 3398–3440. 10.1257/aer.20191766.

Arieli, Amos, Yaniv Ben-Ami, and Ariel Rubinstein. 2011. "Tracking decision makers under uncertainty." *American Economic Journal: Microeconomics* 3 (4): 68–76.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. ""Coherent arbitrariness": Stable demand curves without stable preferences." *The Quarterly Journal of Economics* 118 (1): 73–106.

Augenblick, Ned, Eben Lazarus, and Michael Thaler. 2025. "Overinference from weak signals and underinference from strong signals." *The Quarterly Journal of Economics* 140 (1): 335–401.

Barberis, Nicholas C. 2013. "Thirty years of prospect theory in economics: A review and assessment." *Journal of economic perspectives* 27 (1): 173–196.

Bastianello, Francesca, and Alex Imas. 2025. "Biases in Belief Updating Within and Across Domains." *Working Paper*.

Behaghel, Luc, and David M. Blau. 2012. "Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age." *American Economic Journal: Economic Policy* 4 (4): 41–67.

Bhargava, Saurabh, and Ray Fisman. 2014. "Contrast effects in sequential decisions: Evidence from speed dating." *Review of Economics and Statistics* 96 (3): 444–457.

Bhui, Rahul, and Yang Xiang. 2021. "A rational account of the repulsion effect."

Bindler, Anna, and Randi Hjalmarsson. 2019. "Path Dependency in Jury Decision Making." *Journal of the European Economic Association* 19 (3): 1698–1744.

Bondi, Tommaso, Daniel Csaba, Evan Friedman, and Salvatore Nunnari. 2025. "Range Effects in Economic Choice: The Role of Complexity."Technical report, CESifo Working Paper.

Bordalo, Pedro, Nicola Gennaioli, Giacomo Lanzani, and Andrei Shleifer. 2025. "A cognitive theory of reasoning and choice."Technical report, National Bureau of Economic Research.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, attention, and choice." *The Quarterly Journal of Economics*.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality." *The Quarterly Journal of Economics* 133 (2): 611–663. 10.1093/qje/qjx041.

Bushong, Benjamin, and Tristan Gagnon-Bartsch. 2023. "Reference Dependence and Attribution Bias: Evidence from Real-Effort Experiments." *American Economic Journal: Microeconomics* 15 (2): 271–308. 10.1257/mic.20210031.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics* 112 (2): 407–441. 10.1162/003355397555153.

Charles, Constantin, Cary Frydman, and Mete Kilic. 2024. "Insensitive investors." *The Journal of Finance* 79 (4): 2473–2503.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision-Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *The Quarterly Journal of Economics* 131 (3): 1181–1242. 10.1093/qje/qjw017.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2002. "Defined contribution pensions: Plan rules, participant choices, and the path of least resistance." *Tax policy and the economy* 16 67–113.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush. 2020. "Using goals to motivate college students: Theory and evidence from field experiments." *Review of Economics and Statistics* 102 (4): 648–663.

Coffman, Lucas C, Clayton R Featherstone, and Judd B Kessler. 2024. "A model of information nudges." *Working Paper*.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. "Social comparison and effort provision: Evidence from a field experiment." *Journal of the European Economic Association* 12 (4): 877–898.

Dean, Mark, Benjamin Enke, Thomas Graeber, and Pietro Ortoleva. 2026. "Reference Points as Information." *Working Paper*.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt. 2019. "Predict science to improve science." *Science* 366 (6464): 428–429.

Dhar, Ravi, and Ning Zhu. 2006. "Up close and personal: Investor sophistication and the disposition effect." *Management Science* 52 (5): 726–740.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive uncertainty." *The Quarterly Journal of Economics* 138 (4): 2021–2067.

Enke, Benjamin, Thomas Graeber, Ryan Oprea, and Jeffrey Yang. 2024. "Behavioral attenuation."Technical report, National Bureau of Economic Research.

Farber, Henry S. 2008. "Reference-dependent preferences and labor supply: The case of New York City taxi drivers." *American Economic Review* 98 (3): 1069–1082.

Frydman, Cary, and Lawrence Jin. 2021. "Efficient Coding and Risky Choice." *Quarterly Journal of Economics*.

Gabaix, Xavier. 2019. "Behavioral inattention." In *Handbook of Behavioral Economics-Foundations and Applications 2*, edited by Bernheim, Douglas, Stefano DellaVigna, and David Laibson 261, Elsevier.

Gächter, Simon, and Christian Thöni. 2010. "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis." *Journal of Economic Behavior & Organization* 76 (3): 531–543.

Gallagher, Geoffrey K, and Christopher P Benton. 2022. "Stimulus uncertainty predicts serial dependence in orientation judgements." *Journal of Vision* 22 (1): 6–6.

Genesove, David, and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *The Quarterly Journal of Economics* 116 (4): 1233–1260. 10.1162/003355301753265561.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics* 112 (2): 631–645.

Grether, David M. 1980. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *Quarterly Journal of Economics* 95 537–557.

Hsee, Christopher K. 1996. "The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives." *Organizational behavior and human decision processes* 67 (3): 247–257.

Imas, Alex. 2016. "The realization effect: Risk-taking after realized versus paper losses." *American Economic Review* 106 (8): 2086–2109.

Imhoff, Roland, and Paul Barker. 2023. "Connecting the dots: Nonlinear patterns in the presence of symbolic and nonsymbolic numerical standards." *Judgment and Decision Making* 18 e34.

Jin, Lawrence, Rui Tang, Han Ye, Junjian Yi, and Songfa Zhong. 2024. "Path dependency in physician decisions." *Review of Economic Studies* 91 (5): 2916–2953.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1990. "Experimental tests of the endowment effect and the Coase theorem." *Journal of Political Economy* 98 (6): 1325–1348.

Kahneman, Daniel, and Amos Tversky. 1973. "On the psychology of prediction." *Psychological Review* 80 (4): 237–251.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–292.

Kamenica, Emir. 2008. "Contextual inference in markets: On the informational content of product lines." *American Economic Review* 98 (5): 2127–49.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2021. "Cognitive imprecision and small-stakes risk aversion." *The review of economic studies* 88 (4): 1979–2013.

Knetsch, Jack L. 1989. "The endowment effect and evidence of nonreversible indifference curves." *American Economic Review* 79 (5): 1277–1284.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." *The Quarterly Journal of Economics* 121 (4): 1133–1165.

List, John A. 2003. "Does market experience eliminate market anomalies?" *The quarterly journal of economics* 118 (1): 41–71.

List, John A. 2004. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica* 72 (2): 615–625.

Loewenstein, George F. 1988. "Frames of mind in intertemporal choice." *Management Science* 34 (2): 200–214.

Madrian, Brigitte C, and Dennis F Shea. 2001. "The power of suggestion: Inertia in 401 (k) participation and savings behavior." *The Quarterly journal of economics* 116 (4): 1149–1187.

Meyer, Andrew, and Sean Hundtofte. 2023. "The longshot bias is a context effect." *Management Science* 69 (11): 6954–6968.

Moon, Jongmin, and Oh-Sang Kwon. 2022. "Attractive and repulsive effects of sensory history concurrently shape visual perception." *BMC biology* 20 (1): 247.

Pope, Devin G., and Maurice E. Schweitzer. 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review* 101 (1): 129–157. 10.1257/aer.101.1.129.

Pope, Devin G., and Uri Simonsohn. 2011. "Round Numbers as Goals: Evidence from Baseball, SAT Takers, and the Lab." *Psychological Science* 22 (1): 71–79. 10.1177/0956797610391098.

Radbruch, Jonas, and Amelie Schiprowski. 2025. "Interview sequences and the formation of subjective assessments." *Review of Economic Studies* 92 (2): 1226–1256.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–1278.

Sadil, Patrick, Rosemary A Cowell, and David E Huber. 2024. "The push–pull of serial dependence effects: Attraction to the prior response and repulsion from the prior stimulus." *Psychonomic Bulletin & Review* 31 (1): 259–273.

Seibold, Arthur. 2021. "Reference points for retirement behavior: Evidence from German pension discontinuities." *American Economic Review* 111 (4): 1126–1165.

Shubatt, Cassidy, and Jeffrey Yang. 2023. "Tradeoffs and Comparison Complexity." *Working Paper*.

Simonsohn, Uri, and George Loewenstein. 2006. "Mistake #37: The Effect of Previously Encountered Prices on Current Housing Demand." *The Economic Journal* 116 (508): 175–199. 10.1111/j.1468-0297.2006.01052.x.

Thakral, Neil, and Linh T. Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review* 111 (8): 2417–43. 10.1257/aer.20170768.

Thaler, Richard H. 2016. "Behavioral economics: Past, present, and future." *American Economic Review* 106 (7): 1577–1600.

Villas-Boas, J Miguel. 2024. "Toward an information-processing theory of loss aversion." *Marketing Science* 43 (3): 523–541.

Webb, Ryan, Paul W Glimcher, and Kenway Louie. 2021. "The normalization of consumer valuations: Context-dependent preferences from neurobiological constraints." *Management Science* 67 (1): 93–125.

Wedell, Douglas H, Susanne K Hicklin, and Laura O Smarandescu. 2007. "Contrasting models of assimilation and contrast." *Assimilation and contrast in social psychology* 45–74.

Woodford, Michael. 2020. "Modeling imprecision in perception, valuation, and choice." *Annual Review of Economics* 12 579–601.

Wu, Keyu. 2024. "Context-dependent perceptions and decision-making." *Working Paper*.

Yang, Jeffrey. 2023. "On the Decision-Relevance of Subjective Beliefs." *Working Paper*.

ONLINE APPENDIX

A Proofs and Derivations

A.1 Proof of Prediction 1

In the linear model, $a^\star(x,\theta) = \theta x$, so observing (a_c, x_c) reveals $\theta_c = a_c/x_c$ exactly. The DM chooses

$$a(x; a_c, x_c) = m(\theta_c) \cdot x, \qquad \text{where } m(\theta_c) := E[\theta \mid \theta_c],$$

with $m'(\theta_c) > 0$ by informativeness. Using $\theta_c = a_c/x_c$:

$$\frac{\partial a}{\partial a_c} = \frac{\partial a}{\partial \theta_c} \cdot \frac{\partial \theta_c}{\partial a_c} = m'(\theta_c) \cdot x \cdot \frac{1}{x_c} > 0,$$

since $m'(\theta_c) > 0$, $x > 0$, and $x_c > 0$. Likewise,

$$\frac{\partial a}{\partial x_c} = \frac{\partial a}{\partial \theta_c} \cdot \frac{\partial \theta_c}{\partial x_c} = m'(\theta_c) \cdot x \cdot \left(-\frac{a_c}{x_c^2}\right) < 0,$$

since $a_c, x_c > 0$ and $m'(\theta_c) > 0$. $\square$

A.2 Proof of Prediction 2

Suppose $\theta \sim N(\mu_\theta, \sigma_\theta^2)$ and $\theta_c = \theta + \epsilon$ with $\epsilon \sim N(0, \sigma_\epsilon^2)$ independent of θ. Standard Bayesian updating yields:

$$m(\theta_c) = E[\theta \mid \theta_c] = \mu_\theta + \lambda(\theta_c - \mu_\theta), \qquad \lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}.$$

Thus $m'(\theta_c) = \lambda \in (0,1)$.

From the proof of Prediction 1, the comparison effects are:

$$\frac{\partial a}{\partial a_c} = \lambda \cdot \frac{x}{x_c}, \qquad \frac{\partial a}{\partial x_c} = -\lambda \cdot \frac{a_c}{x_c^2} \cdot x.$$

(Part a) Mapping uncertainty. Taking the derivative of λ with respect to σ_θ^2:

$$\frac{\partial \lambda}{\partial \sigma_\theta^2} = \frac{\sigma_\epsilon^2}{(\sigma_\theta^2 + \sigma_\epsilon^2)^2} > 0.$$

Since both $|\partial a/\partial a_c|$ and $|\partial a/\partial x_c|$ are proportional to λ, they increase in σ_θ^2.

(Part b) Informativeness. Taking the derivative of λ with respect to σ_ϵ^2:

$$\frac{\partial \lambda}{\partial \sigma_\epsilon^2} = -\frac{\sigma_\theta^2}{(\sigma_\theta^2 + \sigma_\epsilon^2)^2} < 0.$$

Hence λ strictly decreases in σ_ϵ^2 and strictly increases in signal precision $1/\sigma_\epsilon^2$. Since both comparison effects are proportional to λ, their magnitudes inherit these properties. □

A.3 Extension: Only One Comparator Observed by Decision Maker

Let $\varepsilon_{z,w} := \frac{\partial \log z}{\partial \log w}$ denote the elasticity of z with respect to w.

Case 1: only the input comparator x_c is observed. When the DM observes only x_c, the conditional expectation is

$$\mathbb{E}[\theta \mid x_c] = \mu_\theta + \lambda\big(\mathbb{E}[\theta_c \mid x_c] - \mu_\theta\big), \qquad \text{where } \mathbb{E}[\theta_c \mid x_c] = \frac{1}{x_c}\mathbb{E}[a_c \mid x_c].$$

Note that the weight λ is the same as in the two-comparator case: it governs the informativeness of θ_c about θ, not how much of the comparison context is directly observed, and therefore remains unchanged regardless of which comparator the DM sees. The decision rule is

$$a(x; x_c) = \mathbb{E}[\theta \mid x_c]\, x,$$

so the sign of $\frac{\partial a}{\partial x_c}$ is determined by $\frac{\partial \mathbb{E}[\theta|x_c]}{\partial x_c}$. A sufficient condition for contrast is that the expected comparator output does not increase more than proportionally in x_c:

$$\frac{\partial \mathbb{E}[\theta \mid x_c]}{\partial x_c} < 0 \quad \Longleftrightarrow \quad \varepsilon_{\mathbb{E}[a_c|x_c], x_c} < 1.$$

Case 2: only the output comparator a_c is observed. When the DM observes only a_c, the conditional expectation is

$$\mathbb{E}[\theta \mid a_c] = \mu_\theta + \lambda\big(\mathbb{E}[\theta_c \mid a_c] - \mu_\theta\big), \qquad \text{where } \mathbb{E}[\theta_c \mid a_c] = a_c\, \mathbb{E}\left[\frac{1}{x_c} \,\middle|\, a_c\right].$$

The decision rule is

$$a(x; a_c) \;=\; \mathbb{E}[\theta \mid a_c]\, x,$$

so the sign of $\frac{\partial a}{\partial a_c}$ is determined by $\frac{\partial \mathbb{E}[\theta|a_c]}{\partial a_c}$. A sufficient condition for assimilation is that the expected inverse input does not decrease too steeply in a_c:

$$\frac{\partial \mathbb{E}[\theta \mid a_c]}{\partial a_c} > 0 \quad \Leftrightarrow \quad \varepsilon_{\mathbb{E}[1/x_c|a_c],a_c} > -1.$$

Remark (Only one comparator observed)**.** *In the linear benchmark:*

(a) *(*Input comparator only*) If the DM observes only x_c and the elasticity of $\mathbb{E}[a_c \mid x_c]$ with respect to x_c satisfies $\varepsilon_{\mathbb{E}[a_c|x_c],x_c} < 1$, then input comparators create contrast, $\frac{\partial a}{\partial x_c} < 0$.*

(b) *(*Output comparator only*) If the DM observes only a_c and the elasticity of $\mathbb{E}[1/x_c \mid a_c]$ with respect to a_c satisfies $\varepsilon_{\mathbb{E}[1/x_c|a_c],a_c} > -1$, then output comparators create assimilation, $\frac{\partial a}{\partial a_c} > 0$.*

We provide proofs for this remark in turn.

(Part a) Only input comparator x_c observed. Since $\theta_c = a_c/x_c$ and x_c is observed, $\mathbb{E}[\theta_c \mid x_c] = \frac{1}{x_c}\mathbb{E}[a_c \mid x_c]$. The posterior expectation of θ is

$$E[\theta \mid x_c] = \mu_\theta + \lambda\big(E[\theta_c \mid x_c] - \mu_\theta\big).$$

Since $\lambda > 0$, the sign of $\partial E[\theta \mid x_c]/\partial x_c$ equals the sign of $\partial E[\theta_c \mid x_c]/\partial x_c$. The decision rule is $a(x; x_c) = E[\theta \mid x_c] \cdot x$, so it suffices to sign the derivative of $E[\theta_c \mid x_c] = \frac{1}{x_c}E[a_c \mid x_c]$:

$$\frac{\partial}{\partial x_c}\left(\frac{E[a_c \mid x_c]}{x_c}\right) = \frac{x_c \cdot \frac{\partial E[a_c|x_c]}{\partial x_c} - E[a_c \mid x_c]}{x_c^2}.$$

This is negative if and only if

$$\frac{\partial E[a_c \mid x_c]}{\partial x_c} < \frac{E[a_c \mid x_c]}{x_c}.$$

In elasticity terms, defining $\varepsilon_{E[a_c|x_c],x_c} := \frac{\partial \log E[a_c|x_c]}{\partial \log x_c}$:

$$\frac{\partial a}{\partial x_c} < 0 \quad \Leftrightarrow \quad \varepsilon_{E[a_c|x_c],x_c} < 1.$$

The condition $\varepsilon_{E[a_c|x_c],x_c} < 1$ is sufficient for strict contrast.

(Part b) Only output comparator a_c observed. Since $\theta_c = a_c/x_c$ and a_c is observed, $\mathbb{E}[\theta_c \mid a_c] = a_c \cdot \mathbb{E}[1/x_c \mid a_c]$. The posterior expectation of θ is

$$E[\theta \mid a_c] = \mu_\theta + \lambda\big(E[\theta_c \mid a_c] - \mu_\theta\big).$$

Since $\lambda > 0$, the sign of $\partial E[\theta \mid a_c]/\partial a_c$ equals the sign of $\partial E[\theta_c \mid a_c]/\partial a_c$. The decision rule is $a(x; a_c) = E[\theta \mid a_c] \cdot x$, so it suffices to sign the derivative of $E[\theta_c \mid a_c] = a_c \cdot E[1/x_c \mid a_c]$:

$$\frac{\partial}{\partial a_c}\left(a_c \cdot E\left[\frac{1}{x_c}\,\middle|\, a_c\right]\right) = E\left[\frac{1}{x_c}\,\middle|\, a_c\right] + a_c \cdot \frac{\partial E[1/x_c \mid a_c]}{\partial a_c}.$$

This is positive if

$$E\left[\frac{1}{x_c}\,\middle|\, a_c\right] + a_c \cdot \frac{\partial E[1/x_c \mid a_c]}{\partial a_c} > 0.$$

Dividing by $E[1/x_c \mid a_c] > 0$ and rearranging:

$$1 + \frac{a_c}{E[1/x_c \mid a_c]} \cdot \frac{\partial E[1/x_c \mid a_c]}{\partial a_c} > 0.$$

In elasticity terms, defining $\varepsilon_{E[1/x_c|a_c],a_c} := \frac{\partial \log E[1/x_c|a_c]}{\partial \log a_c}$:

$$\frac{\partial a}{\partial a_c} > 0 \quad \Leftrightarrow \quad \varepsilon_{E[1/x_c|a_c],a_c} > -1.$$

The condition $\varepsilon_{E[1/x_c|a_c],a_c} > -1$ is sufficient for strict assimilation. □

A.4 Bunching at Reference Points

A distinct empirical phenomenon documented in prior work is *bunching*: excess mass of decisions at specific reference values, such as round numbers (4-hour marathon times), symbolic thresholds (.300 batting averages), or targets (daily earnings goals).[11]

As noted in the main text, our framework can accommodate bunching under some additional assumptions, though we reiterate that we don't view the stylized framework in this paper as a particularly natural formalization of how value uncertainty or decision uncertainty can generate bunching (see Dean et al., 2026, for an alternative model).

Bunching requires two additional ingredients: inattention to input differences and outcome targeting.

[11] See, e.g., Pope and Simonsohn (2011) on round numbers, Allen et al. (2017) on marathon times, Camerer et al. (1997) on taxi driver income targets.

Inattention to input differences. In the baseline model, the DM recognizes that their input x may differ from the comparison input x_c, and scales the comparison information accordingly. Suppose instead that the DM is inattentive to this difference.

Formally, let $\alpha \in [0,1]$ parameterize the degree of attention to input differences, where the DM behaves as if facing input

$$\tilde{x} = \alpha x + (1-\alpha) x_c. \tag{4}$$

Under full inattention ($\alpha = 0$) and high informativeness ($\lambda \to 1$), the DM's action becomes:

$$a = E[\theta \mid \theta_c] \cdot x_c \approx \theta_c \cdot x_c = a_c, \tag{5}$$

generating bunching at the action comparator a_c.

For intermediate values, the action is

$$a = \underbrace{(1-\lambda)\mu_\theta}_{\text{prior}} \cdot \tilde{x} + \lambda \cdot \frac{a_c}{x_c} \cdot \tilde{x}. \tag{6}$$

With $\alpha = 0$ (full inattention) and $\tilde{x} = x_c$:

$$a = (1-\lambda)\mu_\theta x_c + \lambda a_c. \tag{7}$$

This represents *partial bunching*: the action is a weighted average of the prior-based action $\mu_\theta x_c$ and the comparison action a_c, with weight λ on the latter. Full bunching at a_c obtains when $\lambda = 1$.

Outcome targeting. In many applications, the salient comparator is an outcome y_c rather than an action a_c. Marathon runners target finishing times; taxi drivers target daily earnings; batters are aware of batting average thresholds.

When the DM targets an outcome comparator y_c while being inattentive to input differences, bunching occurs in outcome space rather than action space. Consider the deterministic case $y = g(a,x)$ with g strictly monotonic in a. The DM who assumes $x = x_c$ and targets outcome y_c chooses

$$a = g^{-1}(y_c, x_c). \tag{8}$$

This generates bunching at outcome y_c, with all inattentive DMs achieving approximately the same outcome despite potentially different true inputs x.

Importantly, when DMs *do* attend to their own input $x \neq x_c$, they choose different actions to achieve the same outcome target:

$$a = g^{-1}(y_c, x). \tag{9}$$

This generates bunching in outcome space but *dispersion* in action space, conditional on inputs. For example, taxi drivers targeting earnings y_c with different hourly demand rates x would work different hours $a = y_c/x$ to achieve the same earnings.

Stochastic outcomes. When outcomes are stochastic, $y = g(a, x, \eta)$ with η a random variable, the DM cannot perfectly control the realized outcome. A DM targeting outcome y_c chooses a to solve

$$\max_a \mathbb{E}[u(y) \mid a, x] \quad \text{subject to} \quad y = g(a, x, \eta), \tag{10}$$

or, under quadratic loss around the target,

$$\min_a \mathbb{E}\left[(y - y_c)^2 \mid a, x\right]. \tag{11}$$

In the linear-normal case with $y = a \cdot x + \eta$ and $\eta \sim N(0, \sigma_\eta^2)$, targeting y_c in expectation yields $a = y_c/x$. Realized outcomes are then distributed as

$$y \sim N(y_c, \sigma_\eta^2). \tag{12}$$

This generates *imperfect bunching*: intended outcomes cluster at y_c, but realized outcomes are distributed around y_c with variance σ_η^2. This is consistent with the empirical pattern of excess mass *near* reference points rather than exact mass *at* them.

A.5 General (Nonlinear) Model

Consider a general decision problem with optimal action $a^\star(x, \theta) = f(x, \theta)$, where $f : \mathbb{R}_{++} \times \Theta \to \mathbb{R}_{++}$ is twice continuously differentiable with $f_x > 0$ and $f_\theta > 0$. The DM is uncertain about θ, holds a prior over it, and chooses

$$a(x; \text{info}) = f\big(x, E[\theta \mid \text{info}]\big).$$

The comparison context is generated by $a_c = f(x_c, \theta_c)$, with (x_c, a_c, θ_c) drawn from some joint distribution. We assume θ_c is informative about θ in the sense of posterior monotonicity: $\partial E[\theta \mid \theta_c]/\partial \theta_c > 0$.

Since $f_\theta(x_c, \theta_c) > 0$, for each fixed x_c the map $\theta_c \mapsto f(x_c, \theta_c)$ is invertible. Define $g(a_c; x_c)$ as the unique solution to $f(x_c,\, g(a_c; x_c)) = a_c$. By the inverse function theorem, g is continuously differentiable with

$$\frac{\partial g}{\partial a_c} = \frac{1}{f_\theta(x_c, \theta_c)} > 0 \qquad \text{and} \qquad \frac{\partial g}{\partial x_c} = -\frac{f_x(x_c, \theta_c)}{f_\theta(x_c, \theta_c)} < 0.$$

The regularity conditions referenced in the main text are:

1. **Smoothness:** $f(x, \theta)$ is twice continuously differentiable with $f_x > 0$ and $f_\theta > 0$.
2. **Invertibility:** For each x_c, the map $\theta_c \mapsto f(x_c, \theta_c)$ is invertible (guaranteed by $f_\theta > 0$ and appropriate domain restrictions).
3. **Informativeness:** θ_c is informative about θ in the sense that $E[\theta \mid \theta_c]$ is strictly increasing in θ_c. This is satisfied, for example, when $\theta_c = \theta + \epsilon$ with ϵ independent noise, or more generally when (θ, θ_c) are affiliated.
4. **Differentiability of posterior:** The posterior expectation $m(\theta_c) = E[\theta \mid \theta_c]$ is differentiable with $m'(\theta_c) > 0$. (This is used in the proof of Prediction 2.)

Conditions 1–3 suffice for Prediction 1 (via a monotonicity argument). The proof of Prediction 2 additionally requires condition 4. The derivative representation used in the proof of Prediction 1 below also invokes condition 4 for convenience, though a direct monotonicity argument using only Conditions 1–3 yields the same sign predictions. The proof of the one-comparator extension (Section A.5.3) requires further sufficient conditions on the joint distribution of (x_c, θ_c), stated there.

A.5.1 Proof of Prediction 1

When both comparators are observed, the DM infers $\theta_c = g(a_c; x_c)$ from the comparison context and chooses

$$a(x; a_c, x_c) = f\big(x, m(\theta_c)\big), \qquad \text{where } m(\theta_c) := E[\theta \mid \theta_c].$$

Assimilation in output comparator. By the chain rule:

$$\frac{\partial a}{\partial a_c} = f_\theta\big(x, m(\theta_c)\big) \cdot m'(\theta_c) \cdot \frac{\partial g}{\partial a_c}.$$

Since $f_\theta > 0$ (by smoothness), $m'(\theta_c) > 0$ (by informativeness), and $\partial g/\partial a_c > 0$ (by the inverse function theorem), we have

$$\frac{\partial a}{\partial a_c} > 0.$$

Contrast in input comparator. Similarly:

$$\frac{\partial a}{\partial x_c} = f_\theta\big(x, m(\theta_c)\big) \cdot m'(\theta_c) \cdot \frac{\partial g}{\partial x_c}.$$

Since $f_\theta > 0$, $m'(\theta_c) > 0$, and $\partial g/\partial x_c = -f_x/f_\theta < 0$, we have

$$\frac{\partial a}{\partial x_c} < 0. \quad \square$$

A.5.2 Prediction 2 in the General Model

Prediction 2 is proved for the linear-Gaussian benchmark in Section A.2. Here we formalize the conditions under which the same comparative statics extend to the general model.

Setup. Let $\theta \sim F_\theta$ with mean μ_θ and variance σ_θ^2. The comparison parameter θ_c is a signal about θ with conditional distribution $\theta_c \mid \theta \sim H(\cdot \mid \theta)$. Define the *informativeness* of the signal structure by how much posterior variance is reduced: higher informativeness means $\mathrm{Var}(\theta \mid \theta_c)$ is smaller on average. We maintain the regularity conditions stated at the beginning of this section, including differentiability of $m(\theta_c) = E[\theta \mid \theta_c]$ (condition 4).

From the proof of Prediction 1, the magnitudes of both comparison effects are proportional to $m'(\theta_c)$:

$$\frac{\partial a}{\partial a_c} = f_\theta \cdot m'(\theta_c) \cdot \frac{\partial g}{\partial a_c} \quad \text{and} \quad \left|\frac{\partial a}{\partial x_c}\right| = f_\theta \cdot m'(\theta_c) \cdot \left|\frac{\partial g}{\partial x_c}\right|.$$

Both effects are therefore amplified whenever $m'(\theta_c)$ increases. The two parts of Prediction 2 correspond to conditions under which this occurs.

(Part a) Mapping uncertainty. Compare two priors F_θ and $\tilde{F}_\theta$ with the same mean but $\tilde{\sigma}_\theta^2 > \sigma_\theta^2$ (a mean-preserving spread). Under conjugate signal structures, a higher prior variance σ_θ^2 increases the weight on the signal θ_c in the posterior mean, raising $m'(\theta_c)$ and thus amplifying both comparison effects. In the Gaussian case this is immediate: $m'(\theta_c) = \lambda = \sigma_\theta^2/(\sigma_\theta^2 + \sigma_\epsilon^2)$, which is strictly increasing in σ_θ^2.

(Part b) Informativeness. Compare two signal structures H and $\tilde{H}$ where $\tilde{H}$ is more informative in the Blackwell sense (or has lower noise variance in the Gaussian case). Under conjugate signal structures, higher informativeness increases the weight on θ_c in the posterior mean, raising $m'(\theta_c)$ and thus amplifying both comparison effects. In the Gaussian case: λ is strictly decreasing in signal noise σ_ϵ^2, so both effects increase in signal precision.

A.5.3 Proof of Predictions When Only One Comparator Observed

(Part a) Only input comparator x_c observed. Without observing a_c, the DM cannot invert to recover θ_c exactly. By the law of iterated expectations:

$$E[\theta \mid x_c] = E\big[E[\theta \mid \theta_c] \mid x_c\big] = E\big[m(\theta_c) \mid x_c\big].$$

The DM chooses $a(x; x_c) = f\big(x, E[\theta \mid x_c]\big)$, so

$$\frac{\partial a}{\partial x_c} = f_\theta\big(x, E[\theta \mid x_c]\big) \cdot \frac{\partial E[\theta \mid x_c]}{\partial x_c}.$$

Since $f_\theta > 0$, contrast ($\partial a/\partial x_c < 0$) holds if and only if $\partial E[\theta \mid x_c]/\partial x_c < 0$.

Sufficient condition: Suppose the joint distribution of (x_c, θ_c) is such that $\theta_c \mid x_c$ is stochastically decreasing in x_c in the sense of first-order stochastic dominance. That is, higher inputs are associated with lower mapping parameters. Then $E[\theta_c \mid x_c]$ is decreasing in x_c, and by informativeness, $E[\theta \mid x_c]$ is also decreasing in x_c.

A more primitive condition in the linear case: if $\varepsilon_{E[a_c|x_c],x_c} < 1$, then $E[\theta_c \mid x_c] = E[a_c \mid x_c]/x_c$ is decreasing in x_c.

(Part b) Only output comparator a_c observed. By the law of iterated expectations:

$$E[\theta \mid a_c] = E\big[m(\theta_c) \mid a_c\big].$$

The DM chooses $a(x; a_c) = f\big(x, E[\theta \mid a_c]\big)$, so

$$\frac{\partial a}{\partial a_c} = f_\theta\big(x, E[\theta \mid a_c]\big) \cdot \frac{\partial E[\theta \mid a_c]}{\partial a_c}.$$

Since $f_\theta > 0$, assimilation ($\partial a/\partial a_c > 0$) holds if and only if $\partial E[\theta \mid a_c]/\partial a_c > 0$.

Sufficient condition: Suppose θ_c is stochastically increasing in a_c conditional on the joint distribution. Since higher outputs are associated with higher mapping parameters, this is the natural case. Then $E[\theta_c \mid a_c]$ is increasing in a_c, and by informativeness, $E[\theta \mid a_c]$ is also increasing in a_c.

In the linear case, this reduces to $\varepsilon_{E[1/x_c|a_c],a_c} > -1$. $\square$

A.6 Omitted Variable Bias: Only One Comparator Observed by Analyst

When the DM observes both comparators, a first-order Taylor approximation of the true decision rule can be locally written as

$$a \;=\; \beta_0 + \beta_x x + \beta_{a_c} a_c + \beta_{x_c} x_c + \nu, \tag{13}$$

with $\beta_{a_c} > 0$ and $\beta_{x_c} < 0$ as implied by Prediction 1. In many applications, however, the analyst may only observe one of the two comparators and will therefore estimate a misspecified regression that omits the other. Because input and output comparators are typically positively correlated (higher inputs lead to higher outputs), this generates omitted-variable bias that can be strong enough to reverse the sign of the estimated comparator effect. The result is a direct application of the omitted variable bias formula and does not depend on the functional form of the model.

(Part a) Omitting x_c. If the analyst regresses a on (x, a_c) only, the OLS estimator for β_{a_c} is:

$$\tilde{\beta}_{a_c} = \beta_{a_c} + \beta_{x_c} \cdot \frac{\mathrm{Cov}(x_c, a_c \mid x)}{\mathrm{Var}(a_c \mid x)}.$$

Since $\beta_{x_c} < 0$ and $\mathrm{Cov}(x_c, a_c \mid x) > 0$, the bias term is negative. If $|\beta_{x_c}| \cdot \mathrm{Cov}(x_c, a_c \mid x)/\mathrm{Var}(a_c \mid x)$ is sufficiently large, we can have $\tilde{\beta}_{a_c} < 0$ even though the true effect β_{a_c} is positive.

(Part b) Omitting a_c. Analogously, if the analyst regresses a on (x, x_c) only:

$$\tilde{\beta}_{x_c} = \beta_{x_c} + \beta_{a_c} \cdot \frac{\mathrm{Cov}(a_c, x_c \mid x)}{\mathrm{Var}(x_c \mid x)}.$$

Since $\beta_{a_c} > 0$ and $\mathrm{Cov}(a_c, x_c \mid x) > 0$, the bias term is positive. If it is sufficiently large, we can have $\tilde{\beta}_{x_c} > 0$ even though $\beta_{x_c} < 0$.

In sum, when both comparators affect behavior but only one is observed by the analyst, standard omitted-variable bias can generate spurious evidence of contrast in output comparators or assimilation in input comparators. □

We illustrate this omitted variable bias empirically using deliberately misspecified regressions in our own experimental data in Appendix C.

B Additional Tables and Figures

Table B1: Pre-registration links by paradigm and treatment condition

Paradigm	Treatment condition	Pre-registration link
Effort Valuation Decision: WTA Input: workload	No CP	https://aspredicted.org/vqqm-8dc2.pdf
	Both CP (social)	https://aspredicted.org/vqqm-8dc2.pdf
	Input CP (social)	https://aspredicted.org/vqqm-8dc2.pdf
	Output CP (social)	https://aspredicted.org/vqqm-8dc2.pdf
	Both CP (expectations)	https://aspredicted.org/vqqm-8dc2.pdf
	Deliberation (social)	https://aspredicted.org/nn2gw7.pdf
	Time (social)	https://aspredicted.org/vqqm-8dc2.pdf
	Informativeness (expectations)	https://aspredicted.org/rj5er5.pdf
Investment Decision: amount wagered Input: winning chance	No CP	https://aspredicted.org/b5vr-p3hh.pdf
	Both CP (expectations)	https://aspredicted.org/b5vr-p3hh.pdf
Beliefs Decision: posterior belief Inputs: prior, accuracy, test result	No CP	https://aspredicted.org/zc5r-rt4s.pdf
	Both CP (social)	https://aspredicted.org/zc5r-rt4s.pdf
Effort Choice Decision: workload Input: piece rate	No CP	https://aspredicted.org/c298-6tbq.pdf
	Both CP (expectations)	https://aspredicted.org/c298-6tbq.pdf

Table B3: Regression results for *Effort Choice* paradigm

Paradigm:	**Effort choice**	
Dependent variable:	Chosen workload	
Treatment:	*No CP*	*Both CP (expectations)*
	(1)	(2)
Input (piece rate)	36.07***	38.38***
	(6.095)	(4.454)
Input comparator (piece rate)		-6.990**
		(2.814)
Output comparator (workload)		0.121***
		(0.0156)
Observations	2175	3690
Subjects	145	246
R-squared	0.687	0.655

Notes. Main pre-registered specification. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects and task-type fixed effects.

C The Problem of Misspecification

When both input and output comparators influence people's decisions, the analyst potentially encounters an omitted variables problem if he only observes one of the two comparison points in the data. The reason is that input and output comparators will often be correlated. For example, another person's chosen workload or labor supply might be higher when their hourly wage is higher. This is potentially problematic because it can create a problem of misspecification. Concretely, when (i) input and output comparator are positively correlated; (ii) the true effect of the input comparator is negative; and (iii) the analyst only observes the output comparator in the data, then the estimated coefficient of the output comparator will be biased downward, and can even have the wrong sign. A similar logic holds when the analyst only observes the input comparator.

We illustrate this omitted variable bias in our own data, in which we 'know' that both input and output comparator matter. In Table C1, we present regression analyses that are deliberately misspecified. In treatments in which subjects encountered both comparators, we only include one of them in the regression. In treatments in which subjects only encountered one comparator, we only include the comparison point that was not actually observed by subjects (and thus cannot have an effect).

The results effectively show that, due to omitted variable bias, anything goes. Relative to the 'correct' estimates from regressions in which the correct comparison points are included, the estimated coefficients change in a fairly unpredictable manner. Sometimes, a true effect is incorrectly estimated as a zero or even opposite effect. Similarly, a true zero effect is sometimes estimated as a significant positive or negative effect.

To illustrate, in the *Investment* experiment, the input comparator (the comparison winning chance) has a significant negative effect on investment levels when the output comparator (the comparison investment amount) is controlled for. Yet in a deliberately-misspecified regression in which the output comparator is omitted, the coefficient of the input comparator becomes positive (i.e., switches signs) and statistically insignificant. Similarly, in the *Effort Valuation* experiment in which only the output comparator is shown, we find that in misspecified regressions in which only the input comparator is included as independent variable, the estimated effect is positive and highly statistically significant, although it cannot plausibly have a 'true' effect. These are illustrations of classic omitted variable bias that arises when the researcher only observes one of two correlated comparison points.

Table C1: Results of deliberately-misspecified regressions

Paradigm:	Effort valuation							Investment			Beliefs		
Dependent variable:	Minimum acceptable compensation (WTA)							Amount wagered			Log posterior odds		
Treatment:	*Both CP (social)*			*Input CP only (social)*		*Output CP only (social)*		*Both CP (expectations)*			*Both CP (social)*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Input (workload)	0.209***	0.203***	0.158***	0.183***	0.183***	0.150***	0.147***						
	(0.0134)	(0.0149)	(0.0109)	(0.0113)	(0.0113)	(0.0110)	(0.0116)						
Input comparator (workload)	-0.107***	-0.0619***		-0.0302***			0.0221**						
	(0.0142)	(0.0151)		(0.00976)			(0.00946)						
Output comparator (WTA)	0.320***		0.310***		-0.00159	0.250***							
	(0.0175)		(0.0171)		(0.00942)	(0.0169)							
Input (winning chance)								8.230***	8.226***	8.234***			
								(0.290)	(0.290)	(0.290)			
Input comparator (winning chance)								-0.371**	0.177				
								(0.176)	(0.149)				
Output comparator (amount)								0.0642***		0.0402***			
								(0.0147)		(0.0125)			
Input (log prior odds)											0.403***	0.412***	0.377***
											(0.0297)	(0.0295)	(0.0276)
Input (log likelihood ratio)											0.773***	0.775***	0.773***
											(0.0319)	(0.0316)	(0.0319)
Input comparator (log prior)											-0.0678**	-0.00502	
											(0.0272)	(0.0259)	
Input comparator (log likelihood ratio)											-0.0439*	0.0895***	
											(0.0232)	(0.0185)	
Output comparator (log posterior)											0.155***		0.124***
											(0.0242)		(0.0184)
Observations	3660	3660	3660	3735	3735	3690	3690	3705	3705	3705	3610	3674	3610
Subjects	244	244	244	249	249	246	246	247	247	247	246	246	246
R-squared	0.694	0.602	0.687	0.735	0.733	0.665	0.605	0.682	0.680	0.682	0.550	0.542	0.549

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for *Effort valuation* also include task-type fixed effects.

D Additional Treatment: Reducing Uncertainty Through a Familiar Currency

Arguably, a main source of difficulty or uncertainty in determining a decision in the *Effort Valuation* paradigm is that translating the abstract features of an unfamiliar real effort task (such as transcribing 30 Greek letters) into money is unfamiliar. Our second treatment to reduce uncertainty consists of making the "exchange rate" between the input and the decision more familiar.

Design. Treatment *Time* is a variant of the *Social Both CP* treatment. Subjects are again asked to state their WTA for a specified workload, and we provide subjects with the expected total duration of the workload in each round.[12] Moreover, prior to the start of the actual experiment (but after subjects have read the instructions), we elicit each participant's non-incentivized WTA for working 10, 20, . . . , 70 minutes (without specifying a task type) to further familiarize them with their valuation of different work durations, see Appendix Figure 14 for a screenshot.

The idea behind this treatment manipulation is that subjects are likely less uncertain about their exchange rate between time and money (a somewhat familiar 'currency') than about their exchange rate between the number of real-effort tasks and money, in particular after they have been prodded to think about their WTA for a range of different completion times. Conceptually, we think of this manipulation as providing information that reduces subjects' internal 'mapping' uncertainty.

Results. Table 4 and Appendix Figure 8 summarize the results. We see that both the positive effect of the output comparator (the comparison compensation) and the negative effect of the input comparator (the comparison workload) weaken substantially in this treatment.[13] The uncertainty manipulation decreases the effect of the input comparator by 44% and that of the output comparator by 42%.

[12]For instance, subjects might be told that the workload is to count 6,360 table cells and that the predicted completion time is 44 minutes. See Appendix Figure 39 for a screenshot.

[13]Moreover, the treatment increases the effect size of the input parameter itself (the workload in a given round). This result is reminiscent of several results in the literature and shows that there is less behavioral attenuation when internal uncertainty is lower (e.g. Enke and Graeber, 2023; Enke et al., 2024).

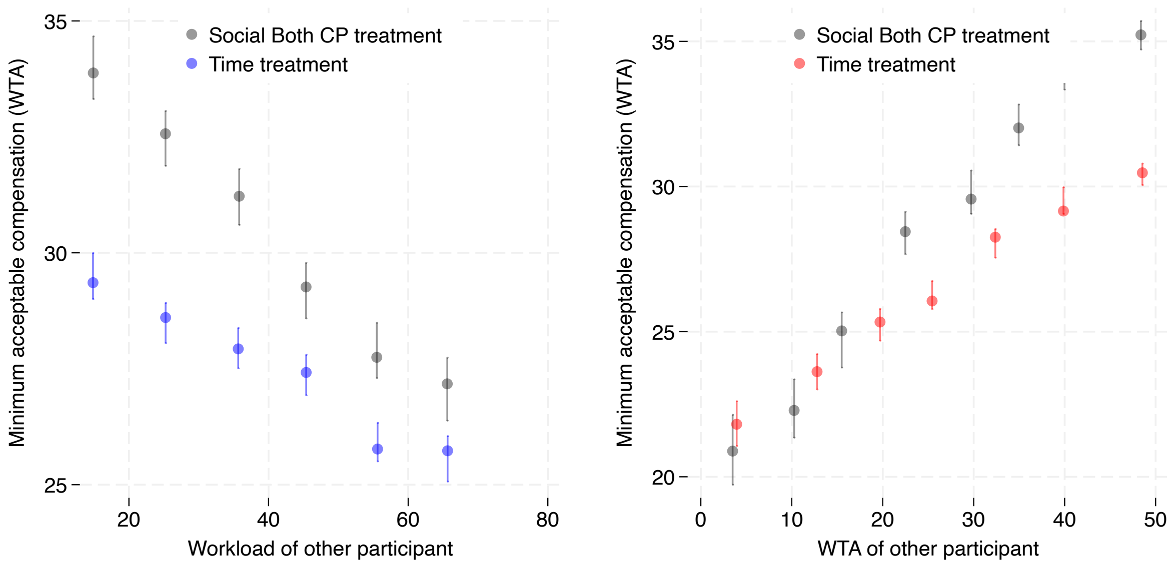


Figure 7: Results for treatment *Time* in the *Effort Valuation* paradigm. The left-hand side panels show average minimum acceptable compensations as a function of the input comparator (other subjects' workload). The right-hand side panel shows average minimum acceptable compensations as a function of the output comparator (other subjects' WTA).

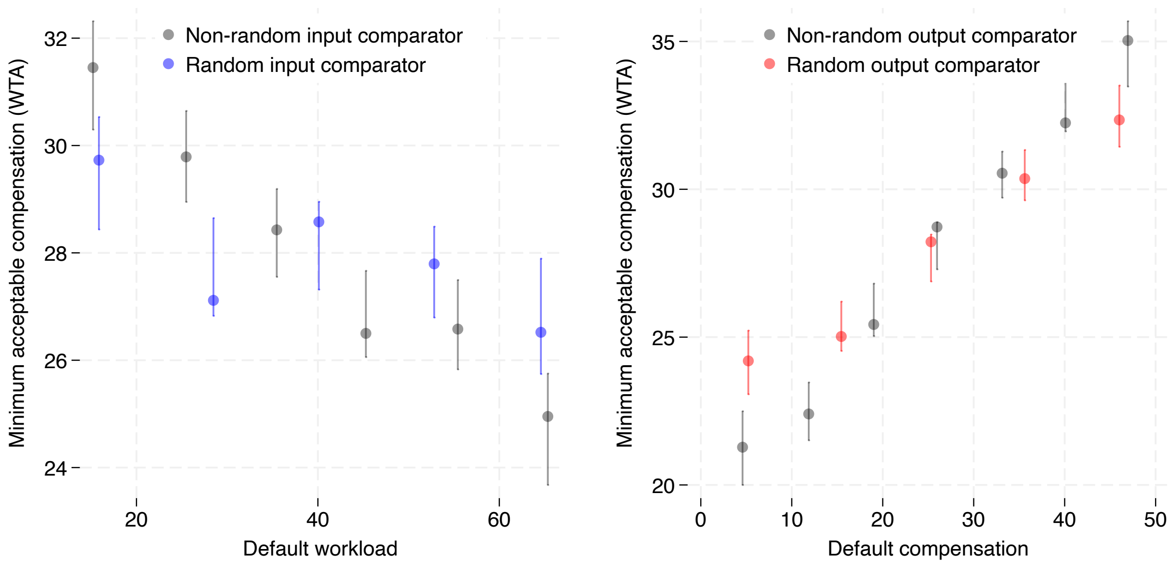


Figure 8: Results for treatment *Informativeness* in the *Effort Valuation* paradigm. The left-hand side panels show average minimum acceptable compensations as a function of the input comparator (other subjects' workload). The right-hand side panel shows average minimum acceptable compensations as a function of the output comparator (other subjects' WTA).

E Results of Forecaster Survey

Table E1: Forecaster prediction accuracy by paradigm and comparator type

Paradigm	Comparator	N	Share predicting (%)				Correct (%)
			Sig. pos.	Sig. neg.	Not sig.	Don't know	
Panel A: Output comparators							
Effort Valuation	WTA of other (output)	49	**77.6**	4.1	14.3	4.1	**77.6**
Effort Choice	Default rounds (output)	105	**52.4**	21.0	23.8	2.9	**52.4**
Investment	Amount wagered (output)	105	**40.0**	25.7	24.8	9.5	**40.0**
Beliefs	Others' log-odds (output)	105	**49.5**	1.9	32.4	16.2	**49.5**
All output comparators							**51.4**
Panel B: Input comparators							
Effort Valuation	Workload of other (input)	56	25.0	**41.1**	30.4	3.6	**41.1**
Effort Choice	Default piece rate (input)	105	46.7	**26.7**	21.9	4.8	**26.7**
Investment	Winning chance (input)	105	40.0	**21.9**	25.7	12.4	**21.9**
Beliefs	Others' prior (input)	105	37.1	**6.7**	41.0	15.2	**6.7**
Beliefs	Others' LLR (input)	105	26.7	**12.4**	44.8	16.2	**12.4**
All input comparators							**19.8**

Notes. Each row corresponds to one prediction slot in block 2 of the forecaster survey ($N = 105$ complete responses; Effort Valuation was randomly assigned, so N differs). Forecasters chose among {significantly positive, significantly negative, not significantly different from zero, I don't know}. Bold entries indicate the correct answer. The correct answer for all output comparators is significantly positive; for all input comparators it is significantly negative. Chance level is 33% if guessing randomly among the three signed options, or 25% if including the "don't know" option.

F Classification of Prior Studies

Table F1: Classification of prior studies

Paper	Relevant inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Abeler et al. (2011)	Piece rate and effort task characteristics	When to stop working	Earnings and minutes worked	Fixed-payment amount if decision doesn't count	Accumulated earnings at stopping increase in fixed payment	Output (outcome)	Assimilation	Consistent
Agnew et al. (2008)	Product attributes, risk/return characteristics	Choose annuity or investment option	Retirement portfolio (annuity vs. investment held)	Default option (pre-selected choice)	Choice of annuity increases when annuity is the default	Output	Assimilation	Consistent
Allen et al. (2017)	Race conditions, effort costs	Effort/pacing in final miles	Finishing time	Round-number finish-time goal (e.g., 3:00:00, 3:30:00)	Finishing times bunch just below round-number goals	Output (outcome)	Assimilation	Consistent
Anagol et al. (2018)	Information about stock value	Decision to hold IPO stock	Ownership of IPO stock	Ownership status (randomly assigned via IPO lottery)	Probability of holding increases in initial ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Andersen et al. (2022)	Current market value; property characteristics	Potential sale; price	Sale or not; sale price	Purchase price	(1) Likelihood of listing decreases in purchase price, (2) Listing price increases in purchase price	(1) Input, (2) Output (decision)	(1) Contrast, (2) Assimilation	Consistent
Ariely et al. (2003)	Product attributes	Stated valuation	Good ownership and payment made	Randomized numeric anchor prompt (SSN-based number)	Valuation increases in numeric anchor	Output (decision)	Assimilation	Consistent
Backus et al. (2022)	Won / lost past auction	Participate in another auction	Time until participate in auction again	Probability of winning past auction	Higher probability of winning past auction reduces probability of participating in another one (while actually winning past auction increases it)	Input	Contrast	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Baillon et al. (2020)	Lottery payoffs and probabilities	Choice between risky prospects	Lottery payoff received	Reference point (status quo = 0, or MaxMin = maximum of minimum outcomes across prospects)	Higher reference point decreases subjective valuation of a given lottery payoff	Input	Contrast	Consistent
Bartling and Schmidt (2015)[14]	Delivery day, seller costs, buyer value	Seller's price demand in renegotiation	Final (renegotiated) price	Initial contract price	Renegotiated price moves toward initial contract price	Output (decision)	Assimilation	Consistent
Bartling et al. (2015)	Match state (current score)	Effort intensity, aggression, substitution strategy	Cards received	Expected match state (from pre-play betting odds)	Higher expected match state (relative to current state) increases aggressive play and offensive substitutions	Input	Contrast	Consistent
Bateman et al. (2009)	Features of a hypothetical public goods project	Accept or reject project at stated cost	Project implementation	Prior cost amount in sequence	Higher prior cost increases acceptance at current cost	Input	Contrast	Consistent

Continued on next page

[14]We classify the within-treatment effect of the initial contract price level on renegotiated prices, not the between-treatment comparison of contract presence versus absence.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Bateman et al. (1997)	Consumption bundles, prices, budgets, and feasible consumption sets.	Valuation for bundles	Final consumption bundle held	Endowed bundle designated as current consumption state	Valuations shift toward endowed bundle	Output (outcome)	Assimilation	Consistent
Beggs and Graddy (2009)[15]	Painting quality/ characteristics	Expert pre-sale valuation	Auction outcome	Previous auction sale price	Current expert valuations increase in previous sale price	Output (decision)	Assimilation	Consistent
Behaghel and Blau (2012)	Financial position, health, pension benefit schedule	When to claim Social Security benefits	Retirement timing, benefit amount	Full Retirement Age (FRA)	Benefit claiming hazard spike moves in lockstep with FRA changes	Output (outcome)	Assimilation	Consistent
Bhargava and Fisman (2014)[16]	Current partner's attributes	(1) Accept vs reject current partner and (2) subjective rating of partner attractiveness	Potential match	Prior partner's attractiveness (a measure of "objective attribute quality")	Higher prior partner attractiveness decreases acceptance/rating of current partner	Input	Contrast	Consistent

Continued on next page

[15]The paper also studies pass through of the effect on final auction sale prices (and finds assimilation, there, too).

[16]Paper explicitly evaluates effect as contrast on objective attributes/stimuli. The measure of prior partner attractiveness could potentially be seen as an output comparator for the attractiveness evaluation of the current data–but note that the participants do not actually see the proxy for attractiveness used in

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Bindler and Hjalmarsson (2019)[17]	Current case evidence, defendant characteristics	Convict vs. acquit	Verdict	Previous convict/ acquit decision	Previous guilty verdict increases probability of subsequent guilty verdict	Output (decision)	Assimilation	Consistent
Blumenstock et al. (2018)	Financial considerations (income, savings needs, match incentives)	Decision to enroll/participate in mobile salary-linked savings account	Savings account enrollment	Default enrollment status	Enrollment increases in default enrollment status	Output	Assimilation	Consistent
Bordalo et al. (2016)	True distribution of types in target group	Belief about target group's type distribution		Reference group's type distribution	Higher reference group mean pushes beliefs about the target group down	Input	Contrast	Consistent
Brandts and Solà (2001)	Player 1's choice in strategic game	Player 2's response (punish or not)	Final game payoffs	Player 1's foregone alternative (what P1 could have chosen but didn't)	Better foregone alternative by Player 1 increases punishment by Player 2	Input	Contrast	Consistent

Continued on next page

the analysis; participants only see partner attributes.

[17]Judges might also use previous defendants' characteristics as input comparator, but not studied in this paper.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Breza et al. (2018)	Own wage (flat daily wage)	Effort / work intensity	Worker output (productivity) and attendance	Co-worker wages (same vs. different wages within production units)	Output decreases in co-worker wages	Input	Contrast	Consistent
Burson et al. (2013)[18]	Characteristics of goods	Valuation (WTA for sellers, WTP for buyers)	Final ownership (goods vs. money)	Assigned ownership	Assigned ownership increases valuation and final ownership	Output (outcome)	Assimilation	Consistent
Bushong and Gagnon-Bartsch (2023)	Task ease the design	Willingness to work (WTW)	Task completion / effort provision	Prior probability with which participant expected easy vs hard task rather than the other (certain vs 50-50 vs 0.99/0.01)	WTW decreases in expected task ease	Input	Contrast	Consistent

Continued on next page

[18]Also studies the effect of holding multiple units.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Cao et al. (2020)[19]	Piece rate in additional task	Work effort / participation in additional task	Productivity, work quality	Prior compensation per unit (lump-sum pay + gift, divided by copies completed in phase 1)	Higher gift amount (raising the reference wage) decreases effort on the lower-paid additional task	Input	Contrast	Consistent
Card and Dahl (2011)	Home team performance in game	Violent behavior	Domestic violence incidents	Expected home team performance from betting markets	Violence increases in expected performance	Input	Contrast	Consistent
Carrera et al. (2025)	The restaurant's quality rating	Whether or not to visit restaurant	Dining experience / restaurant visited	The consumer's quality reference point, formed from past experiences	Restaurant choice probability decreases in quality reference point	Input	Contrast	Consistent
Casey (1995)	Lottery characteristics	Valuation of a lottery ticket (WTA or WTP)	Ownership	Initially assigned ownership status	Valuations increase with initially assigned ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

[19] We classify the effect of gift amount on effort within a given gift type (e.g., Small Cash vs. Large Cash); the paper explicitly models the prior compensation (lump sum + gift) as a reference wage against which the lower piece rate in the additional task is evaluated.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Cerulli-Harms et al. (2019)[20]	Characteristics of goods	Keep vs. exchange	Final ownership	Expected ownership	Prob. of keeping / acquiring object increases in probability of forced ownership	Output (outcome)	Assimilation	Consistent
Chang and Kirgios (2024)	Candidate pool with candidates of different demographic identities	Choice of replacement candidate	Identity of new group member	Demographic identity of departing group member	Replacements disproportionately match the demographic identity of the departing member	Output	Assimilation	Consistent
Chang et al. (2019)[21]	Characteristics of dictator game	Tokens allocated to self	Final allocation to self	Dictator's initial endowment (0-10 tokens)	Higher dictator endowment increases tokens allocated to self	Output	Assimilation	Consistent
Chang et al. (2019)	Adjustment factor in a stock split, stock fundamentals	Ex-day stock valuation (following a stock split)	Ex-day market price/return	Cum-day price (pre-split price)	Ex-day price increases in cum-day price	Output (decision)	Assimilation	Consistent

Continued on next page

[20]We ignore the experiments with null results.
[21]Paper primarily studies an additional framing manipulation.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Chen et al. (2016)[22]	Case/application merits; pitch location	Grant/deny asylum; approve/ reject loan; call ball/strike	Case outcome (asylum granted, loan approved, pitch called)	Previous decision in the sequence	Current decision moves away from previous decision (negative autocorrelation)	Output (decision)	Contrast	Inconsistent
Clark et al. (2020)	Course material, practice exams, time constraints	Effort exertion (completing online practice exams, studying for course)	Task completion, course grade	Self-set goal	Performance increases in performance goal	Output (outcome)	Assimilation	Consistent
Cohn et al. (2014)[23]	Worker's own wage	Effort / work intensity	Productivity / work output	Peer's wage	Effort decreases in peer's wage	Input	Contrast	Consistent
Corgnet et al. (2018)	Task parameters, wage contract	Effort provision	Production	Goal (target production level)	Production increases in goal	Output (outcome)	Assimilation	Consistent

Continued on next page

[22]Prior case's true quality (input comparator) is unobserved and plausibly correlated with the lagged decision.
[23]We ignore treatment with null effect.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Corgnet et al. (2015)[24]	Piece rate	Effort level	Task performance (number of problems solved)	Exogenously assigned performance goal	Higher goal increases effort/performance (for attainable goals)	Output	Assimilation	Consistent
Crawford and Meng (2011)	Trip characteristics such as wage rate	When to stop working	Daily earnings and hours worked	Daily income and hours targets	Stopping probability decreases until income/hours reach target	Output (outcome)	Assimilation	Consistent
Czibor et al. (2022)[25]	Task, opportunities for theft/helping	Effort allocation (task, theft, helping)	Final wealth	Initial wealth	More stealing when initial wealth higher	Output (outcome)	Assimilation	Consistent
DellaVigna et al. (2017)	Current UI benefit level, labor market conditions	Search effort (job-finding rate)	Income / employment	Recent income (prior benefit level)	Job-finding rate increases when current / anticipated income falls below recent income	Output (outcome)	Assimilation	Consistent

Continued on next page

[24]We ignore null result for non-reasonable goals.
[25]We ignore the null results for other outcome variables.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Demeré et al. (2019)	Supervisor's initial rating of employee	Adjustment of rating (up/down/ none) by calibration committee	Final adjusted rating	Average rating of supervisor (across candidates)	Downward adjustment increases in average rating	Input	Contrast	Consistent
Dhar and Zhu (2006)[26]	Current price, stock characteristics, market conditions	Sell or hold stock	Stock ownership	Purchase price	Likelihood of selling generally increases in current price but decreases in purchase price	Input	Contrast	Consistent
Dufwenberg et al. (2011)	Characteristics of public goods game	Contribution level to public good	Payoffs from public goods game	Default contribution (give vs. take frame)	Contributions increase in default level	Output (decision)	Assimilation	Consistent
Dunn (1996)	Wage rate, job characteristics	WTW and WTA as a function of job amenities	Income-leisure bundle	Current income-leisure bundle (status quo, varies across workers)	Chosen outcomes cluster around the current income-leisure bundle	Output	Assimilation	Consistent

Continued on next page

[26] Paper focused on effect of sophistication on disposition effect.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Eil and Lien (2014)	Current session balance	Continue playing vs. quit	Final balance	Initial balance	Final balance assimilates to initial balance (zero profit); players quit when ahead and keep playing when behind	Output (outcome)	Assimilation	Consistent
Ericson and Fuster (2011)	Item characteristics	Trade/keep decision, WTA valuation	Final ownership	Expectation of ownership	Higher expected probability of owning increases chosen ownership	Output (outcome)	Assimilation	Consistent
Exley and Kessler (2024)[27]	Trivia question	Belief report	Belief	Random anchor (20 or 80)	Beliefs increase in anchor	Output (decision)	Assimilation	Consistent
Farber (2008)	Wage rate, hours available, conditions	When to stop driving	Daily income earned	Reference income level	Stopping probability increases as earnings approach income target	Output (outcome)	Assimilation	Consistent

Continued on next page

[27]Additional experiments reported in the paper; only the anchoring study qualifies as a comparator study.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Fehr et al. (2022)[28]	Attributes of goods	Keep vs. exchange	Ownership	Assigned initial ownership	Prob. keeping increases in assigned ownership	Output (outcome)	Assimilation	Consistent
Fehr and Tyran (2008)	Money supply, costs, demand conditions	Nominal price setting	Payoffs / firm profits	Pre-shock nominal price	Post-shock prices are pulled toward pre-shock price level (creating nominal inertia)	Output (decision)	Assimilation	Consistent
Ferraro et al. (2024)	Conservation contract characteristics, auction context	Bid submitted in procurement auction	Outcome of auction	Starting value of bid (0% vs. 100%)	Bids increase in starting bid	Output (decision)	Assimilation	Consistent
Fiedler and Hillenbrand (2020)	Payoff options for self and other	Choice between two options, one more selfish than the other	Final wealth of self and other	Initial wealth (endowment)	Final wealth increases in initial wealth (framing effect)	Output (outcome)	Assimilation	Consistent

Continued on next page

[28]We classify the baseline effect. Paper also studies effect of scarcity on size of endowment effect.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Fischer and Grosch (2025)	Realized interim payoff, contract terms	Contract breach decision	Final payoff	Buyer's expected payoff, partly determined by seller's performance	Bauyer's contract breach increases in seller's expected performance, conditional on seller's actual performance	Input	Contrast	Consistent
Freund and Özden (2008)[29]	World price of steel	Level of trade protection (tariff/ quota)	Domestic price / industry profits	Historical domestic price (reference point)	When world price falls below the reference, protection increases to maintain domestic price at the reference level	Output (outcome)	Assimilation	Consistent
Genesove and Mayer (2001)	Predicted current market value from hedonic controls (property attributes and time effects)	Listing price	Days on market and eventual sale price	Purchase price	Listing price increases in purchase price	Output (decision)	Assimilation	Consistent

Continued on next page

[29]No variation in comparison point

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
González-Díaz et al. (2024)	Opponent strength, game conditions	Effort / participation in tournaments	Elo chess rating, game outcome	Personal best (all-time max) Elo rating	Higher personal best pulls performance towards it (higher effort / better performance when close to personal best)	Output (outcome)	Assimilation	Consistent
Grolleau et al. (2016)	Task characteristics	Cheating	Final wealth	Initial wealth	Cheating and final wealth increase in initial wealth (framing manipulation)	Output (outcome)	Assimilation	Consistent
Grossman (2014)	Characteristics of dictator game	Information acquisition (reveal vs. remain ignorant)	Revealed information about payoffs	Default information state (ignorance vs. informed)	Default of no reveal increases ignorance	Output (decision)	Assimilation	Consistent
Haenni (2019)	Match outcome (win/loss)	Time to next tournament enrollment	Time to next tournament enrollment	Expected match outcome (proxied by opponent's ranking)	Losing to opponent with lower ranking increases delay to next signup	Input	Contrast	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Hartman et al. (1991)	Service reliability (outage frequency, duration), price (monthly bill)	Choice among reliability-rate contracts	Service reliability level, monthly bill	Current service reliability (status quo)	Customers disproportionately choose the contract matching their current reliability level; this holds for both high- and low-reliability status quo groups	Output (outcome)	Assimilation	Consistent
Hastings and Shapiro (2013)	Gasoline prices (by grade)	Gasoline grade choice	Gasoline grade and fuel expenditure	Recent habitual gas expenditure	When prices of (all grades of) gasoline rise, consumers downgrade octane to keep expenditure near their habitual budget.	Output	Assimilation	Consistent
Heffetz and List (2014)[30]	Characteristics of goods	Keep vs. exchange	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

[30] We ignore the null effect of the expectations manipulation.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Herz and Taubinsky (2018)	Proposer's offer in ultimatum game	Receiver's minimum acceptable offer	Ultimatum game payoffs	Previously experienced market offer level	Higher experienced offers increase minimum acceptable offers	Output (decision)	Assimilation	Consistent
Hossain and List (2012)	Characteristics of work at manufacturing facility	Effort	Output; final bonus	Initial bonus allocation (gain vs. loss frame)	Starting with bonus increases effort and likelihood of ending up with bonus	Output (outcome)	Assimilation	Consistent
Imas et al. (2017)[31]	Task characteristics	Effort provision	Earnings / final bonus	Initial bonus allocation (gain vs. loss framing)	Effort and final bonus increase in initial bonus allocation	Output (outcome)	Assimilation	Consistent
Iturbe-Ormaetxe et al. (2011)[32]	Characteristics of threshold public goods game	Whether or not to contribute to public good	Payoffs from public goods game	Default outcome (gain-loss framing manipulation): public good is provided vs. not provided	Contributions higher when default is that public good will be provided	Output (outcome)	Assimilation	Consistent

Continued on next page

[31] Paper also studies whether people prefer loss vs. gain contracts.
[32] We ignore effect in minimum threshold condition.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Jin et al. (2023)	Current patient's characteristics/ medical condition	Treatment decision for current patient (e.g., admit or not)	Patient outcomes	Physician's own previous treatment decision (for prior patient)	Likelihood of admitting increases in having admitted the previous patient	Output (decision)	Assimilation	Consistent
Kahneman et al. (1990)	Good characteristics	Whether to sell (if owner) or buy (if non-owner)	Final ownership of good	Initially assigned ownership	WTA exceeds WTP; final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent
Keys and Wang (2019)	Debt balance requiring repayment decision	Credit card debt repayment amount	Credit card debt repayment amount	Minimum payment amount (varies with formula changes)	Repayment amount increases in minimum payment amount	Output	Assimilation	Consistent
Kiessling et al. (2022)	Characteristics of task ("suicide run" in a PE class)	Effort (running)	Performance (time)	Peer effort	Performance regresses towards peer performance	Output (decision)	Assimilation	Consistent
Knetsch (1989)	Good characteristics	Keep vs. exchange	Final ownership	Initially assigned ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Lalive et al. (2023)	Age, financial incentives	When to claim pension / when to retire	Retirement age, pension benefit level	Full retirement age (FRA)	Higher FRA increases the chosen retirement/claiming age	Output	Assimilation	Consistent
Leib et al. (2021)[33]	Property attributes	Counteroffer	Price paid	Asking price	Counteroffers gravitate toward asking price	Output (decision)	Assimilation	Consistent
Lien and Zheng (2015)	Random slot machine outcomes	When to stop / at what earnings level to stop	Final balance	Initial balance	Spike at break-even point, implying that final balances assimilate towards initial balances	Output (outcome)	Assimilation	Consistent
List (2003)[34]	Attributes of goods	Keep vs. exchange	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

[33] We classify baseline effect; paper also studies effect of round vs. precise asking prices.
[34] Baseline endowment effect for inexperienced traders. Paper shows effect diminishes with market experience.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
List (2004)[35]	Attributes of goods	Keep vs. exchange	Ownership status	Initially assigned ownership	Prob. keeping increases in assigned ownership	Output (outcome)	Assimilation	Consistent
Loewenstein (1988)	Outcome timing, outcome sign (gain/loss), question frame	WTP / WTA for time-dated rewards	Final dated payoffs	Temporal status quo - which outcome (sooner or later) you currently "have"	Valuations move in the direction of keeping status quo	Output (outcome)	Assimilation	Consistent
Madrian and Shea (2001)	Income, financial situation	Retirement savings contribution rate	Retirement savings balance	Default contribution rate	Workers cluster around default contribution rates; higher defaults lead to higher contributions	Output	Assimilation	Consistent
Maniadis et al. (2014)	Aversive sound characteristics (duration, type)	Willingness-to-accept payment for listening to sounds	Payment and sounds listened to	Anchor price (high 50¢ vs. low 10¢)	Higher anchor increases WTA	Output (decision)	Assimilation	Consistent

Continued on next page

[35]We ignore the null result in the experts sample.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Mas (2006)	Actual wage received	Effort / work intensity	Police performance (arrests, clearance rates)	Reference wage (union demand)	Performance decreases in reference wage (union demand)	Input	Contrast	Consistent
Meyer and Hundtofte (2023)	Gamble payoffs, probabilities	Bet choice / gamble valuation	Monetary payoff	Contextual gambles' payoffs	Valuation of longshots increases when contextual gambles have lower payoffs	Input	Contrast	Consistent
Müller and Rau (2019)[36]	Lottery characteristics, own endowment	Lottery choice	Payoff/income	Peer's income	Higher peer income increases choice of lotteries with higher expected payoffs, moving expected income towards peer.	Output (outcome)	Assimilation	Consistent

Continued on next page

[36]Subjects who target a higher lottery upside in response to higher peer earnings also accept greater downside risk, meaning they may end up with less money than if they had chosen the safer option.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Plott and Zeiler (2007)[37]	Attributes of goods	Keep vs. exchange	Final ownership	Initially assigned ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent
Pope and Schweitzer (2011)[38]	Characteristics and conditions of current attempt	Putt attempt / effort on current shot	Total number of strokes on a golf hole	Par number for the present hole	Make probability higher for par putts than birdie putts	Output (outcome)	Assimilation	Consistent
Pope and Simonsohn (2011)[39]	Conditions/ characteristics of current choice, such as weather, field situation etc.	Effort/ aggressiveness in final plate appearance	Final batting average	.300 batting-average goal (round-number threshold)	Effort increases when batting average is just below .300 threshold	Output (outcome)	Assimilation	Consistent
Quispe-Torreblanca et al. (2025)	Current stock price / returns, other market conditions	Decision to sell or keep stock	Stock ownership	Purchase price; Price at last login	Likelihood of selling decreases in purchase price	Input	Contrast	Consistent

Continued on next page

[37] Baseline endowment effect with traditional procedures. Paper shows effect disappears when procedures are modified.
[38] Analysis does not exploit variation in par (uses par-fixed effects).
[39] No variation in threshold

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Radbruch and Schiprowski (2025)	Quality/attributes of current candidate	Interview rating/ assessment of current candidate	Scholarship/ hiring decision	Quality of previous candidate (measured as average of other evaluators' decisions)	Rating of current candidate decreases in previous candidate's quality	Input	Contrast	Consistent
Ramachandran et al. (2018)	Demand uncertainty, production cost structure	Selling price and quantity	Profit	Mean demand quantity for quantity, and cost of production for price	Quantity and price increase in their anchors	Output	Assimilation	Consistent
Ray et al. (2015)	Current price of focal product	Purchase decision for a product at online hardware retailer	Quantity purchased	Recent sale price of focal product	Purchase quantity decreases in recent sale price of focal product	Input	Contrast	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Rees-Jones (2018)	Inputs to the tax liability declaration such as receipts etc.	How much effort to put into finding deductions / when to stop declaring deductions	Tax liability declaration	Withholding amount	Reported tax liability clusters around withholding amount	Output (outcome)	Assimilation	Consistent
Riley et al. (2020)	Current price, stock characteristics	Selling price at which investor would feel neutral	None (hypothetical)	Salient past prices (purchase price, maximum, minimum, 52-week high/low)	Stated selling price increases in salient past prices	Output (decision)	Assimilation	Consistent
Rizzo and Zeckhauser (2003)	Current income, career stage, specialty, practice characteristics	Labor supply / effort	Subsequent income/ income growth	Self-reported target income	Future income increases in self-reported target income.	Output (outcome)	Assimilation	Consistent
Schurr and Ritov (2014)[40]	Characteristics of goods	Valuation (WTA or WTP)	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent

Continued on next page

[40]Also studies variation in the degree to which endowed goods are given up when endowed with multiple ones.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Schwerter (2024)[41]	Lottery payoffs, probabilities	Risk-taking level: choice between two lotteries with different upsides (and corresponding risk)	Lottery payoffs	Randomly determined peer earnings (not from lottery)	Chosen lottery upside increases in peer earnings	Output (decision)	Assimilation	Consistent
Seibold (2021)	Financial incentives (pension adjustment rates, benefit levels)	When to retire	Worker's retirement timing: age of benefit claiming/exit	Statutory pension thresholds (varies across cohorts due to reforms)	Retirement age increases in statutory pension threshold	Output	Assimilation	Consistent
Shelley (1993)	Outcome timing, outcome sign (gain/loss), question frame	Valuation of time-dated rewards	Final dated payoffs	Temporal status quo - which outcome (sooner or later) you currently "have"	Valuations move in the direction of keeping status quo	Output (outcome)	Assimilation	Consistent

Continued on next page

[41]Subjects who target a higher lottery upside in response to higher peer earnings also accept greater downside risk, meaning they may end up with less money than if they had chosen the safer option.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Simonsohn and Loewenstein (2006)	Housing options in destination city	Housing choice / how much to spend on rent	Rent/housing expenditure of movers to a new city	Origin city average housing prices	Housing expenditure increases in origin city prices	Output	Assimilation	Consistent
Simonsohn (2006)	Conditions and options in the new city	Where to live / job-housing location choice	Commute duration	Origin city average commute length	Commute length increases in origin city average commute	Output (outcome)	Assimilation	Consistent
Sprenger (2015)	Lottery payoffs and probabilities	Lottery valuations (certainty equivalents and probability equivalents)	Risk level of chosen option	Risk level of constant option in price list (certain vs. stochastic)	Final chosen risk increases in reference risk level	Output (outcome)	Assimilation	Consistent
Tereyağoğlu et al. (2018)	Ticket price (negative) and seat quality proxied by utilization (positive)	Purchase decision	Ticket obtained	Previous price and utilization	Demand increases in previous price and decreases in previous utilization	Input	Contrast	Consistent

Continued on next page

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Thakral and Tô (2021)	Market wage, hours worked, time/location characteristics	Stop vs. continue work	Labor supply / accumulated earnings at end of shift	Recent earnings target (income reference point)	Probability of stopping increases as earnings approach target	Output (outcome)	Assimilation	Consistent
Vihriälä (2025)	Debt balance, interest rate, available income (controlled for in analysis)	Credit card payment amount chosen by account holders	Credit card payment amount chosen by account holders	Minimum payment amount displayed on statement	Repayment amount increases in minimum payment amount	Output	Assimilation	Consistent
Weingarten et al. (2019)[42]	Actual outcome (e.g., grade received, exercises completed)	Emotional rating (happiness/ satisfaction)	Emotional rating (happiness/ satisfaction)	Goal on each dimension (e.g., target grade, target exercises)	Higher goal reduces happiness rating	Input	Contrast	Consistent

Continued on next page

[42]Goal classified as Input (not Output as in other goal papers) because the decision is a happiness judgment, not the goal-directed action itself.

Table F1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classification
Yu et al. (2021)[43]	Actual wait time, customers opportunity cost of time	Customer's patience decision in a call-center queue: abandon vs continue waiting	Final wait time	Announcement of estimated wait time	Higher announced wait time induces customers to abandon less	Input	Contrast	Consistent

[43]Classification of results from treatment with generally accurate delay information

References Classified Papers

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–492. 10.1257/aer.101.2.470.

Agnew, Julie R., Lisa R. Anderson, Jeffrey R. Gerlach, and Lisa R. Szykman. 2008. "Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults." *American Economic Review* 98 (2): 418–422.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science* 63 (6): 1657–1672. 10.1287/mnsc.2015.2417.

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai. 2018. "Endowment Effects in the Field: Evidence from India's IPO Lotteries." *Review of Economic Studies* 85 (4): 1971–2004.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai. 2022. "Reference Dependence in the Housing Markets." *American Economic Review* 112 (10): 3398–3440. 10.1257/aer.20191766.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. ""Coherent arbitrariness": Stable demand curves without stable preferences." *The Quarterly Journal of Economics* 118 (1): 73–106.

Backus, Matthew, Thomas Blake, Dimitriy Masterov, and Steven Tadelis. 2022. "Expectation, disappointment, and exit: evidence on reference point formation from an online marketplace." *Journal of the European Economic Association* 20 (1): 116–149.

Baillon, Aurélien, Han Bleichrodt, and Vitalie Spinu. 2020. "Searching for the Reference Point." *Management Science* 66 (1): 93–112.

Bartling, Björn, Leif Brandes, and Daniel Schunk. 2015. "Expectations as Reference Points: Field Evidence from Professional Soccer." *Management Science* 61 (11): 2646–2661. 10.1287/mnsc.2014.2048.

Bartling, Björn, and Klaus M. Schmidt. 2015. "Reference Points, Social Norms, and Fairness in Contract Renegotiations." *Journal of the European Economic Association* 13 (1): 98–129. 10.1111/jeea.12109.

Bateman, Ian J., Brett H. Day, Diane P. Dupont, and Stavros Georgiou. 2009. "Procedural Invariance Testing of the One-and-One-Half-Bound Dichotomous Choice Elicitation Method." *Review of Economics and Statistics* 91 (4): 806–820.

Bateman, Ian, Alistair Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. 1997. "A Test of the Theory of Reference-Dependent Preferences." *Quarterly Journal of Economics* 112 (2): 479–505.

Beggs, Alan, and Kathryn Graddy. 2009. "Anchoring Effects: Evidence from Art Auctions." *American Economic Review* 99 (3): 1027–1039.

Behaghel, Luc, and David M. Blau. 2012. "Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age." *American Economic Journal: Economic Policy* 4 (4): 41–67.

Bhargava, Saurabh, and Ray Fisman. 2014. "Contrast effects in sequential decisions: Evidence from speed dating." *Review of Economics and Statistics* 96 (3): 444–457.

Bindler, Anna, and Randi Hjalmarsson. 2019. "Path Dependency in Jury Decision Making." *Journal of the European Economic Association* 19 (3): 1698–1744.

Blumenstock, Joshua, Michael Callen, and Tarek Ghani. 2018. "Why Do Defaults Affect Behavior? Experimental Evidence from Afghanistan." *American Economic Review* 108 (10): 2868–2901.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics* 131 (4): 1753–1794.

Brandts, Jordi, and Carles Solà. 2001. "Reference Points and Negative Reciprocity in Simple Sequential Games." *Games and Economic Behavior* 36 (2): 138–157.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality." *The Quarterly Journal of Economics* 133 (2): 611–663. 10.1093/qje/qjx041.

Burson, Katherine A., David Faro, and Yuval Rottenstreich. 2013. "Multiple-Unit Holdings Yield Attenuated Endowment Effects." *Management Science* 59 (3): 545–555.

Bushong, Benjamin, and Tristan Gagnon-Bartsch. 2023. "Reference Dependence and Attribution Bias: Evidence from Real-Effort Experiments." *American Economic Journal: Microeconomics* 15 (2): 271–308. 10.1257/mic.20210031.

Cao, Cong J., Sherry X. Li, and Tracy Xiao Liu. 2020. "A Gift with Thoughtfulness: A Field Experiment on Work Incentives." *Games and Economic Behavior* 176 311–332.

Card, David, and Gordon B. Dahl. 2011. "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." *The Quarterly Journal of Economics* 126 (1): 103–143. 10.1093/qje/qjq029.

Carrera, Clara, Victor Martínez-de Albéniz, and Manuel E. Sosa. 2025. "The Bright Side of Lower Quality: Evidence from Restaurant Exploration." *Management Science* 71. 10.1287/mnsc.2022.04141.

Casey, Jeff T. 1995. "Predicting Buyer-Seller Pricing Disparities." *Management Science* 41 (6): 979–999.

Cerulli-Harms, Annette, Lorenz Goette, and Charles Sprenger. 2019. "Randomizing Endowments: An Experimental Study of Rational Expectations and Reference-Dependent Preferences." *American Economic Journal: Microeconomics* 11 (1): 185–207.

Chang, Daphne, Roy Chen, and Erin Krupka. 2019. "Rhetoric Matters: A Social Norms Explanation for the Anomaly of Framing." *Games and Economic Behavior* 116 158–178. 10.1016/j.geb.2019.04.011.

Chang, Elaine H., and Erika L. Kirgios. 2024. "Demographic Stickiness: The Demographic Identity of Departing Group Members Influences Who Is Chosen to Replace Them." *Management Science* 180 104296.

Chang, Eric C, Tse-Chun Lin, Yan Luo, and Jinjuan Ren. 2019. "Ex-day returns of stock distributions: An anchoring explanation." *Management Science* 65 (3): 1076–1095.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision-Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *The Quarterly Journal of Economics* 131 (3): 1181–1242. 10.1093/qje/qjw017.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush. 2020. "Using goals to motivate college students: Theory and evidence from field experiments." *Review of Economics and Statistics* 102 (4): 648–663.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. "Social comparison and effort provision: Evidence from a field experiment." *Journal of the European Economic Association* 12 (4): 877–898.

Corgnet, Brice, Joaquín Gómez-Miñambres, and Roberto Hernán-González. 2015. "Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough." *Management Science* 61 (12): 2926–2944.

Corgnet, Brice, Joaquín Gómez-Miñambres, and Roberto Hernán-González. 2018. "Goal Setting in the Principal-Agent Model: Weak Incentives for Strong Performance." *Games and Economic Behavior* 109 311–326.

Crawford, Vincent P., and Juanjuan Meng. 2011. "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." *American Economic Review* 101 (5): 1912–1932. 10.1257/aer.101.5.1912.

Czibor, Eszter, David Hsu, David Jimenez-Gomez, Susanne Neckermann, and Bedre-Defolie Subasi. 2022. "Loss-Framed Incentives and Employee (Mis-)Behavior." *Management Science* 68 (12): 8509–8532.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder. 2017. "Reference-Dependent Job Search: Evidence from Hungary." *The Quarterly Journal of Economics* 132 (4): 1969–2018. 10.1093/qje/qjx015.

Demeré, B. W., Karen L. Sedatole, and Amelia Woods. 2019. "The Role of Calibration Committees in Subjective Performance Evaluation Systems." *Management Science* 65 (4): 1562–1585.

Dhar, Ravi, and Ning Zhu. 2006. "Up close and personal: Investor sophistication and the disposition effect." *Management Science* 52 (5): 726–740.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. 2011. "The Framing of Games and the Psychology of Play." *Games and Economic Behavior* 73 (2): 459–478.

Dunn, Lucia F. 1996. "Loss Aversion and Adaptation in the Labor Market: Empirical Indifference Functions and Labor Supply." *Review of Economics and Statistics* 78 (3): 441–450.

Eil, David, and Jaimie W. Lien. 2014. "Staying Ahead and Getting Even: Risk Attitudes of Experienced Poker Players." *Games and Economic Behavior* 87 50–69.

Ericson, Keith Marzilli M., and Andreas Fuster. 2011. "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments." *Quarterly Journal of Economics* 126 (4): 1879–1907.

Exley, Christine L., and Judd B. Kessler. 2024. "Motivated Errors." *American Economic Review* 114 (4): 961–87. 10.1257/aer.20191849.

Farber, Henry S. 2008. "Reference-dependent preferences and labor supply: The case of New York City taxi drivers." *American Economic Review* 98 (3): 1069–1082.

Fehr, Dietmar, Günther Fink, and B. Kelsey Jack. 2022. "Poor and Rational: Decision-Making under Scarcity." *Journal of Political Economy* 130 (11): 2862–2897.

Fehr, Ernst, and Jean-Robert Tyran. 2008. "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia." *Econometrica* 76 (2): 353–394.

Ferraro, Paul J., Kent D. Messer, Pallavi Shukla, and Collin Weigel. 2024. "Behavioral Biases among Producers: Experimental Evidence of Anchoring in Procurement Auctions." *The Review of Economics and Statistics* 106 (5): 1381–1392. 10.1162/rest_a_01379.

Fiedler, Susann, and Adrian Hillenbrand. 2020. "Gain-Loss Framing in Interdependent Choice." *Games and Economic Behavior* 121 232–251.

Fischer, Simon, and Kerstin Grosch. 2025. "Contract Breach with Overconfident Expectations: Experimental Evidence on Reference-Dependent Preferences." *Games and Economic Behavior* 149 1–19.

Freund, Caroline, and Çağlar Özden. 2008. "Trade Policy and Loss Aversion." *American Economic Review* 98 (4): 1675–1691.

Genesove, David, and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *The Quarterly Journal of Economics* 116 (4): 1233–1260. 10.1162/003355301753265561.

González-Díaz, Julio, Ignacio Palacios-Huerta, and José Manuel Abuín. 2024. "Personal Bests and Gender." *Review of Economics and Statistics*, Forthcoming.

Grolleau, Gilles, Martin G. Kocher, and Angela Sutan. 2016. "Cheating and Loss Aversion: Do People Cheat More to Avoid a Loss?" *Management Science* 62 (12): 3428–3438.

Grossman, Zachary. 2014. "Strategic Ignorance and the Robustness of Social Preferences." *Management Science* 60 (11): 2659–2665.

Haenni, Simon. 2019. "Ever tried. Ever failed. No matter? On the demotivational effect of losing in repeated competitions." *Games and Economic Behavior* 115 346–362. 10.1016/j.geb.2019.03.012.

Hartman, Raymond S., Michael J. Doane, and Chi-Keung Woo. 1991. "Consumer Rationality and the Status Quo." *Quarterly Journal of Economics* 106 (1): 141–162.

Hastings, Justine S, and Jesse M Shapiro. 2013. "Fungibility and consumer choice: Evidence from commodity price shocks." *The Quarterly Journal of Economics* 128 (4): 1449–1498.

Heffetz, Ori, and John A. List. 2014. "Is the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association* 12 (5): 1396–1422.

Herz, Holger, and Dmitry Taubinsky. 2018. "What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views." *Journal of the European Economic Association* 16 (2): 316–352.

Hossain, Tanjim, and John A. List. 2012. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Science* 58 (12): 2151–2167.

Imas, Alex, Sally Sadoff, and Anya Samek. 2017. "Do People Anticipate Loss Aversion?" *Management Science* 63 (5): 1271–1284.

Iturbe-Ormaetxe, Iñigo, Giovanni Ponti, Josefa Tomás, and Luis Ubeda. 2011. "Framing Effects in Public Goods: Prospect Theory and Experimental Evidence." *Games and Economic Behavior* 72 (2): 439–447.

Jin, Li, Rui Tang, Heng Ye, Jessica J. Yi, and Songfa Zhong. 2023. "Path Dependency in Physician Decisions." *Review of Economic Studies*.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1990. "Experimental tests of the endowment effect and the Coase theorem." *Journal of Political Economy* 98 (6): 1325–1348.

Keys, Benjamin J., and Jialan Wang. 2019. "Minimum Payments and Debt Paydown in Consumer Credit Cards." *Journal of Financial Economics* 131 (3): 528–548.

Kiessling, Lukas, Jonas Radbruch, and Stefanie Schaube. 2022. "Self-Selection of Peers and Performance." *Management Science* 68 (11): 7793–7811.

Knetsch, Jack L. 1989. "The endowment effect and evidence of nonreversible indifference curves." *American Economic Review* 79 (5): 1277–1284.

Lalive, Rafael, Arvind Magesan, and Stefan Staubli. 2023. "How Social Security Reform Affects Retirement and Pension Claiming." *American Economic Journal: Economic Policy* 15 (3): 115–150. 10.1257/pol.20200686.

Leib, M., N. C. Köbis, M. Francke, S. Shalvi, and M. Roskes. 2021. "Precision in a Seller's Market: Round Asking Prices Lead to Higher Counteroffers and Selling Prices." *Management Science* 67 (10): 6061–6072. 10.1287/mnsc.2020.3789.

Lien, Jaimie W., and Jie Zheng. 2015. "Deciding When to Quit: Reference-Dependence over Slot Machine Outcomes." *American Economic Review* 105 (5): 366–370.

List, John A. 2003. "Does market experience eliminate market anomalies?" *The quarterly journal of economics* 118 (1): 41–71.

List, John A. 2004. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica* 72 (2): 615–625.

Loewenstein, George F. 1988. "Frames of mind in intertemporal choice." *Management Science* 34 (2): 200–214.

Madrian, Brigitte C., and Dennis F. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116 (4): 1149–1187.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *American Economic Review* 104 (1): 277–290.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *Quarterly Journal of Economics* 121 (3): 783–821.

Meyer, Andrew, and Sean Hundtofte. 2023. "The longshot bias is a context effect." *Management Science* 69 (11): 6954–6968.

Müller, Stephan, and Holger A. Rau. 2019. "Decisions under Uncertainty in Social Contexts." *Games and Economic Behavior* 116 73–95.

Plott, Charles R., and Kathryn Zeiler. 2007. "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?" *American Economic Review* 97 (4): 1449–1466.

Pope, Devin G., and Maurice E. Schweitzer. 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review* 101 (1): 129–157. 10.1257/aer.101.1.129.

Pope, Devin G., and Uri Simonsohn. 2011. "Round Numbers as Goals: Evidence from Baseball, SAT Takers, and the Lab." *Psychological Science* 22 (1): 71–79. 10.1177/0956797610391098.

Quispe-Torreblanca, Edika, John Gathergood, George Loewenstein, and Neil Stewart. 2025. "Investor Logins and the Disposition Effect." *Management Science*, Forthcoming.

Radbruch, Jonas, and Amelie Schiprowski. 2025. "Interview sequences and the formation of subjective assessments." *Review of Economic Studies* 92 (2): 1226–1256.

Ramachandran, Karthik, Necati Tereyağoğlu, and Yue Stella Xia. 2018. "Multidimensional Decision Making in Operations: An Experimental Investigation of Joint Pricing and Quantity Decisions." *Management Science* 64 (7): 2977–2989.

Ray, Debajyoti, Matthew Shum, and Colin F. Camerer. 2015. "Loss Aversion in Post-Sale Purchases of Consumer Products and Their Substitutes." *American Economic Review* 105 (5): 376–380.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–1278.

Riley, Ciara, Barbara Summers, and Darren Duxbury. 2020. "Capital Gains Overhang with a Dynamic Reference Point." *Management Science* 180 200–220.

Rizzo, John A., and Richard J. Zeckhauser. 2003. "Reference Incomes, Loss Aversion, and Physician Behavior." *Review of Economics and Statistics* 85 (4): 909–922.

Schurr, Amos, and Ilana Ritov. 2014. "The Effect of Giving It All Up on Valuation: A New Look at the Endowment Effect." *Management Science* 60 (3): 628–637.

Schwerter, Frederik. 2024. "Social Reference Points and Risk Taking." *Management Science*, Forthcoming.

Seibold, Arthur. 2021. "Reference points for retirement behavior: Evidence from German pension discontinuities." *American Economic Review* 111 (4): 1126–1165.

Shelley, Marjorie K. 1993. "Outcome Signs, Question Frames and Discount Rates." *Management Science* 39 (7): 806–815.

Simonsohn, Uri. 2006. "New Yorkers Commute More Everywhere: Contrast Effects in the Field." *The Review of Economics and Statistics* 88 (1): 1–9. 10.1162/rest.2006.88.1.1.

Simonsohn, Uri, and George Loewenstein. 2006. "Mistake #37: The Effect of Previously Encountered Prices on Current Housing Demand." *The Economic Journal* 116 (508): 175–199. 10.1111/j.1468-0297.2006.01052.x.

Sprenger, Charles. 2015. "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points." *Journal of Political Economy* 123 (6): 1456–1499. 10.1086/683836.

Tereyağoğlu, Necati, Peter S. Fader, and Senthil Veeraraghavan. 2018. "Multiattribute loss aversion and reference dependence: Evidence from the performing arts industry." *Management Science* 64 (1): 421–436.

Thakral, Neil, and Linh T. Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review* 111 (8): 2417–43. 10.1257/aer.20170768.

Vihriälä, Erkki. 2025. "Intra-Household Frictions, Anchoring, and the Credit Card Debt Puzzle." *Review of Economics and Statistics*.

Weingarten, Evan, Sudeep Bhatia, and Barbara Mellers. 2019. "Multiple Goals as Reference Points: One Failure Makes Everything Else Feel Worse." *Management Science* 65 (10): 4338–4358.

Yu, Qi Pu, Gad Allon, and Achal Bassamboo. 2021. "The Reference Effect of Delay Announcements: A Field Experiment." *Management Science* 67 (12): 7417–7437.

G Robustness Checks and Sensitivity Analyses

All experiments reported in this paper were pre-registered prior to data collection; Appendix Table B1 provides the corresponding links. Each pre-registration specifies a set of robustness checks and secondary analyses, which we report in this appendix. The analyses fall into three categories. First, we examine the sensitivity of our results to alternative sample restrictions that address data quality concerns: requiring the strictest human-verification threshold (reCaptcha score equal to 1), excluding rounds with implausibly fast or slow response times, and excluding participants who self-reported using external help (Section G.1, Tables G1–G5). Second, we assess robustness to the estimation method by replacing OLS with median (quantile) regressions, which are less sensitive to outliers (Section G.2, Tables G6–G7). Third, we report pre-registered secondary analyses that explore heterogeneity along theoretically relevant dimensions: whether the effects evolve over the course of the experiment (Section G.3, Tables G8–G9), whether comparison effects differ depending on whether the comparison is "favorable" or "unfavorable" (Section G.4, Table G10), and additional specifications for the effort choice paradigm and mechanism treatments (Sections G.5–G.6, Tables G11–G14).

Across the board, the conclusions from our main specifications remain unchanged: output comparators produce assimilation effects and input comparators produce contrast effects, with coefficients that are quantitatively similar to those reported in the main text. The pattern is consistent across all four experimental paradigms, all mechanism treatments, and all robustness and sensitivity specifications reported below. The secondary analyses further show that the comparison effects do not meaningfully evolve over the course of the experiment—nearly all round interaction terms are small and statistically insignificant (Tables G8–G9)—and that the direction of the effects does not depend on whether the comparison point is "favorable" or "unfavorable": both input contrast and output assimilation effects appear in both subsamples with similar magnitudes (Table G10).

G.1 Sample restrictions

Sample restrictions: Strict human verification (reCaptcha score = 1), Response-time validity (no speeders or overlong rounds), Protocol compliance (no external help).

Table G1: Robustness (Strict human verification)

Paradigm:	Effort valuation					Investment		Beliefs	
Dependent variable:	Minimum acceptable compensation (WTA)					Amount wagered		Log posterior odds	
Treatment:	*No CP*	*Both CP (social)*	*Both CP (expectations)*	*Output CP only (social)*	*Input CP only (social)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (social)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input (workload)	0.154***	0.245***	0.182***	0.181***	0.155***				
	(0.0172)	(0.0133)	(0.0150)	(0.0130)	(0.0133)				
Output comparator (WTA)		0.334***	0.337***		0.244***				
		(0.0197)	(0.0222)		(0.0200)				
Input comparator (workload)		-0.153***	-0.133***	-0.0318***					
		(0.0136)	(0.0158)	(0.0116)					
Input (winning chance)						8.418***	8.115***		
						(0.335)	(0.308)		
Output comparator (amount)							0.0776***		
							(0.0174)		
Input comparator (winning chance)							-0.460**		
							(0.209)		
Input (log prior odds)								0.445***	0.422***
								(0.0379)	(0.0328)
Input (log likelihood ratio)								0.825***	0.790***
								(0.0472)	(0.0367)
Output comparator (log posterior)									0.150***
									(0.0267)
Input comparator (log prior)									-0.0707**
									(0.0281)
Input comparator (log likelihood ratio)									-0.0477*
									(0.0257)
Observations	1575	2670	2940	2865	2775	1620	2805	1646	2778
Subjects	105	178	196	191	185	108	187	112	189
R-squared	0.747	0.705	0.655	0.723	0.651	0.782	0.691	0.644	0.563

Notes. Main pre-registered specification. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

Table G2: Robustness (Response-time validity)

Paradigm:	**Effort valuation**					**Investment**		**Beliefs**	
Dependent variable:	Minimum acceptable compensation (WTA)					Amount wagered		Log posterior odds	
Treatment:	*No CP*	*Both CP (social)*	*Both CP (expectations)*	*Output CP only (social)*	*Input CP only (social)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (social)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input (workload)	0.148***	0.249***	0.195***	0.189***	0.149***				
	(0.0138)	(0.0113)	(0.0132)	(0.0111)	(0.0110)				
Output comparator (WTA)		0.328***	0.341***		0.253***				
		(0.0174)	(0.0195)		(0.0171)				
Input comparator (workload)		-0.147***	-0.137***	-0.0285***					
		(0.0120)	(0.0141)	(0.00982)					
Input (winning chance)						8.609***	8.294***		
						(0.272)	(0.284)		
Output comparator (amount)							0.0676***		
							(0.0148)		
Input comparator (winning chance)							-0.407**		
							(0.178)		
Input (log prior odds)								0.433***	0.409***
								(0.0328)	(0.0298)
Input (log likelihood ratio)								0.844***	0.777***
								(0.0453)	(0.0323)
Output comparator (log posterior)									0.142***
									(0.0239)
Input comparator (log prior)									-0.0682**
									(0.0278)
Input comparator (log likelihood ratio)									-0.0392
									(0.0238)
Observations	2057	3560	3567	3618	3566	2102	3634	2097	3497
Subjects	145	244	246	249	246	145	247	148	246
R-squared	0.741	0.692	0.639	0.734	0.651	0.794	0.687	0.635	0.553

Notes. Main pre-registered specification. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

Table G3: Robustness (Protocol compliance)

Paradigm:	**Effort valuation**					**Investment**		**Beliefs**	
Dependent variable:	Minimum acceptable compensation (WTA)					Amount wagered		Log posterior odds	
Treatment:	*No CP*	*Both CP (social)*	*Both CP (expectations)*	*Output CP only (social)*	*Input CP only (social)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (social)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input (workload)	0.147***	0.243***	0.190***	0.183***	0.151***				
	(0.0142)	(0.0113)	(0.0131)	(0.0113)	(0.0111)				
Output comparator (WTA)		0.321***	0.336***		0.249***				
		(0.0176)	(0.0195)		(0.0169)				
Input comparator (workload)		-0.144***	-0.133***	-0.0311***					
		(0.0121)	(0.0139)	(0.00978)					
Input (winning chance)						8.501***	8.241***		
						(0.280)	(0.291)		
Output comparator (amount)							0.0641***		
							(0.0147)		
Input comparator (winning chance)							-0.368**		
							(0.177)		
Input (log prior odds)								0.419***	0.393***
								(0.0328)	(0.0303)
Input (log likelihood ratio)								0.862***	0.770***
								(0.0459)	(0.0328)
Output comparator (log posterior)									0.155***
									(0.0246)
Input comparator (log prior)									-0.0691**
									(0.0279)
Input comparator (log likelihood ratio)									-0.0407*
									(0.0238)
Observations	2130	3630	3660	3690	3660	2175	3690	2015	3494
Subjects	142	242	244	246	244	145	246	137	238
R-squared	0.751	0.697	0.654	0.732	0.667	0.791	0.682	0.638	0.545

Notes. Main pre-registered specification. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

Table G4: Combined Effort Choice

Paradigm:	**Effort choice**					
Dependent variable:	Chosen workload					
Sample Restriction:	Strict human verification		Protocol compliance		Response-time validity	
Treatment:	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (expectations)*
	(1)	(2)	(3)	(4)	(5)	(6)
Input (piece rate)	31.77***	34.49***	36.06***	38.11***	38.55***	40.41***
	(6.981)	(4.755)	(6.150)	(4.479)	(6.297)	(4.601)
Input comparator (piece rate)		-6.046**		-7.061**		-7.499***
		(3.059)		(2.853)		(2.880)
Output comparator (Workload)		0.123***		0.121***		0.124***
		(0.0166)		(0.0157)		(0.0159)
Observations	1590	3045	2160	3645	2028	3492
Subjects	106	203	144	243	144	246
R-squared	0.717	0.642	0.687	0.654	0.664	0.641

Notes. Main pre-registered specification. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

Table G5: Determinants of Effort Valuation

Dependent Variable:	Minimum Acceptable Compensation (WTA)								
Paradigm:	**(1) Effort Valuation**			**(2) Effort Valuation**			**(3) Effort Valuation**		
Mechanism (M):	*Time (social)*			*Informativeness (expectations)*			*Deliberation (social)*		
Sample / Spec:	Strict human verification	Response-time validity	Protocol compliance	Strict human verification	Response-time validity	Protocol compliance	Strict human verification	Response-time validity	Protocol compliance
Main Effects									
Input (workload)	0.246***	0.243***	0.249***	0.147***	-0.483***	0.152***	0.231***	0.237***	0.233***
	(0.0133)	(0.0113)	(0.0113)	(0.0163)	(0.0404)	(0.0163)	(0.0173)	(0.0163)	(0.0165)
Input comparator (workload)	-0.153***	-0.143***	-0.146***	-0.0279	-0.127***	-0.0392*	-0.113***	-0.118***	-0.117***
	(0.0136)	(0.0121)	(0.0120)	(0.0155)	(0.0179)	(0.0155)	(0.0128)	(0.0119)	(0.0119)
Output comparator (WTA)	0.338***	0.326***	0.333***	0.214***	0.331***	0.219***	0.319***	0.340***	0.333***
	(0.0192)	(0.0170)	(0.0169)	(0.0250)	(0.0280)	(0.0229)	(0.0272)	(0.0248)	(0.0251)
Interaction Effects (Mechanism × Variable)									
$M \times$ Input (workload)	0.162***	0.169***	0.163***	0.0518*	0.692***	0.0393	0.134***	0.135***	0.132***
	(0.0183)	(0.0158)	(0.0157)	(0.0251)	(0.0475)	(0.0235)	(0.0227)	(0.0206)	(0.0209)
$M \times$ Input comparator (workload)	0.0722***	0.0661***	0.0685***	-0.109***	0	-0.0845***	0.0517***	0.0531***	0.0546***
	(0.0164)	(0.0144)	(0.0144)	(0.0228)	(.)	(0.0215)	(0.0154)	(0.0140)	(0.0139)
$M \times$ Output comparator (WTA)	-0.144***	-0.139***	-0.145***	0.119**	0	0.101**	-0.107***	-0.137***	-0.136***
	(0.0232)	(0.0204)	(0.0203)	(0.0374)	(.)	(0.0339)	(0.0320)	(0.0286)	(0.0288)
Observations	6,690	8,715	8,649	3,631	1,799	4,471	5,470	6,695	6,864
R-squared	0.735	0.730	0.730	0.676	0.674	0.683	0.767	0.766	0.774

Notes. Main pre-registered specification. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects. The "Mechanism" (M) variable changes by block.

G.2 Median (Quantile) Regressions

Table G6: Median regression ($\tau = 0.5$): Effort valuation, investment, beliefs

Paradigm:	**Effort valuation**					**Investment**		**Beliefs**	
Dependent variable:	Median acceptable compensation (WTA)					Amount wagered (median)		Log posterior odds (median)	
Treatment:	*No CP*	*Both CP (social)*	*Both CP (expectations)*	*Output CP only (social)*	*Input CP only (social)*	*No CP*	*Both CP (expectations)*	*No CP*	*Both CP (social)*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input (workload)	0.234***	0.338***	0.267***	0.260***	0.197***				
	(0.0352)	(0.0170)	(0.0205)	(0.0308)	(0.0202)				
Input comparator (workload)		-0.215***	-0.217***	-0.0265					
		(0.0162)	(0.0194)	(0.0258)					
Output comparator (WTA)		0.538***	0.562***		0.475***				
		(0.0243)	(0.0259)		(0.0268)				
Input (winning chance)						8.621***	9.518***		
						(0.440)	(0.323)		
Input comparator (winning chance)							-0.519*		
							(0.231)		
Output comparator (amount)							0.0723***		
							(0.0211)		
Input (log prior odds)								0.383***	0.411***
								(0.0613)	(0.0499)
Input (log likelihood ratio)								0.926***	0.881***
								(0.0429)	(0.0319)
Input comparator (log prior)									-0.0477
									(0.0349)
Input comparator (log likelihood ratio)									-0.0428
									(0.0293)
Output comparator (log posterior)									0.128***
									(0.0283)
Observations	2175	3660	3690	3735	3690	2175	3705	2180	3610
Subjects	145	244	246	249	246	145	247	148	246

Notes. Median (quantile) regressions with $\tau = 0.5$. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Fixed effects as in the corresponding OLS specifications.

Table G7: Median regression ($\tau = 0.5$): Effort Valuation Mechanisms and Effort Choice

Paradigm:	**Effort Valuation (WTA)**			**Effort Choice (Workload)**	
Treatment:	**Deliberation (social)**	**Time (social)**	**Informativeness (expectations)**	**No CP**	**Both CP (expectations)**
	(Table 1)	(Table 2)	(Table 4)	(Table 3)	
	(1)	(2)	(3)	(4)	(5)
Main Effects					
Input (Workload/Piece Rate)	0.319***	0.338***	0.229***	27.78*	40.45***
	(0.0204)	(0.0180)	(0.0372)	(11.67)	(7.949)
Input comparator (Workload/Piece Rate)	-0.126***	-0.215***	-0.0643		-17.33***
	(0.0196)	(0.0162)	(0.0372)		(5.267)
Output comparator (WTA/Workload)	0.553***	0.538***	0.434***		0.247***
	(0.0319)	(0.0215)	(0.0514)		(0.0555)
Interactions					
Treatment × Input (Workload)	0.158***	0.168***	0.0533		
	(0.0252)	(0.0256)	(0.0439)		
Treatment × Input comparator (Workload)	-0.00561	0.120***	-0.153***		
	(0.0226)	(0.0234)	(0.0412)		
Treatment × Output comparator (WTA)	-0.238***	-0.251***	0.133*		
	(0.0362)	(0.0347)	(0.0573)		
Observations	6939	8850	4486	2175	3690
Subjects	463	590	300	145	246

Notes. Median (quantile) regressions with $\tau = 0.5$. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Fixed effects as in the corresponding OLS specifications.

G.3 Round Interactions

Table G8: Round Interactions (Time (social) and Effort Choice)

Paradigm:	**Time (social)**	**Effort Choice**	
	(1)	(2)	(3)
Specification:	All Rounds	No CP	Both CP (expectations)
Panel A: Time (social) mechanism (Effort Valuation)			
Round × workload	0.00104		
	(0.00192)		
Round × Output comparator (WTA)	-0.00345		
	(0.00295)		
Time × Round × Input comparator (workload)	0.00259		
	(0.00240)		
Time × Round × Output comparator (WTA)	-0.00536		
	(0.00345)		
Panel B: Effort Choice Regressors			
Round × Input (piece rate)		3.484***	0.832
		(0.665)	(0.640)
Round × Input comparator (piece rate)			0.623
			(0.580)
Round × Output comparator (workload)			-0.00636*
			(0.00262)
Observations	8,850	2,160	3,645
Subjects	590	144	243
R-squared	0.733	0.717	0.642

Notes. OLS estimates with interactions between round and the listed regressors. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects. "No CP" indicates no comparison points present; "Both CP" indicates that both input and output comparison points are present.

Table G9: Round Interactions (Investment and Beliefs)

Paradigm:	**Investment**		**Beliefs**	
	(1)	(2)	(3)	(4)
Specification:	No CP	Both CP (expectations)	No CP	Both CP (social)
Panel A: Investment Regressors				
Round × Input (winning chance)	-0.00857	0.140***		
	(0.0315)	(0.0322)		
Round × Input comparator (winning chance)		0.0373		
		(0.0348)		
Round × Output comparator (amount wagered)		-0.00242		
		(0.00318)		
Panel B: Beliefs Regressors				
Round × Input (log prior)			0.00849	0.00149
			(0.00560)	(0.00552)
Round × Input (log likelihood ratio)			0.00470	0.00246
			(0.00444)	(0.00360)
Round × Input comparator (log prior)				0.00151
				(0.00585)
Round × Input comparator (log likelihood ratio)				0.00415
				(0.00526)
Round × Output comparator (log posterior)				-0.00369
				(0.00465)
Observations	2,175	3,690	2,015	3,494
Subjects	145	246	137	238

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. "No CP" indicates no comparison points present; "Both CP" indicates that both input and output comparison points are present.

G.4 Differing Effects Depending on Whether the Comparison is "Favorable" or "Unfavorable"

Table G10: Robustness: Favorable vs. Unfavorable Comparisons

	Effort Valuation						Effort Choice		Investment	
	Both CP (social)		Both CP (expectations)		Input CP only (social)		Both CP (expectations)		Both CP (expectations)	
	Fav.	Unfav.	Fav.	Unfav.	Fav.	Unfav.	Fav.	Unfav.	Fav.	Unfav.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Input (workload/piece rate/winning chance)	0.231***	0.157***	0.210***	0.105***	0.206***	0.151***	30.67***	42.57***	8.329***	7.823***
	(0.021)	(0.020)	(0.026)	(0.016)	(0.020)	(0.019)	(8.62)	(6.51)	(0.38)	(0.42)
Input comparator (workload/piece rate/winning chance)	-0.086***	-0.112***	-0.111***	-0.093***	-0.021	-0.024	-6.130	-7.687	0.235	-0.765**
	(0.020)	(0.021)	(0.028)	(0.022)	(0.019)	(0.020)	(7.20)	(4.73)	(0.35)	(0.30)
Output comparator (WTA/workload/amount wagered)	0.332***	0.302***	0.329***	0.358***			0.116***	0.114***	0.067***	0.086***
	(0.022)	(0.024)	(0.027)	(0.026)			(0.020)	(0.020)	(0.025)	(0.019)
Observations	1868	1882	1630	2120	1899	1851	1651	2099	1918	1832
Subjects	250	250	250	250	250	250	250	250	250	250
R-squared	0.717	0.722	0.690	0.687	0.761	0.757	0.730	0.678	0.738	0.706

Notes. This table replicates the main analysis splitting the sample by whether the comparison point was favorable or unfavorable. Columns (1)-(6) show Effort Valuation, columns (7)-(8) show Effort Choice, and columns (9)-(10) show Investment. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns split the sample by whether the comparison point was favorable or unfavorable (e.g., peer had a higher workload or lower wage). All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

G.5 Effort Choice – Additional Specifications

Table G11: Effort Choice – Earnings as DV

Effort choice – Earnings as DV		
	No CP	Both CP (expectations)
	(1)	(2)
	Earnings	Earnings
Input (piece rate)	36.82***	37.49***
	(2.875)	(1.820)
Input comparator (piece rate)		-1.197
		(0.810)
Output comparator (workload)		0.0272***
		(0.00399)
Observations	2160	3645
Subjects	144	243

Notes. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects. Earnings are computed from the chosen workload and piece rate.

Table G12: Effort Choice – Baseline with time-normalized DV

Effort choice, time-normalized		
	No CP	Both CP (expectations)
	(1)	(2)
	Time estimate of chosen workload	Time estimate of chosen workload
Input (piece rate)	29.43***	22.37***
	(4.771)	(2.522)
Input comparator (piece rate)		-3.708*
		(1.642)
Output comparator (workload)		0.124***
		(0.0161)
Observations	2175	3690
Subjects	145	246

Notes. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects. The dependent variable is the time estimate of the chosen workload.

G.6 Additional secondary analysis – Informativeness (expectations) and Deliberation (social)

Table G13: Informativeness (expectations) – decision-making self-report

Group	**Subjects**	**Share considering default workload and/or compensation**
Uninformative (random)	145	68.3%
Informative (control)	155	71.6%

Notes. This table reports subject counts and shares. Shares are percentages.

Table G14: Deliberation (social) – per task analysis

Effort Valuation only: split by task type (OLS)			
	Counting	Sliders	Greek
	(1)	(2)	(3)
Input (workload)	0.313***	0.447***	0.331***
	(0.0206)	(0.0230)	(0.0200)
Input comparator (workload)	-0.0633***	-0.101***	-0.0502***
	(0.0109)	(0.0193)	(0.0101)
Output comparator (WTA)	0.233***	0.195***	0.176***
	(0.0258)	(0.0261)	(0.0224)
Observations	1512	1543	1560
Subjects	101	103	104
R-squared	0.792	0.805	0.836

Notes. OLS estimates. Standard errors in parentheses are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects. Each column restricts to a single task type.

H Experimental Instructions, Comparison Point Announcements, and Decision Screens

H.1 Experimental Instructions

Here we show screenshots of the experimental instructions and comprehension questions for each of the four paradigms (effort valuation, effort choice, investment, and beliefs). These were the same for different treatment conditions within a given paradigm, specifically effort valuation. For the effort valuation time and Deliberation treatments, we in addition present screenshots of the training screens; for the Deliberation treatment, we also include the task type announcement screens.

H.1.1 Effort Valuation

Instructions

You will **decide for which amount of money you're willing to complete a given work offer.**

The work offer:

- Each work offer specifies the task you would have to work on, as well as the total workload you would have to complete.
- There are 3 **different types of tasks** that will be described to you on the following pages.
- If you accept a work offer, you will be paid a Dollar **compensation** if you complete the workload in full.

Your decision and bonus payment:

- **You will tell us how much money we would at least have to pay you to complete the offer. We call this your lowest acceptable compensation.**
- The computer will determine the actual compensation by randomly drawing a **payment between $1 and $49**, with equal probability. You will then work if the amount drawn by the computer is at least as high as your lowest acceptable compensation.
- This means: **stating a lowest acceptable compensation that is high means that you only agree to work if the compensation is high.** For example:
  - If your lowest acceptable compensation is $1, you will definitely work. You agree to work whatever the compensation is, and will be paid the actual compensation.
  - If your lowest acceptable compensation is $50, you will definitely NOT work. You decline the work offer whatever the actual compensation is.
  - If your lowest acceptable compensation is in between $1 and $50, you may or may not work.
- You should **carefully think** about **how the workload offered (i.e., the number of tasks) affects the amount of money you would require to complete it.**

Important: your lowest acceptable compensation is not a hypothetical number but determines whether you actually work and receive a payment at the end of this study!

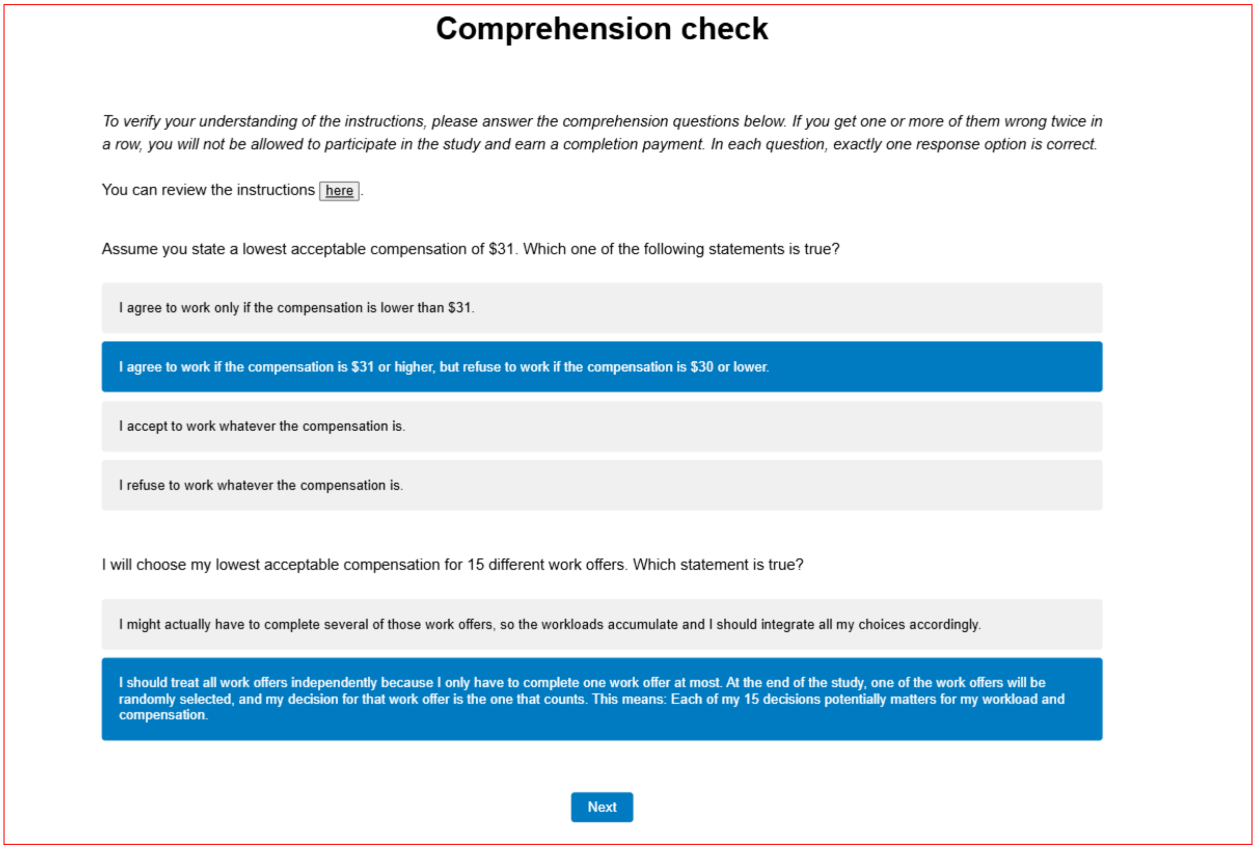
15 decisions:

- In total, you will make a decision for each of **15 different work offers**.
- Across the work offers, **the type of task and the workload** vary and will be shown to you.
- You **do NOT have to complete multiple work offers**! Only one of your 15 decisions will be randomly selected as the "decision that counts" at the end of the study, and this decision of yours will determine your workload and compensation. This means: Each of your 15 decisions might be the one that matters in the end. You should treat them independently because only one of them will count. Each of your decisions is equally likely to be the "decision that counts."

Next

Figure 9: Instruction screen (2) for the *Effort Valuation* paradigm.

Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Informativeness*, *Deliberation*); the decision screen that followed varied.



Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

You can review the instructions here.

Assume you state a lowest acceptable compensation of $31. Which one of the following statements is true?

- I agree to work only if the compensation is lower than $31.
- I agree to work if the compensation is $31 or higher, but refuse to work if the compensation is $30 or lower.
- I accept to work whatever the compensation is.
- I refuse to work whatever the compensation is.

I will choose my lowest acceptable compensation for 15 different work offers. Which statement is true?

- I might actually have to complete several of those work offers, so the workloads accumulate and I should integrate all my choices accordingly.
- I should treat all work offers independently because I only have to complete one work offer at most. At the end of the study, one of the work offers will be randomly selected, and my decision for that work offer is the one that counts. This means: Each of my 15 decisions potentially matters for my workload and compensation.

Next

Figure 10: Comprehension-check screen for the *Effort Valuation* paradigm.
Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Informativeness*, *Deliberation*); the decision screen that followed varied.

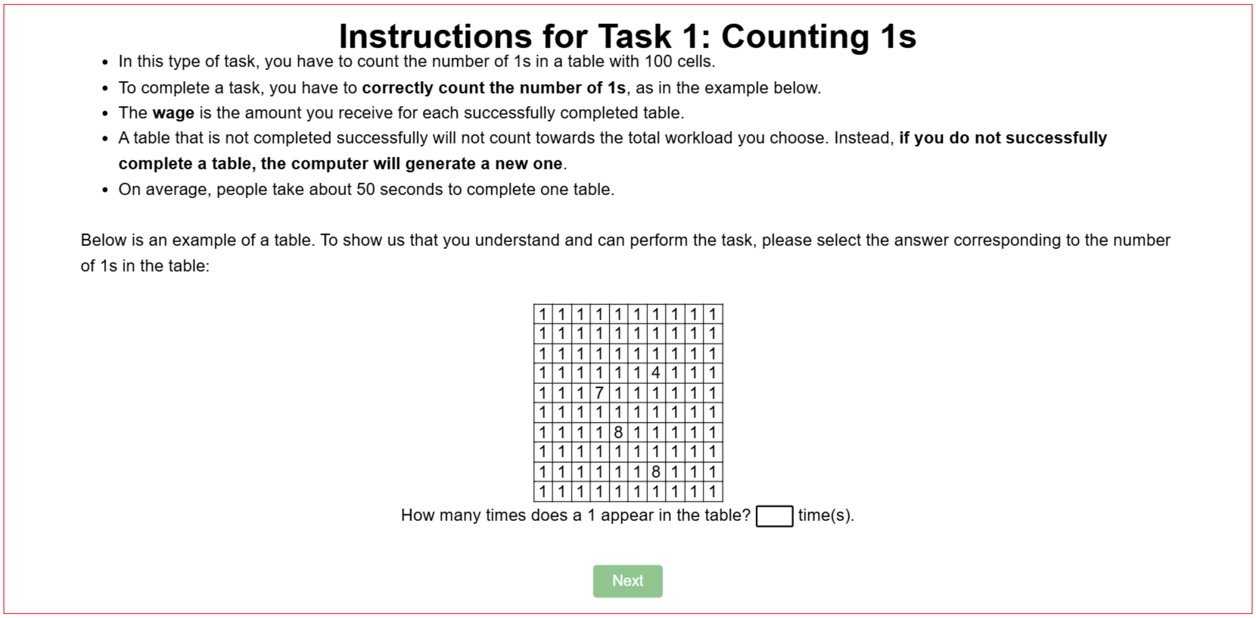


Figure 11: Counting-task instruction screen for the *Effort Valuation* paradigm.

Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Informativeness*, *Deliberation*); the decision screen that followed varied.

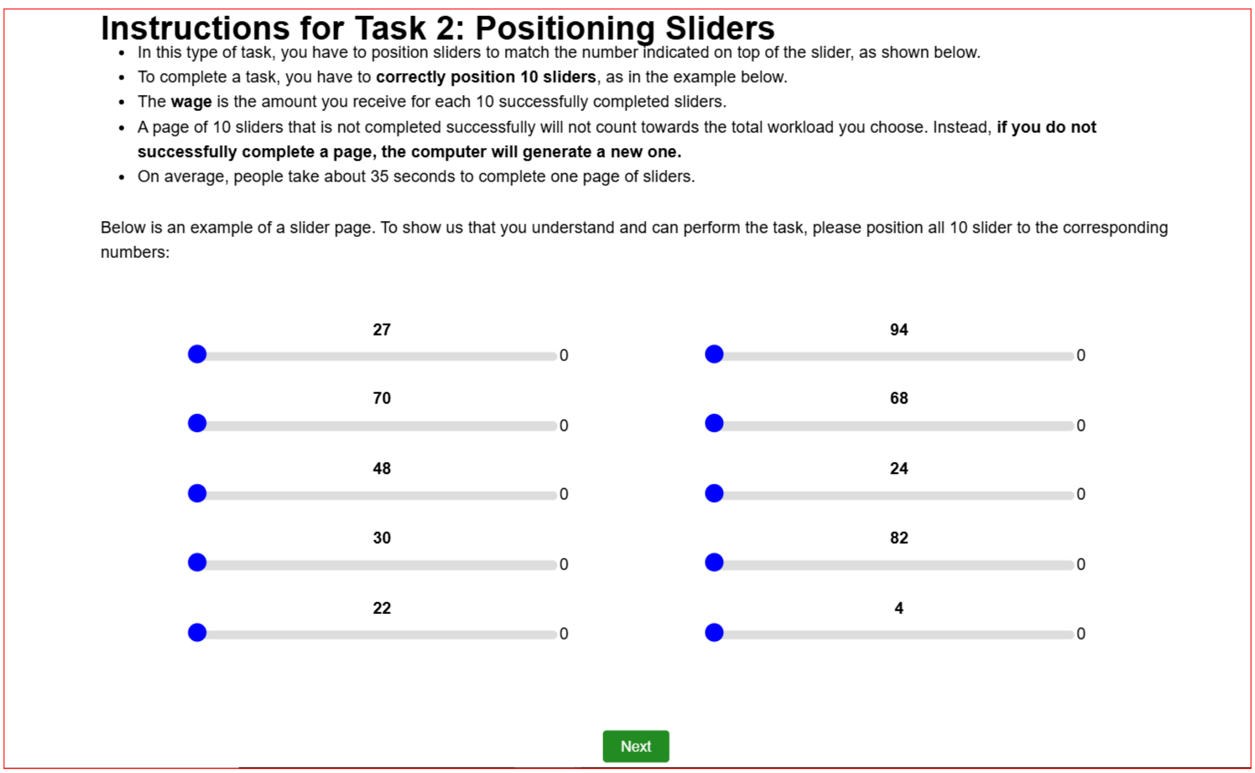


Figure 12: Slider-task instruction screen for the *Effort Valuation* paradigm.

Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Informativeness*, *Deliberation*); the decision screen that followed varied.

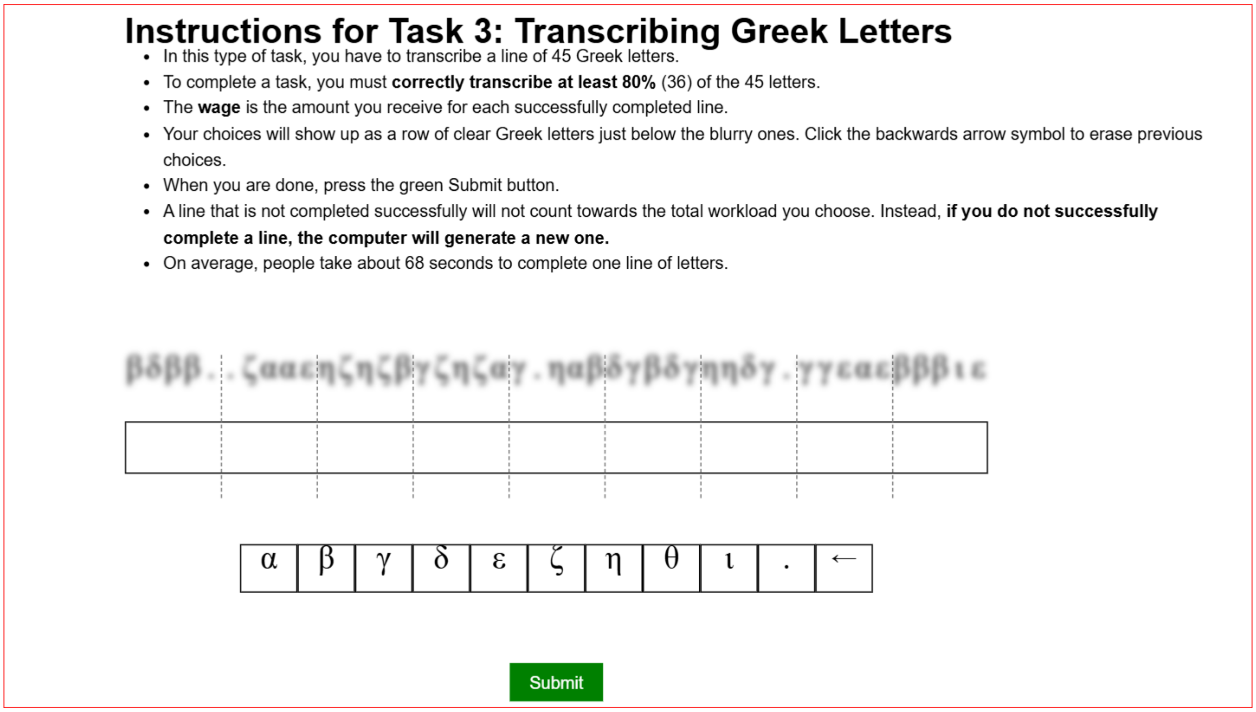
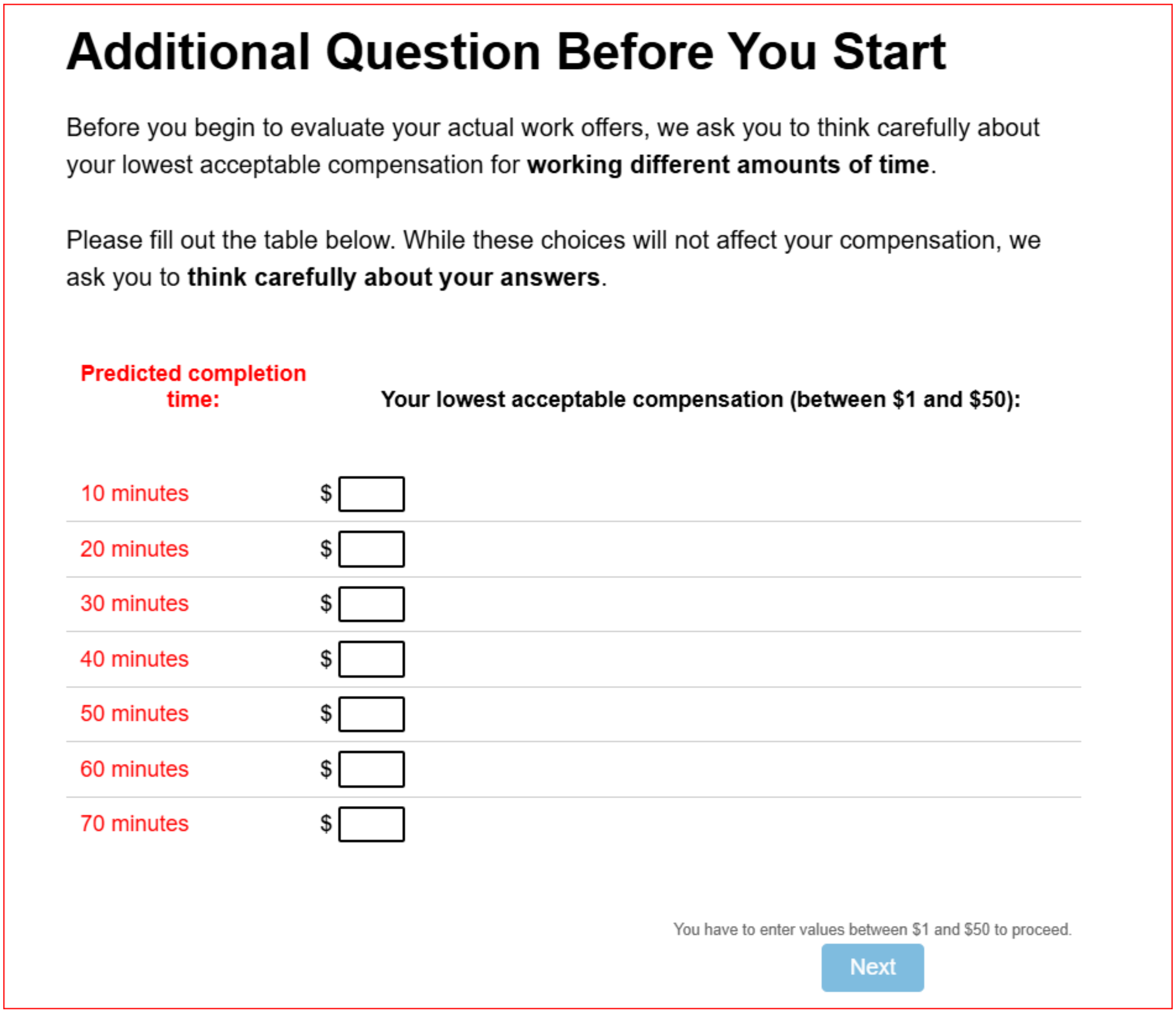


Figure 13: Greek-letters typing-task instruction screen for the *Effort Valuation* paradigm.
Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Informativeness*, *Deliberation*); the decision screen that followed varied.



Additional Question Before You Start

Before you begin to evaluate your actual work offers, we ask you to think carefully about your lowest acceptable compensation for **working different amounts of time**.

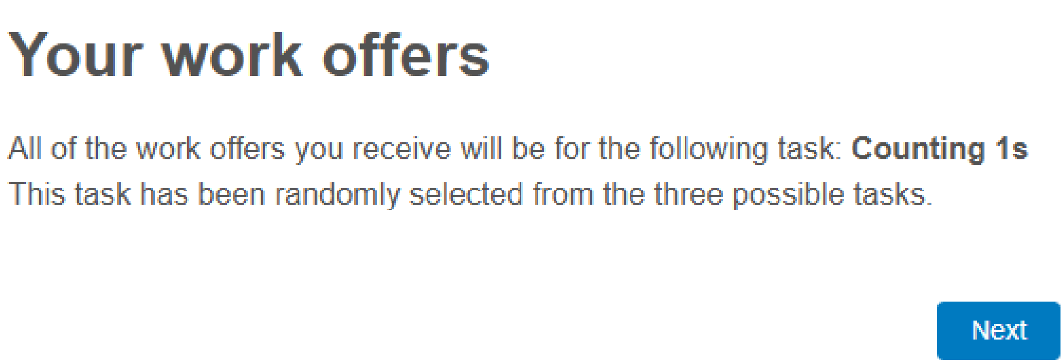
Please fill out the table below. While these choices will not affect your compensation, we ask you to **think carefully about your answers**.

Predicted completion time:	**Your lowest acceptable compensation (between $1 and $50):**
10 minutes	$
20 minutes	$
30 minutes	$
40 minutes	$
50 minutes	$
60 minutes	$
70 minutes	$

You have to enter values between $1 and $50 to proceed.

Next

Figure 14: Additional question shown only in the *Time* treatment.

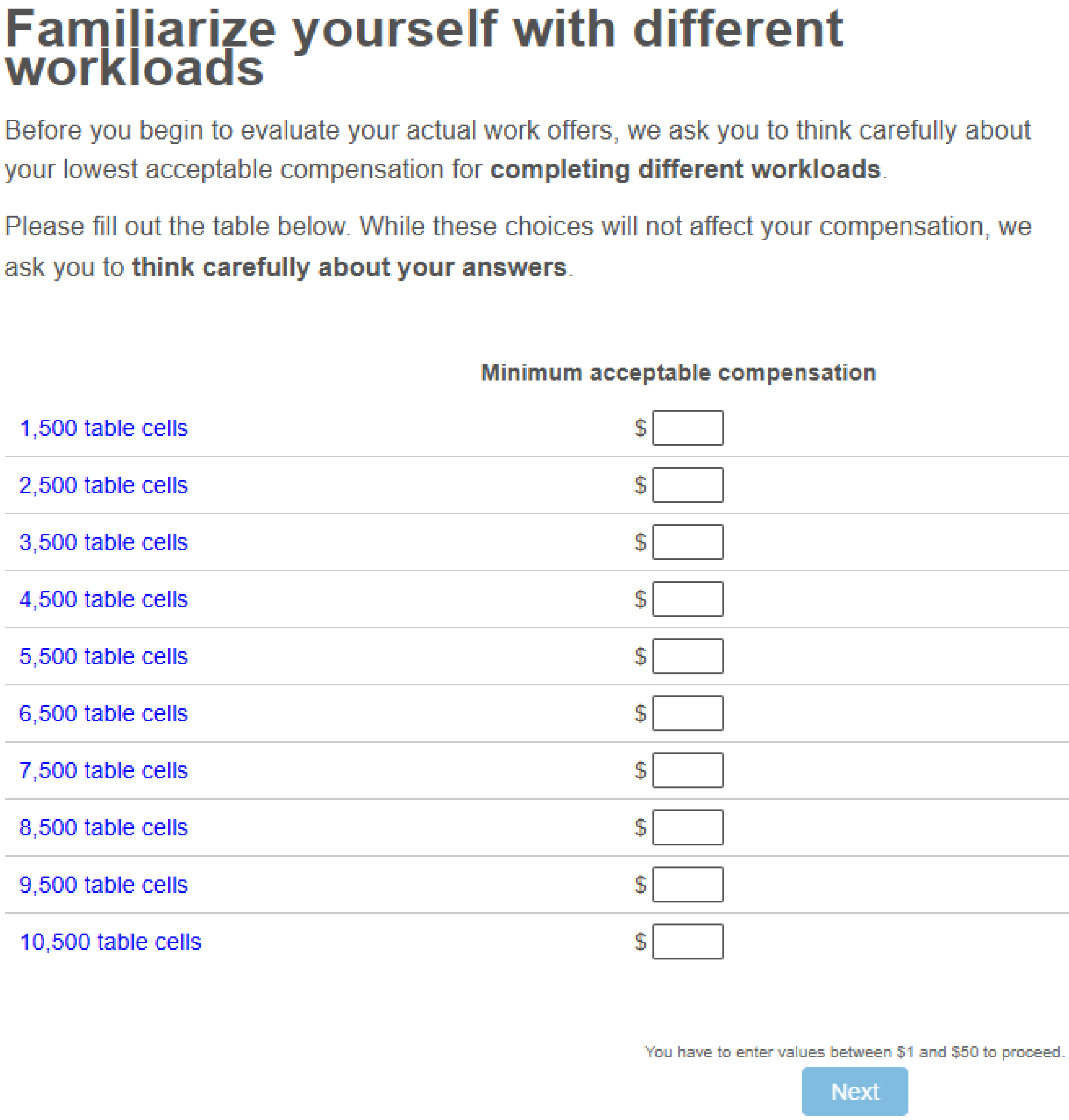


Your work offers

All of the work offers you receive will be for the following task: **Counting 1s**
This task has been randomly selected from the three possible tasks.

Next

Figure 15: Counting ones announcement screen for the *Deliberation* treatment.



Familiarize yourself with different workloads

Before you begin to evaluate your actual work offers, we ask you to think carefully about your lowest acceptable compensation for **completing different workloads**.

Please fill out the table below. While these choices will not affect your compensation, we ask you to **think carefully about your answers**.

	Minimum acceptable compensation
1,500 table cells	$
2,500 table cells	$
3,500 table cells	$
4,500 table cells	$
5,500 table cells	$
6,500 table cells	$
7,500 table cells	$
8,500 table cells	$
9,500 table cells	$
10,500 table cells	$

You have to enter values between $1 and $50 to proceed.

Next

Figure 16: Counting ones training screen for the *Deliberation* treatment.

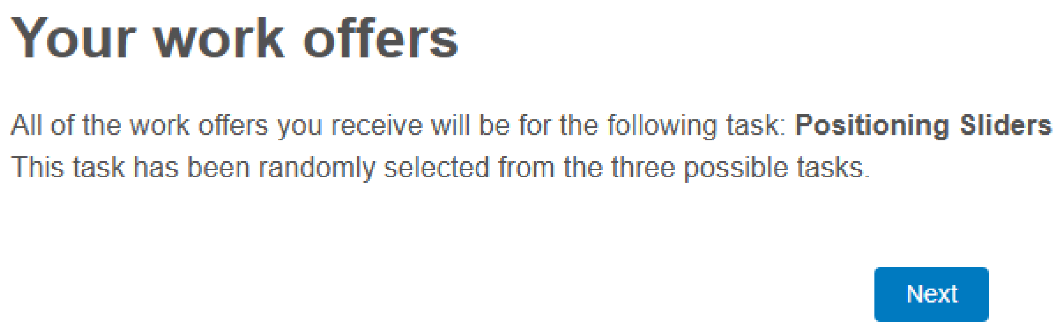


Figure 17: Positioning sliders announcement screen for the *Deliberation* treatment.

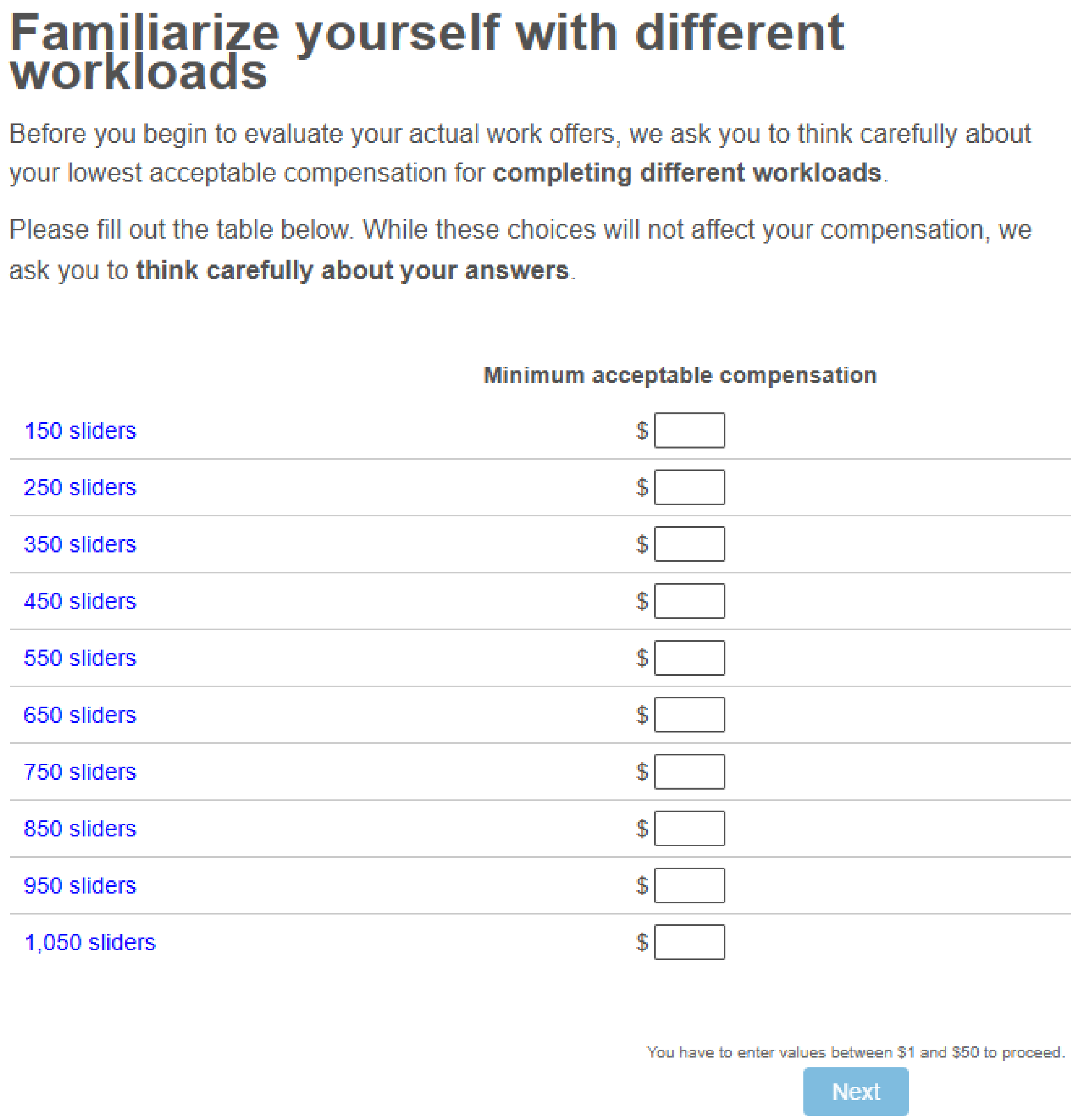


Figure 18: Positioning sliders training screen for the *Deliberation* treatment.

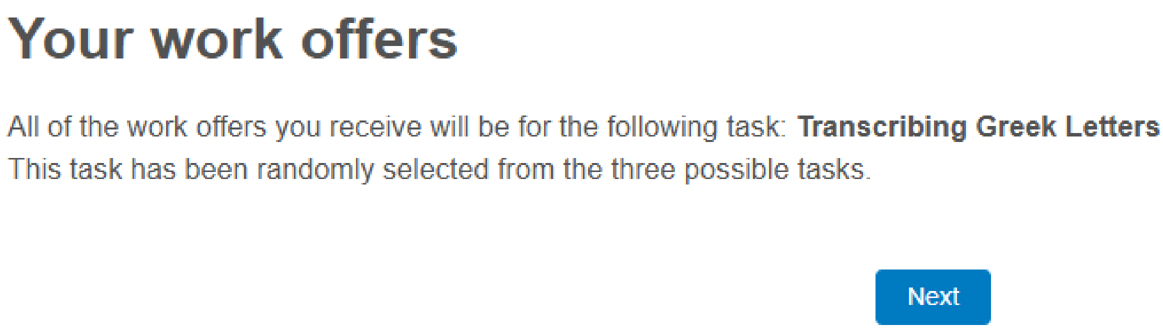


Figure 19: Greek-letters typing-task training screen for the *Deliberation* treatment.

Familiarize yourself with different workloads

Before you begin to evaluate your actual work offers, we ask you to think carefully about your lowest acceptable compensation for **completing different workloads**.

Please fill out the table below. While these choices will not affect your compensation, we ask you to **think carefully about your answers**.

	Minimum acceptable compensation
500 Greek letters	$
750 Greek letters	$
1000 Greek letters	$
1,250 Greek letters	$
1,500 Greek letters	$
1,750 Greek letters	$
2,000 Greek letters	$
2,250 Greek letters	$
2,500 Greek letters	$
2,750 Greek letters	$
3,000 Greek letters	$
3,250 Greek letters	$
3,500 Greek letters	$

You have to enter values between $1 and $50 to proceed.

Next

Figure 20: Greek-letters typing-task training screen for the *Deliberation* treatment.

H.1.2 Effort Choice

Instructions

You will **decide how many tasks you want to complete for a given wage.**

The work decision:

- You will be offered a wage for every task you complete. You then decide how many tasks you would like to be assigned. We call this your chosen workload. Your payment for the full workload you choose equals the **wage times the number of tasks you complete.**
- There are **3 different types of tasks** that will be described to you on the following pages.
- A task that is not completed successfully will not count towards the total of tasks you need to complete for your chosen workload. Instead, **if you do not successfully complete a task, the computer will generate a new one.**

Your bonus payment:

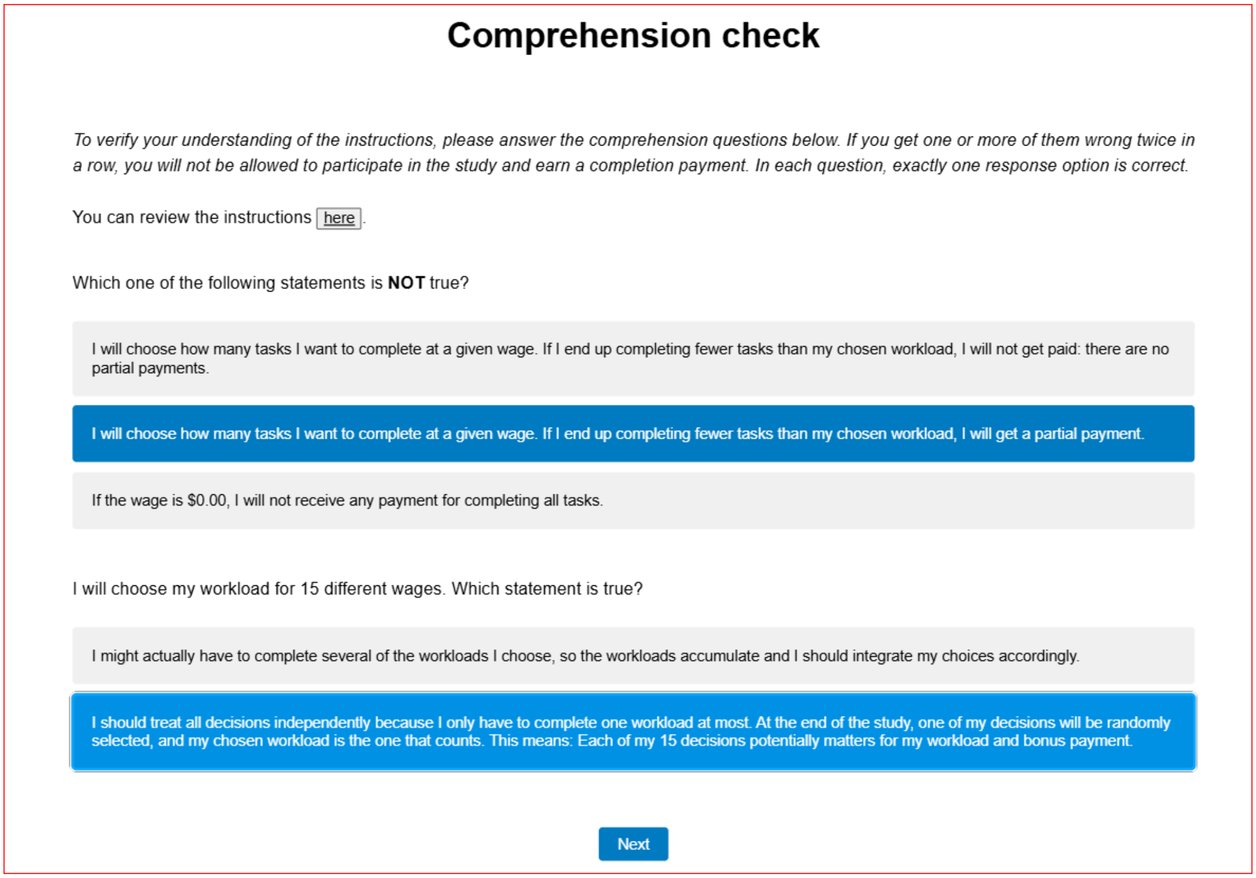
- Your decisions may affect your bonus payment as well as how many tasks you will have to work on to receive your bonus. If you are chosen for a bonus, you will have to complete your workload and you will receive the total payment in return. If you do not complete the total workload you chose, you will not receive any bonus payment. There is no partial payment for partially completed workloads.

15 decisions:

- In each round, you will be told the wage per task. You will then decide how many tasks your workload should include.
- In total, you will make a decision for each of **15 different wages**. If one of the rounds of this task is selected to determine your bonus, only your decision in this one round will determine your bonus.
- You **do NOT have to complete multiple workloads!** Only one of your 15 decisions will be randomly selected as the "decision that counts" at the end of the study, and this decision of yours will determine your workload and compensation. This means: Each of your 15 decisions might be the one that matters in the end. You should treat them independently because only one of them will count. Each of your decisions is equally likely to be the “decision that counts.”

Next

Figure 21: Instruction screen for the *Effort Choice* paradigm.



Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

You can review the instructions here.

Which one of the following statements is **NOT** true?

I will choose how many tasks I want to complete at a given wage. If I end up completing fewer tasks than my chosen workload, I will not get paid: there are no partial payments.

I will choose how many tasks I want to complete at a given wage. If I end up completing fewer tasks than my chosen workload, I will get a partial payment.

If the wage is $0.00, I will not receive any payment for completing all tasks.

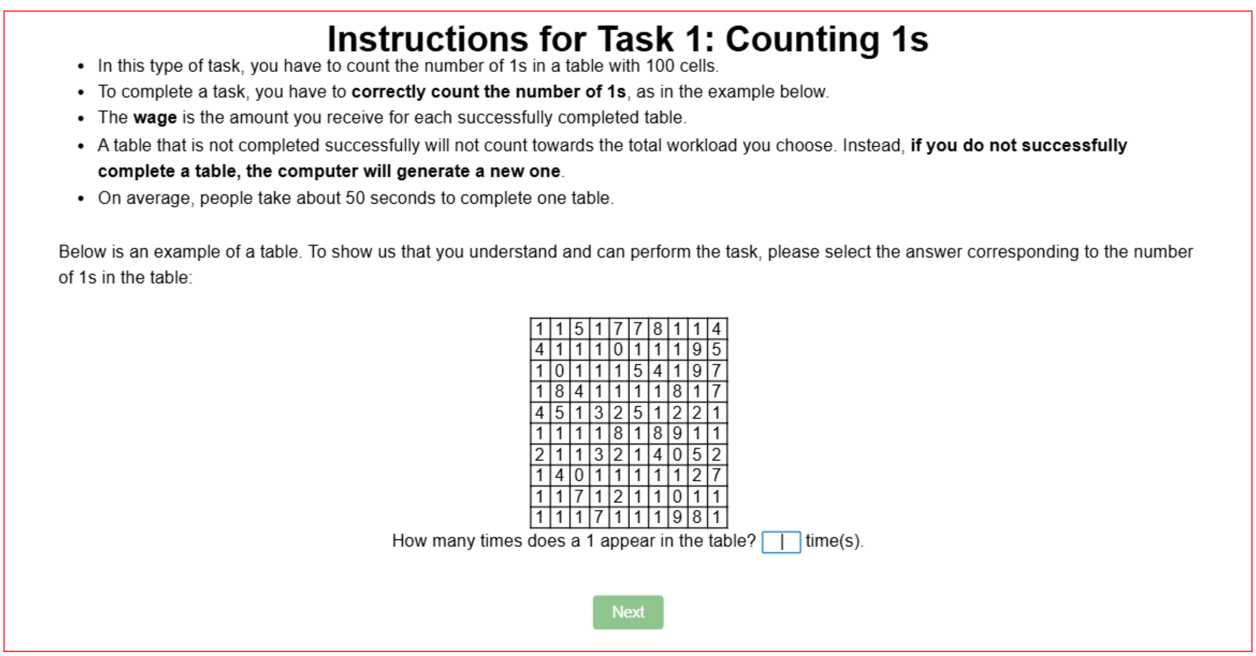
I will choose my workload for 15 different wages. Which statement is true?

I might actually have to complete several of the workloads I choose, so the workloads accumulate and I should integrate my choices accordingly.

I should treat all decisions independently because I only have to complete one workload at most. At the end of the study, one of my decisions will be randomly selected, and my chosen workload is the one that counts. This means: Each of my 15 decisions potentially matters for my workload and bonus payment.

Next

Figure 22: Comprehension-check screen for the *Effort Choice* paradigm.



Instructions for Task 1: Counting 1s

- In this type of task, you have to count the number of 1s in a table with 100 cells.
- To complete a task, you have to **correctly count the number of 1s**, as in the example below.
- The **wage** is the amount you receive for each successfully completed table.
- A table that is not completed successfully will not count towards the total workload you choose. Instead, **if you do not successfully complete a table, the computer will generate a new one**.
- On average, people take about 50 seconds to complete one table.

Below is an example of a table. To show us that you understand and can perform the task, please select the answer corresponding to the number of 1s in the table:

1	1	5	1	7	7	8	1	1	4
4	1	1	1	0	1	1	1	9	5
1	0	1	1	1	5	4	1	9	7
1	8	4	1	1	1	1	8	1	7
4	5	1	3	2	5	1	2	2	1
1	1	1	1	8	1	8	9	1	1
2	1	1	3	2	1	4	0	5	2
1	4	0	1	1	1	1	1	2	7
1	1	7	1	2	1	1	0	1	1
1	1	1	7	1	1	1	9	8	1

How many times does a 1 appear in the table? time(s).

Next

Figure 23: Counting-task instruction screen for the *Effort Choice* paradigm.

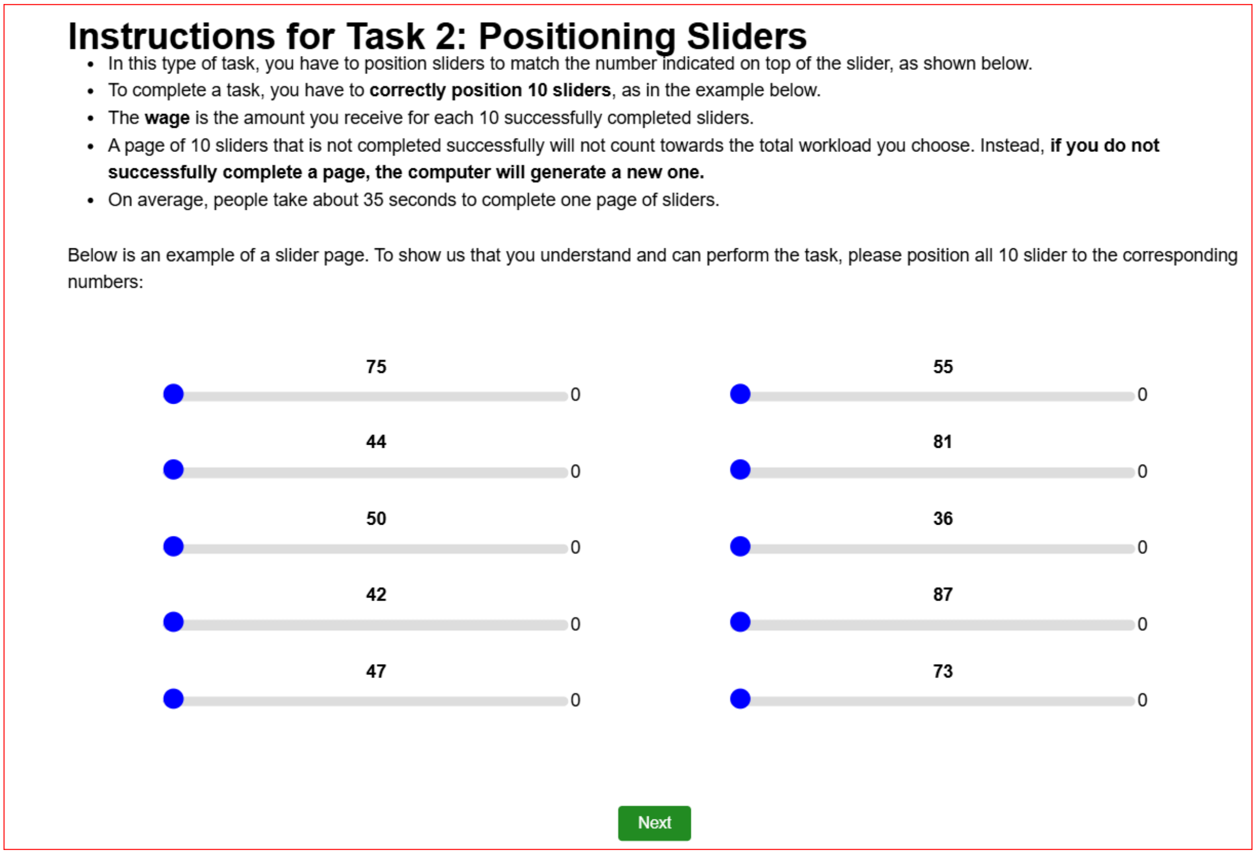


Figure 24: Slider-task instruction screen for the *Effort Choice* paradigm.

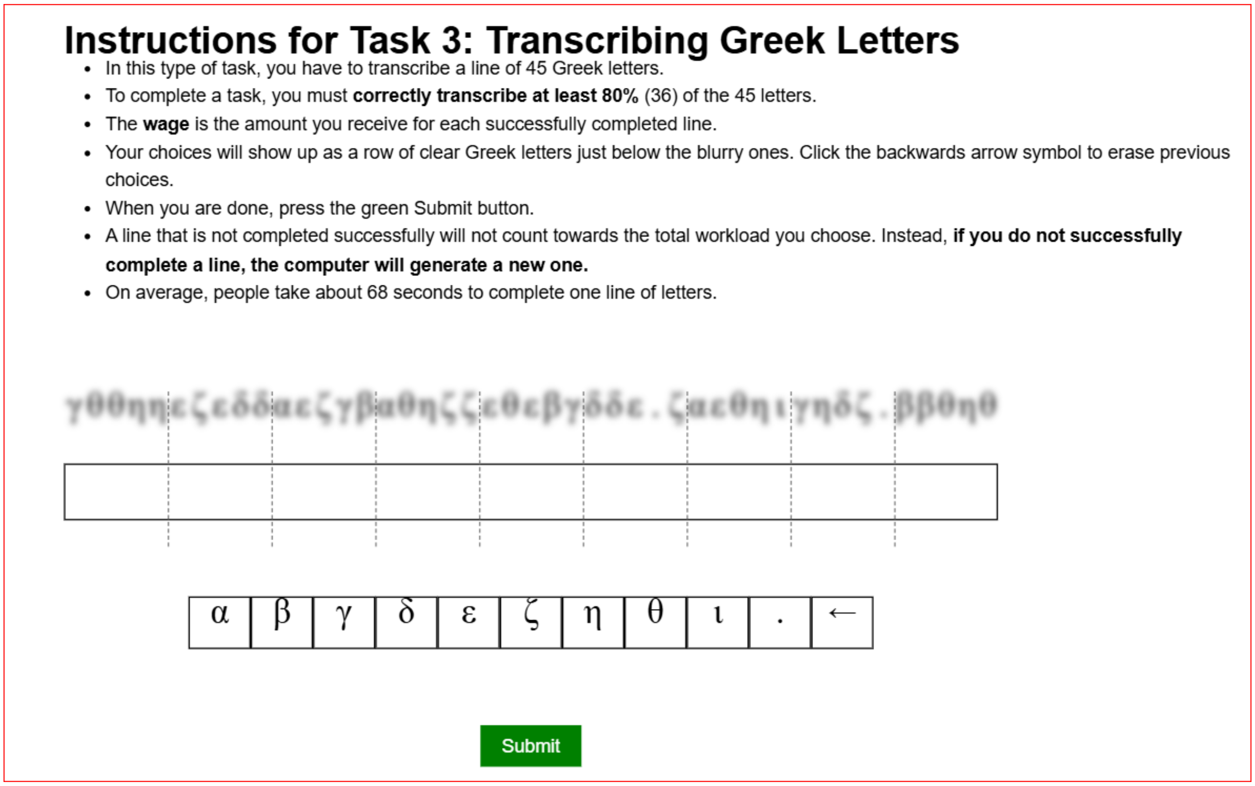
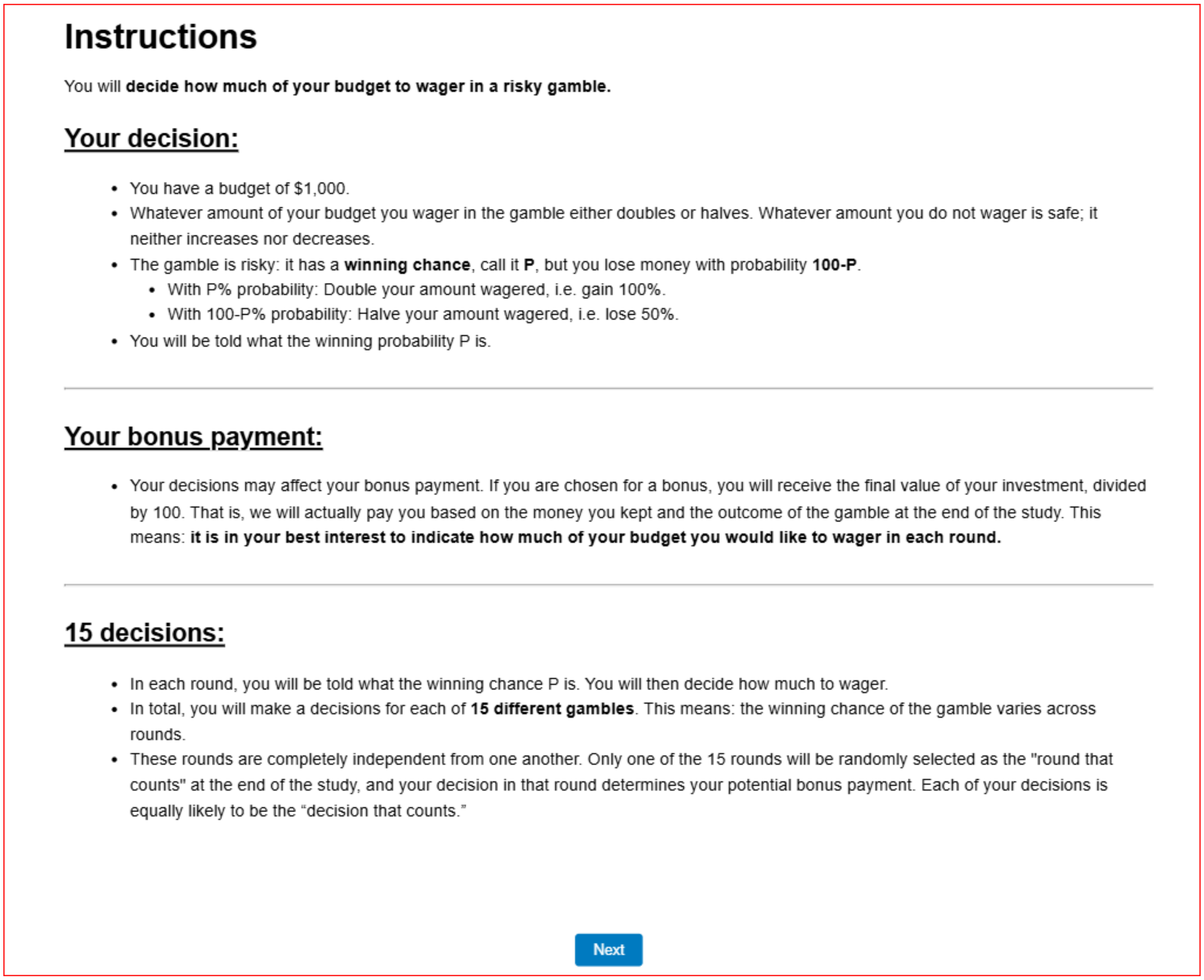


Figure 25: Greek-letters typing-task instruction screen for the *Effort Choice* paradigm.

H.1.3 Investment



Instructions

You will **decide how much of your budget to wager in a risky gamble.**

Your decision:

- You have a budget of $1,000.
- Whatever amount of your budget you wager in the gamble either doubles or halves. Whatever amount you do not wager is safe; it neither increases nor decreases.
- The gamble is risky: it has a **winning chance**, call it **P**, but you lose money with probability **100-P**.
  - With P% probability: Double your amount wagered, i.e. gain 100%.
  - With 100-P% probability: Halve your amount wagered, i.e. lose 50%.
- You will be told what the winning probability P is.

Your bonus payment:

- Your decisions may affect your bonus payment. If you are chosen for a bonus, you will receive the final value of your investment, divided by 100. That is, we will actually pay you based on the money you kept and the outcome of the gamble at the end of the study. This means: **it is in your best interest to indicate how much of your budget you would like to wager in each round.**

---

15 decisions:

- In each round, you will be told what the winning chance P is. You will then decide how much to wager.
- In total, you will make a decisions for each of **15 different gambles**. This means: the winning chance of the gamble varies across rounds.
- These rounds are completely independent from one another. Only one of the 15 rounds will be randomly selected as the "round that counts" at the end of the study, and your decision in that round determines your potential bonus payment. Each of your decisions is equally likely to be the “decision that counts.”

Next

Figure 26: Instruction screen for the *Investment* paradigm.

Comprehension check

To verify your understanding of the instructions, please answer the comprehension questions below. If you get one or more of them wrong twice in a row, you will not be allowed to participate in the study and earn a completion payment. In each question, exactly one response option is correct.

Click here *to re-read the instructions.*

Suppose you wager $600 in a gamble that has a winning chance of 30%. Which one of the following statements is true?

My final outcome will definitely be lower than my original budget of $1,000.

My final outcome can be higher or lower than my original budget of $1,000. The wagered $600 will be worth either $1,200 (doubled) or $300 (halved). Given that the $400 not wagered are safe, my final outcome would be either $1,600 or $700.

My final outcome will definitely be higher than my original budget of $1,000.

Which one of the following statements is true?

The wagered amount either doubles or halves. The winning chance tells me how likely it is that it doubles. The higher my winning chance, the less likely it is that my wagered amount halves.

The winning chance tells me by how much the wagered amount increases if I win.

Which one of the following statements is true? If the winning chance is 70%, my wagered amount…

… doubles with 70% probability and halves with 70% probability.

… doubles with 70% probability and halves with 30% probability.

… doubles with 70% probability and stays the same with 30% probability.

I will choose my wager for 15 different gambles. Which statement is true?

All of my 15 decisions affect my bonus payment jointly, therefore I should integrate my decisions across them.

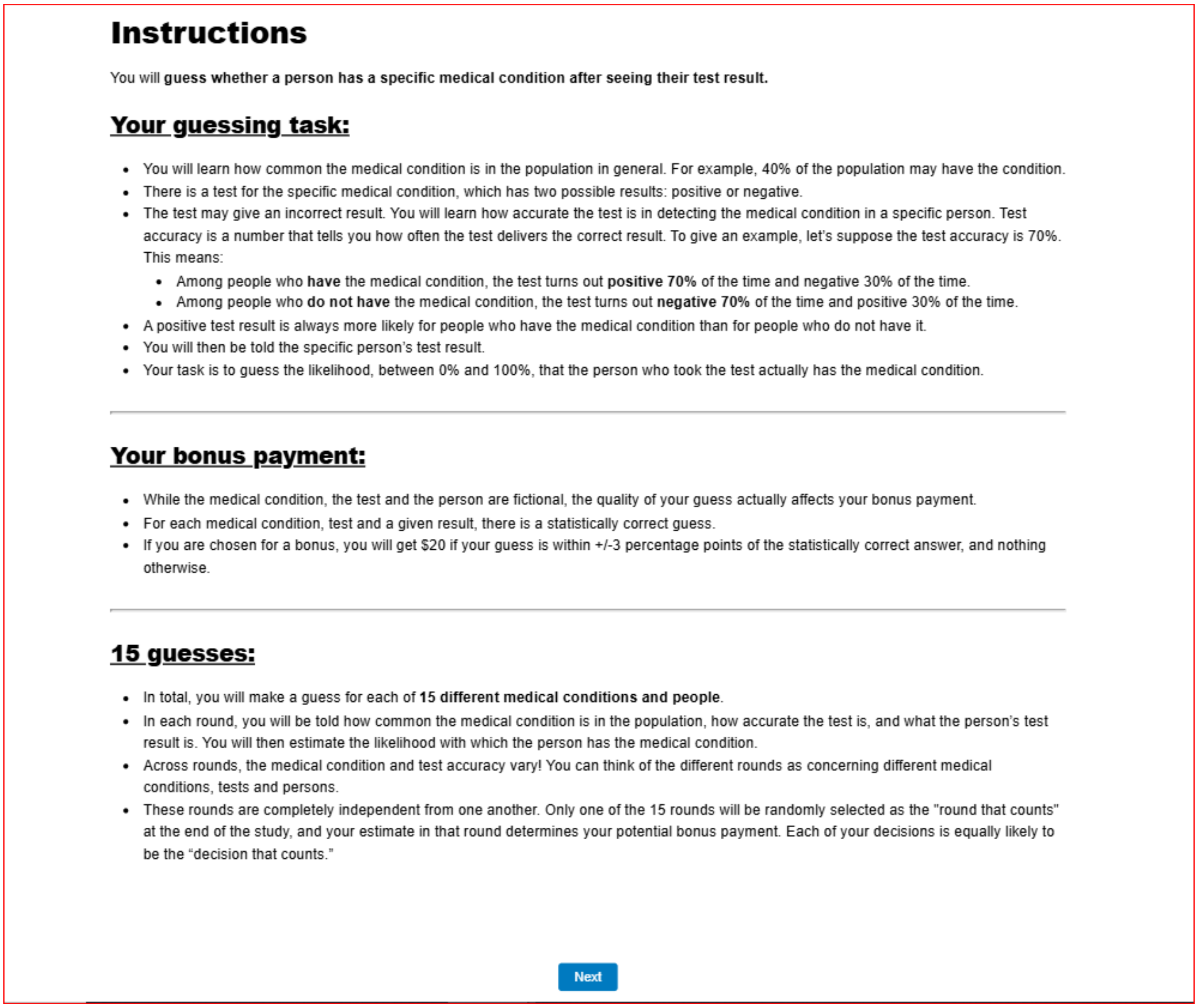
I should treat all 15 decisions independently because, at the end of the study, one of the decisions will be randomly selected to be the "decision that counts", and my wager in that round is the one that determines my bonus.

None of my 15 decisions matters for my bonus payment, therefore it does not matter which wagers I pick.

Next

Figure 27: Comprehension-check screen for the *Investment* paradigm.

H.1.4 Beliefs



Instructions

You will **guess whether a person has a specific medical condition after seeing their test result.**

Your guessing task:

- You will learn how common the medical condition is in the population in general. For example, 40% of the population may have the condition.
- There is a test for the specific medical condition, which has two possible results: positive or negative.
- The test may give an incorrect result. You will learn how accurate the test is in detecting the medical condition in a specific person. Test accuracy is a number that tells you how often the test delivers the correct result. To give an example, let's suppose the test accuracy is 70%. This means:
  - Among people who **have** the medical condition, the test turns out **positive 70%** of the time and negative 30% of the time.
  - Among people who **do not have** the medical condition, the test turns out **negative 70%** of the time and positive 30% of the time.
- A positive test result is always more likely for people who have the medical condition than for people who do not have it.
- You will then be told the specific person's test result.
- Your task is to guess the likelihood, between 0% and 100%, that the person who took the test actually has the medical condition.

Your bonus payment:

- While the medical condition, the test and the person are fictional, the quality of your guess actually affects your bonus payment.
- For each medical condition, test and a given result, there is a statistically correct guess.
- If you are chosen for a bonus, you will get $20 if your guess is within +/-3 percentage points of the statistically correct answer, and nothing otherwise.

---

15 guesses:

- In total, you will make a guess for each of **15 different medical conditions and people**.
- In each round, you will be told how common the medical condition is in the population, how accurate the test is, and what the person's test result is. You will then estimate the likelihood with which the person has the medical condition.
- Across rounds, the medical condition and test accuracy vary! You can think of the different rounds as concerning different medical conditions, tests and persons.
- These rounds are completely independent from one another. Only one of the 15 rounds will be randomly selected as the "round that counts" at the end of the study, and your estimate in that round determines your potential bonus payment. Each of your decisions is equally likely to be the "decision that counts."

Next

Figure 28: Instruction screen for the *Beliefs* paradigm.

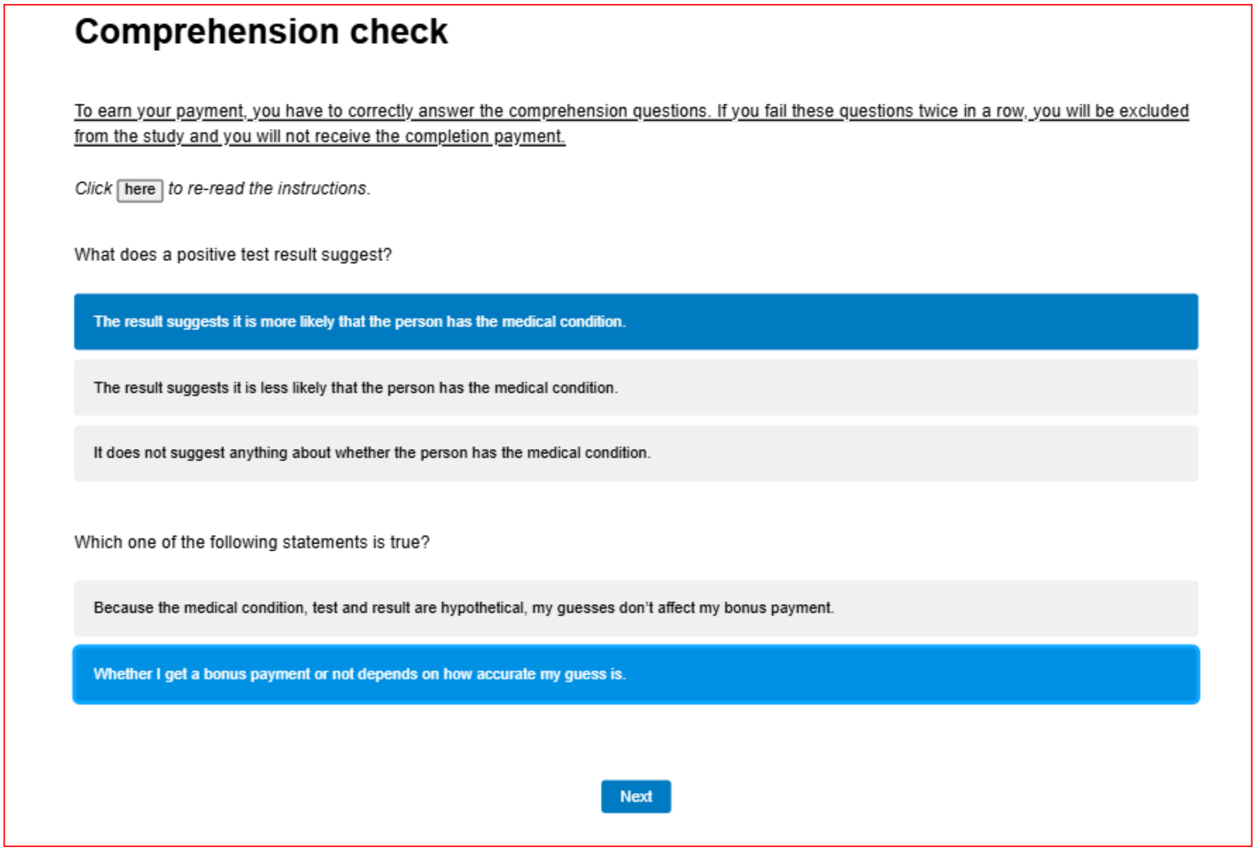


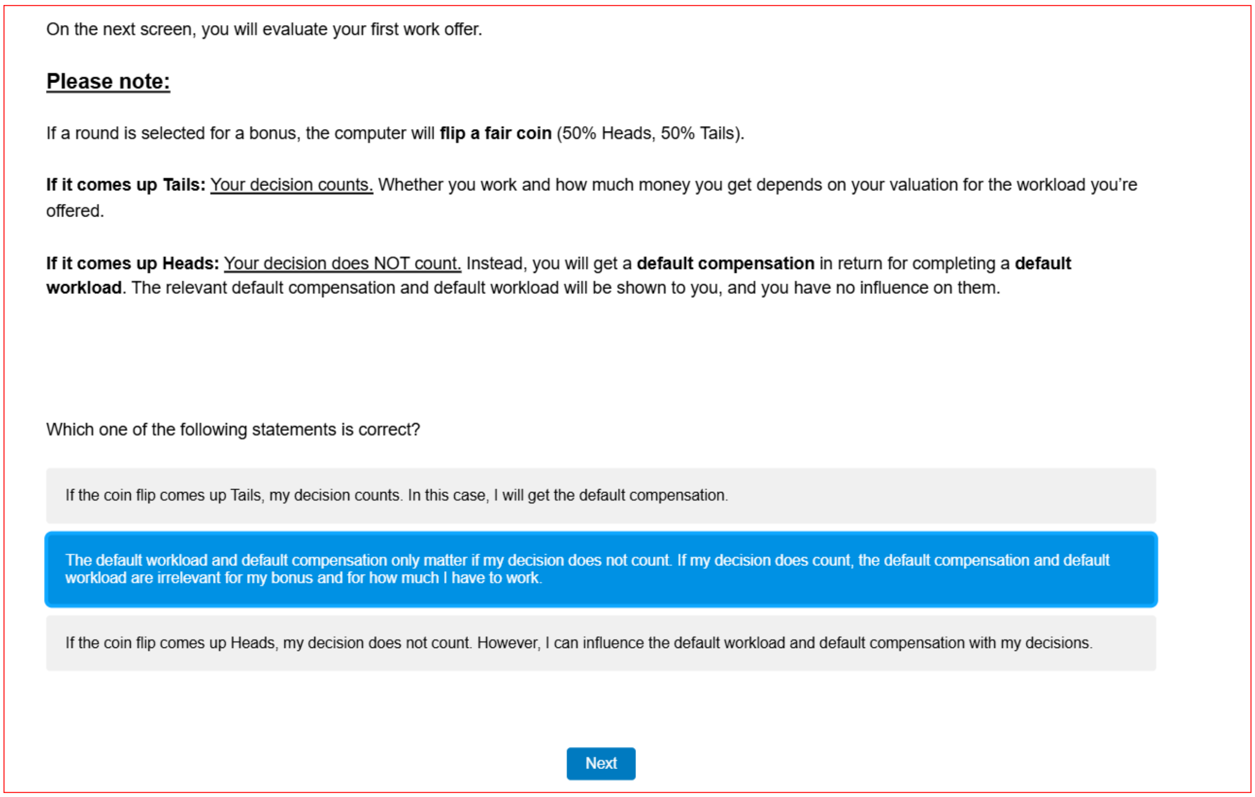
Figure 29: Comprehension-check screen for the *Beliefs* paradigm.

H.2 Comparison Point Announcement and Decision Screens

Here we show screenshots of the comparison point announcement and decision screens.

H.2.1 Effort Valuation

Treatment: Expectations CP



On the next screen, you will evaluate your first work offer.

Please note:

If a round is selected for a bonus, the computer will **flip a fair coin** (50% Heads, 50% Tails).

If it comes up Tails: Your decision counts. Whether you work and how much money you get depends on your valuation for the workload you're offered.

If it comes up Heads: Your decision does NOT count. Instead, you will get a **default compensation** in return for completing a **default workload**. The relevant default compensation and default workload will be shown to you, and you have no influence on them.

Which one of the following statements is correct?

If the coin flip comes up Tails, my decision counts. In this case, I will get the default compensation.

The default workload and default compensation only matter if my decision does not count. If my decision does count, the default compensation and default workload are irrelevant for my bonus and for how much I have to work.

If the coin flip comes up Heads, my decision does not count. However, I can influence the default workload and default compensation with my decisions.

Next

Figure 30: Treatment announcement — *Expectations CP* (*Effort Valuation*).

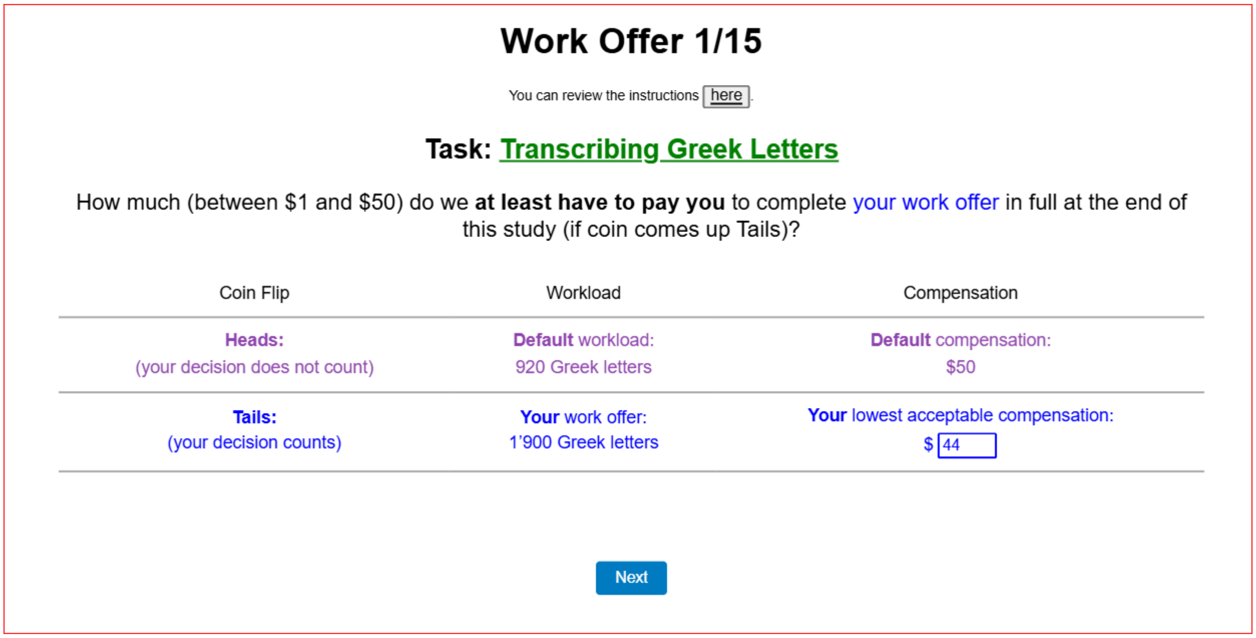


Figure 31: Decision screen — *Expectations CP* (*Effort Valuation*).

Treatment: Social Both CP

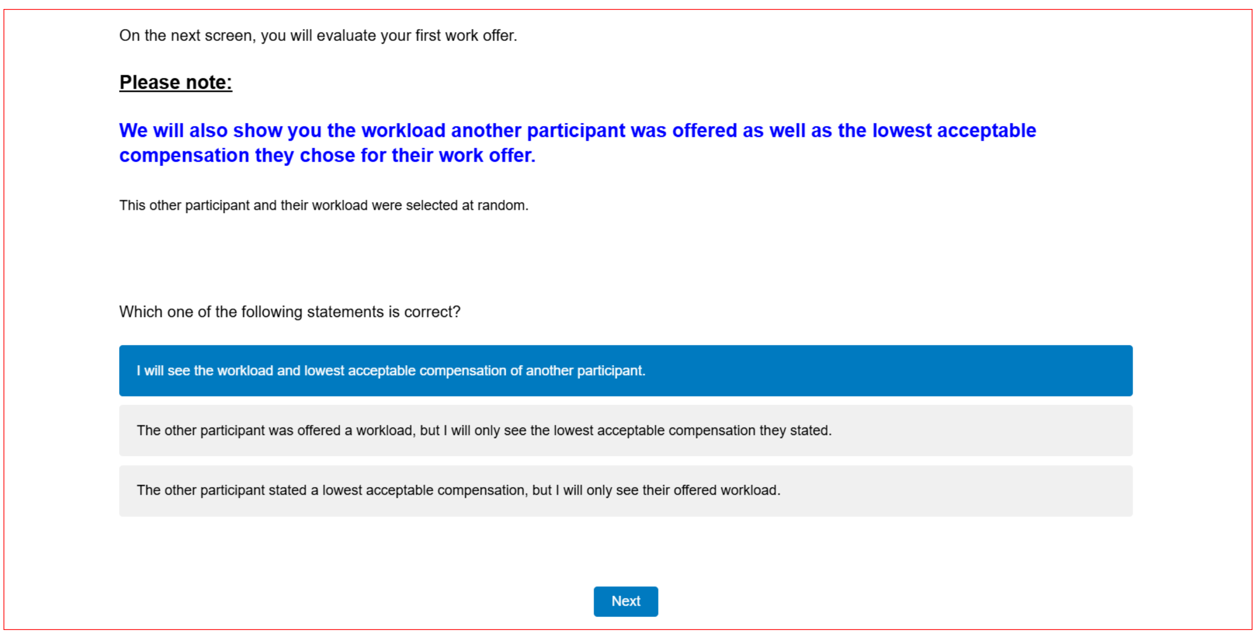


Figure 32: Treatment announcement — *Social Both CP*.

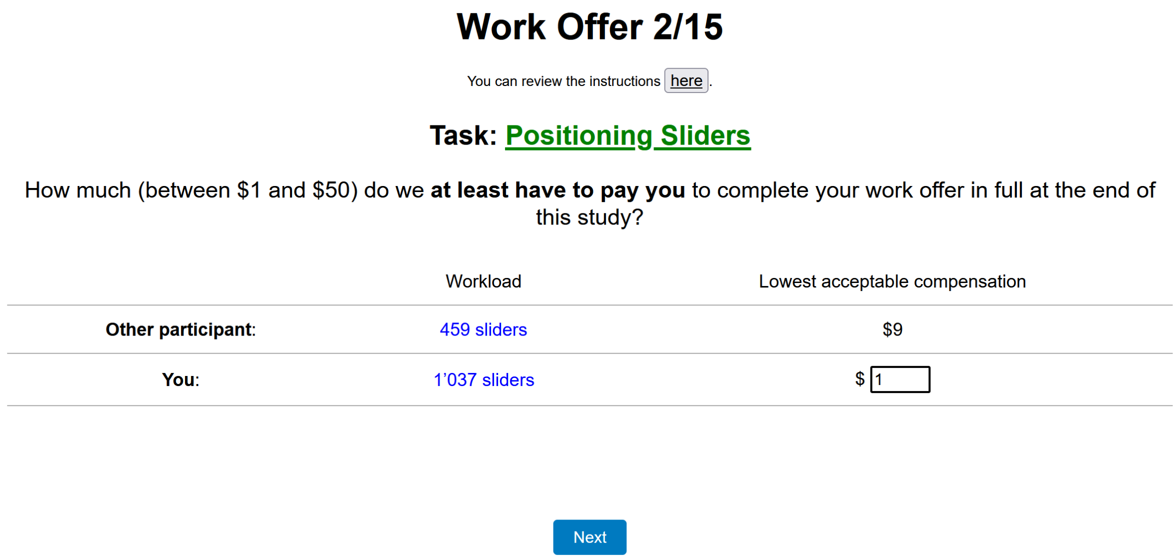


Figure 33: Decision screen — *Social Both CP* (*Effort Valuation*).

Treatment: Social Input CP

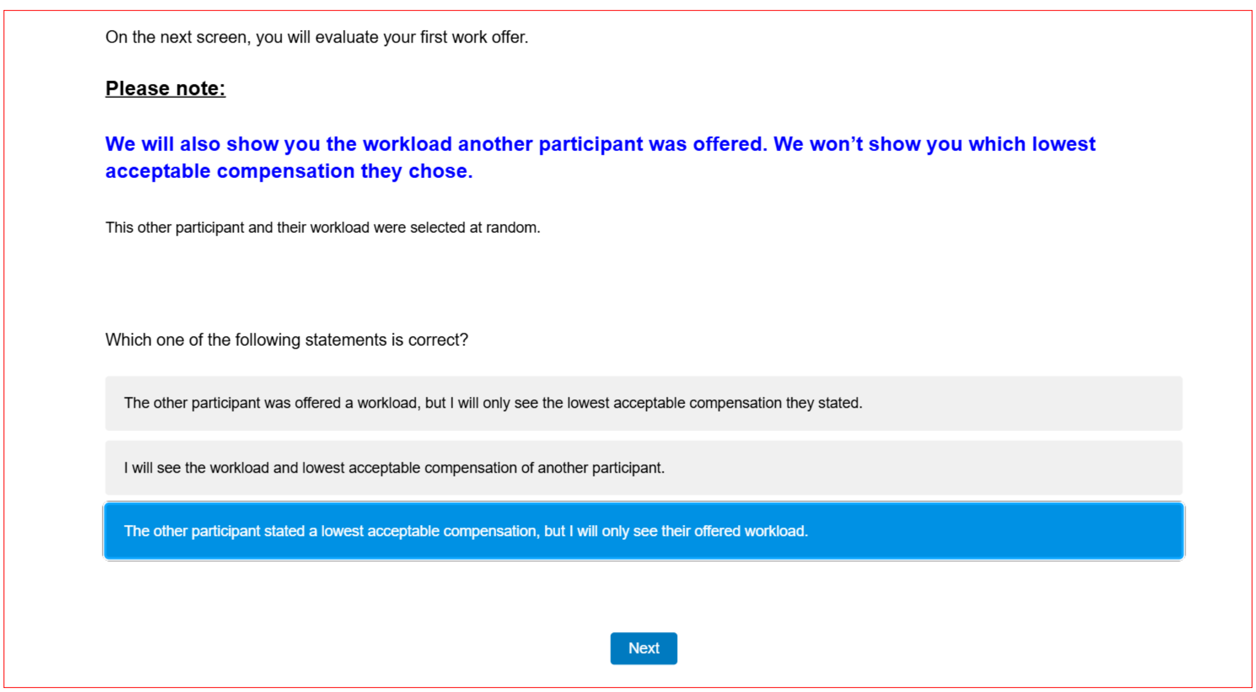


Figure 34: Treatment announcement — *Social Input CP*.

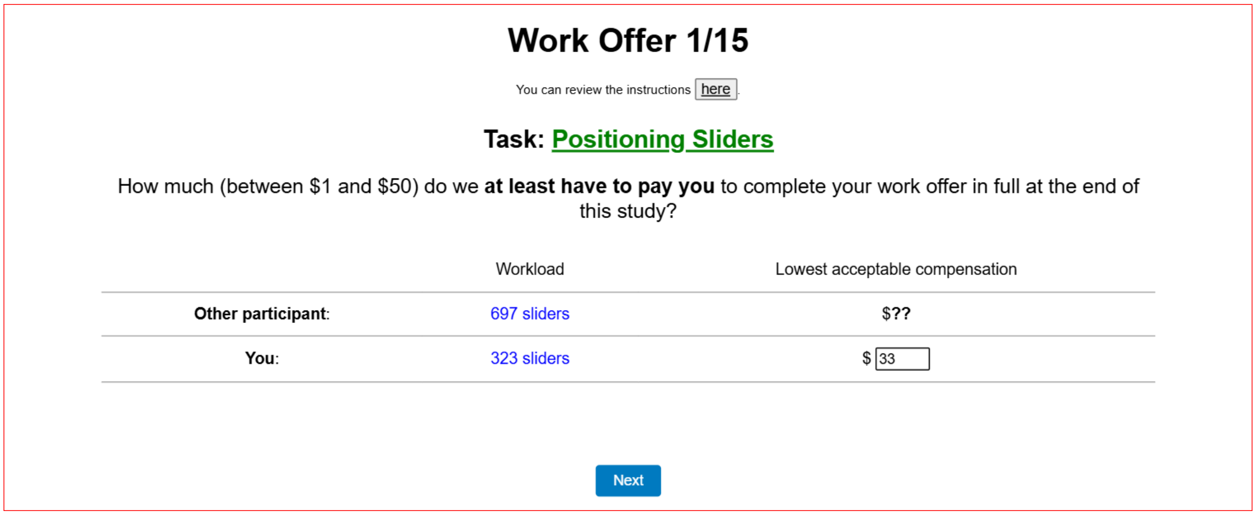


Figure 35: Decision screen — *Social Input CP* (*Effort Valuation*).

Treatment: Social Output CP

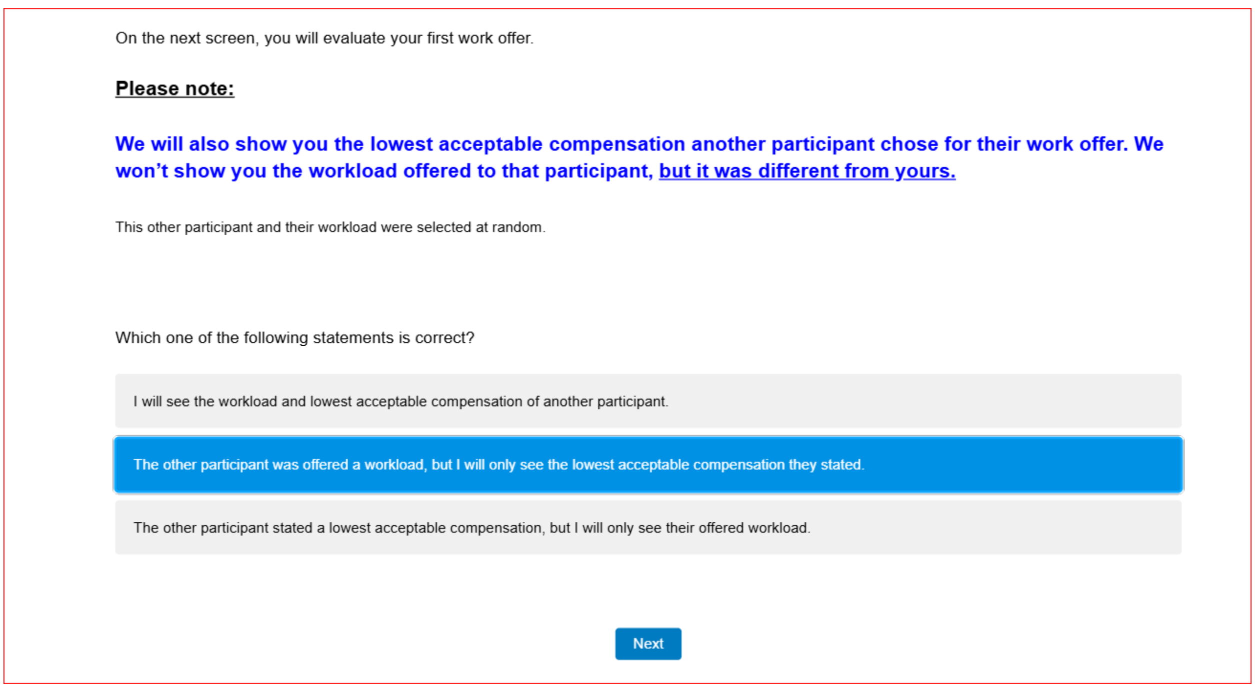
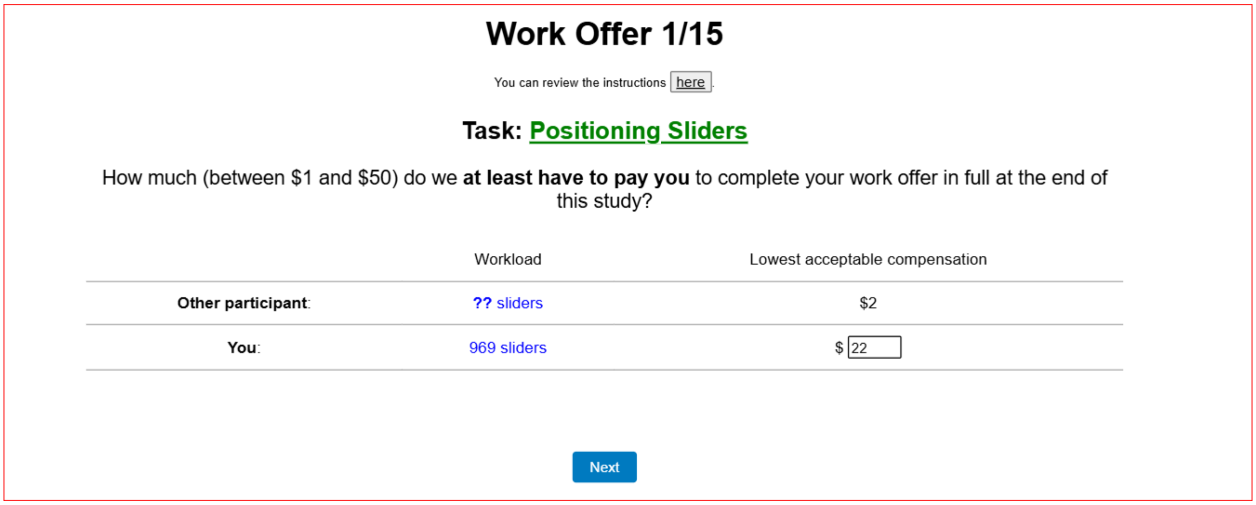


Figure 36: Treatment announcement — *Social Output CP*.



Work Offer 1/15

You can review the instructions here.

Task: Positioning Sliders

How much (between $1 and $50) do we **at least have to pay you** to complete your work offer in full at the end of this study?

	Workload	Lowest acceptable compensation
Other participant:	**??** sliders	$2
You:	969 sliders	$ 22

Next

Figure 37: Decision screen — *Social Output CP* (*Effort Valuation*).

Treatment: Time Treatment

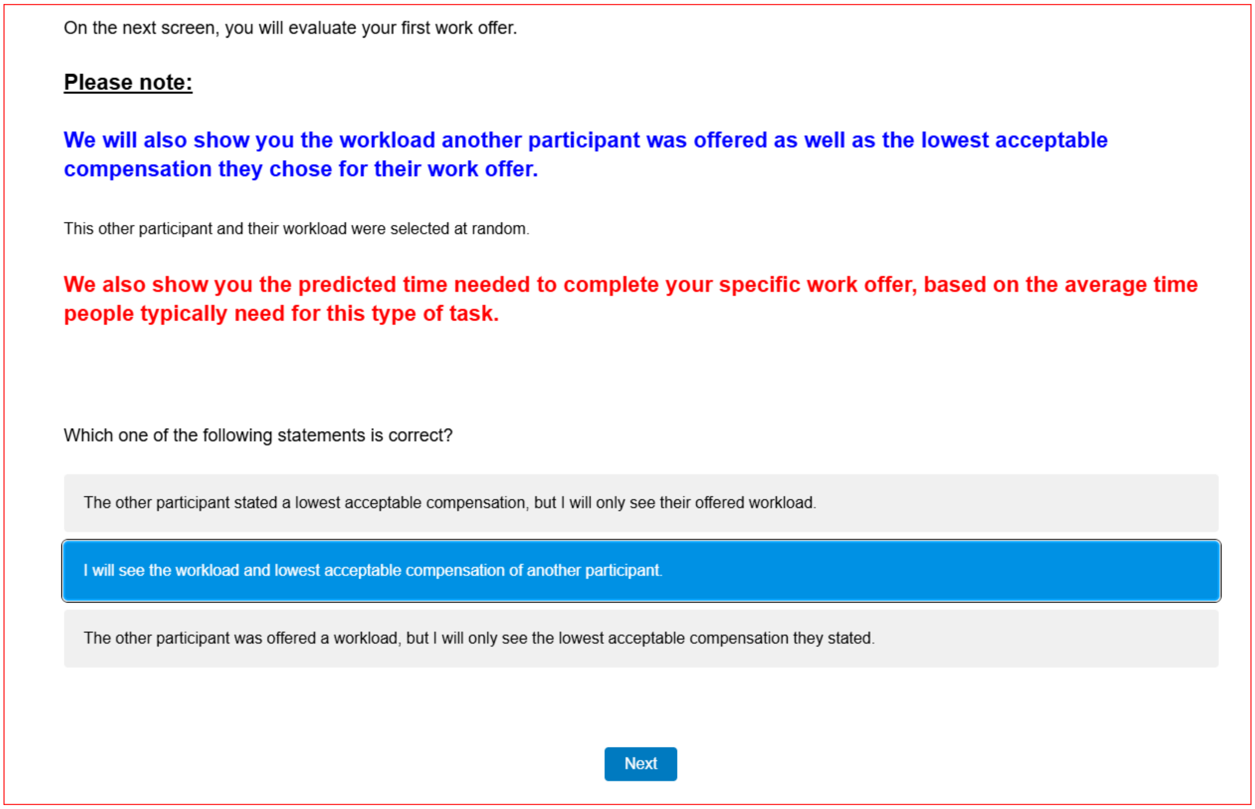


Figure 38: Treatment announcement — *Time* treatment.

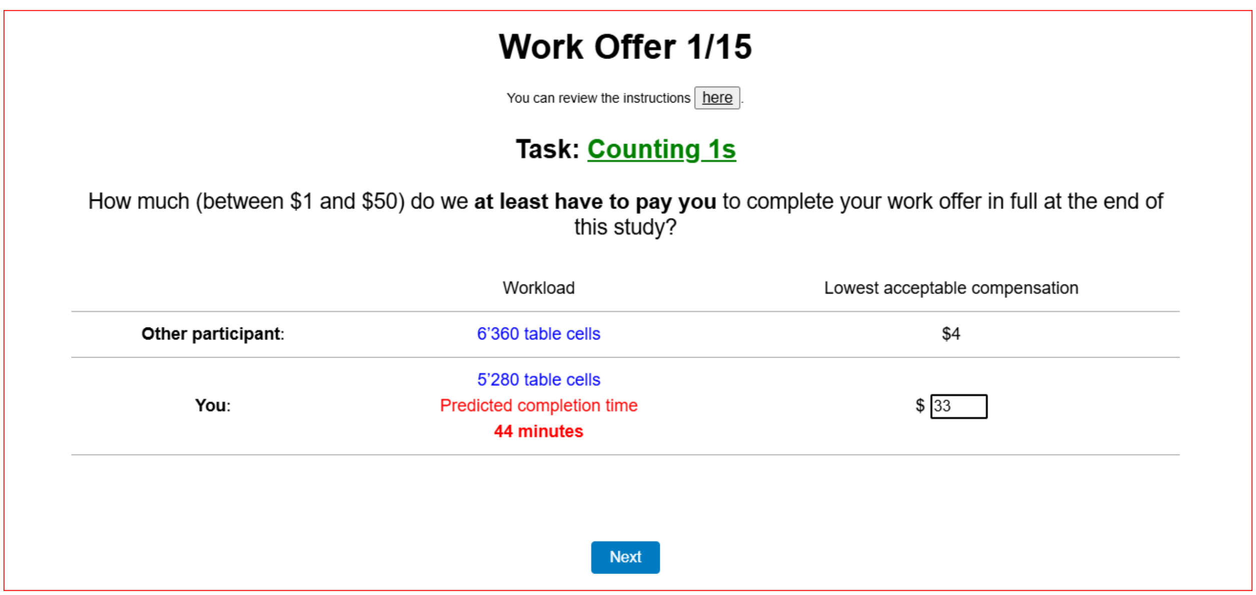


Figure 39: Decision screen — *Time* treatment (*Effort Valuation*).

Treatment: Informativeness

On the next screen, you will evaluate your first work offer. Please note:

If this round is selected for a bonus, the computer will **flip a fair coin** (50% Heads, 50% Tails).

If it comes up Tails: Your decision counts. Whether you work and how much money you get depends on your valuation for the workload you're offered.

If it comes up Heads: Your decision does NOT count. Instead, you will get a randomly drawn, **arbitrary compensation** in return for completing a randomly drawn, **arbitrary workload**. These two random numbers were determined arbitrarily and have no meaning for your decision, and you have no influence on them.

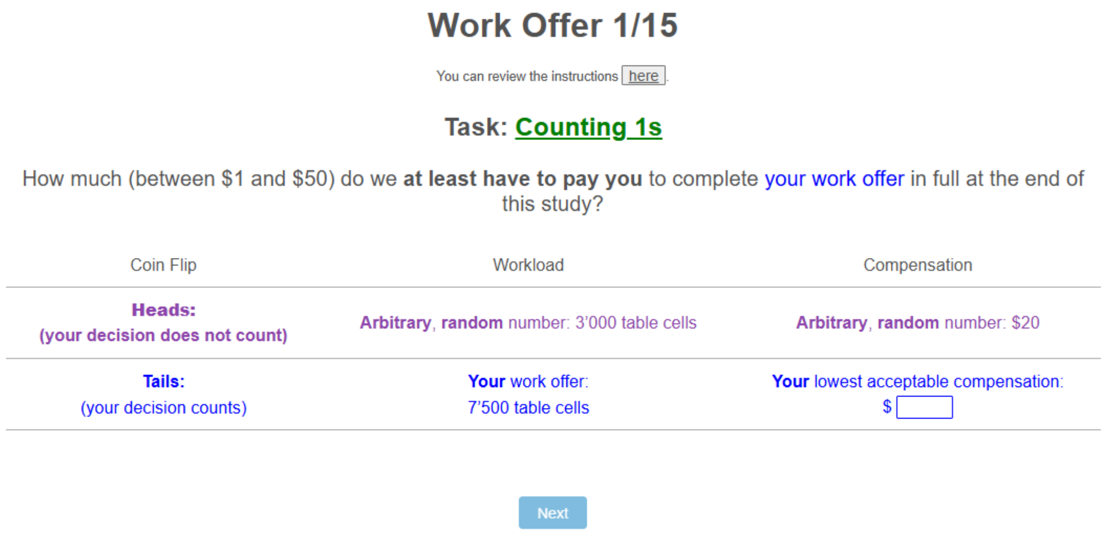
Which one of the following statements is correct?

- The arbitrary workload and arbitrary compensation only matter if my decision does not count. If my decision does count, the arbitrary compensation and arbitrary workload are irrelevant for my bonus and for how much I have to work.
- If the coin flip comes up Heads, my decision does not count. However, I can influence the arbitrary workload and arbitrary compensation with my decisions.
- If the coin flip comes up Tails, my decision counts. In this case, I will get the arbitrary compensation.

Please indicate which statement is correct:

- The arbitrary workload and compensation were chosen to indicate how much most participants usually work and receive.
- The arbitrary workload and compensation were generated at random and have no special meaning for my decision.

Figure 40: Treatment announcement — *Informativeness* treatment (*Effort Valuation*).


Work Offer 1/15

You can review the instructions here

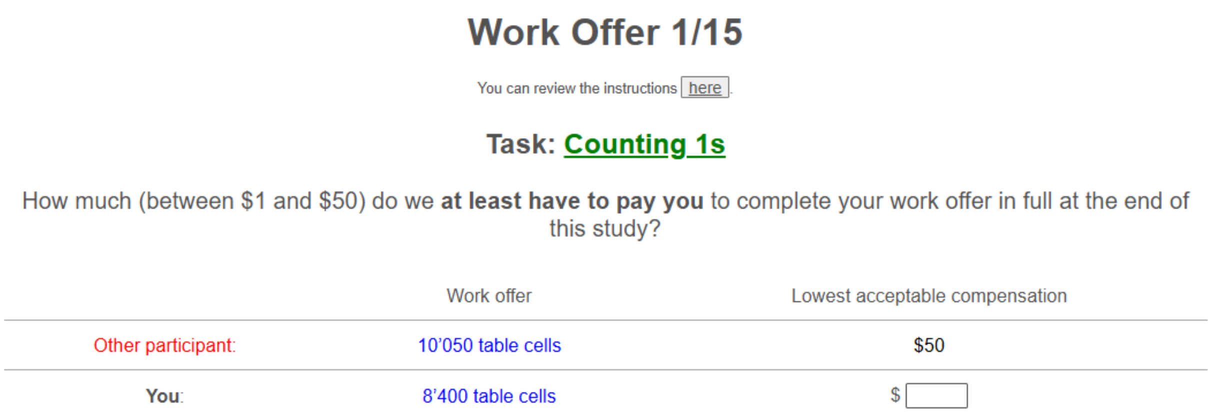
Task: Counting 1s

How much (between $1 and $50) do we **at least have to pay you** to complete your work offer in full at the end of this study?

Coin Flip	Workload	Compensation
Heads: **(your decision does not count)**	**Arbitrary**, **random** number: 3'000 table cells	**Arbitrary**, **random** number: $20
Tails: (your decision counts)	**Your** work offer: 7'500 table cells	**Your** lowest acceptable compensation: $

Next

Figure 41: Decision screen — *Informativeness* treatment (*Effort Valuation*).


Work Offer 1/15

You can review the instructions here

Task: Counting 1s

How much (between $1 and $50) do we **at least have to pay you** to complete your work offer in full at the end of this study?

	Work offer	Lowest acceptable compensation
Other participant:	10'050 table cells	$50
You:	8'400 table cells	$

Figure 42: Decision screen for counting ones — *Deliberation* treatment (*Effort Valuation*).

Figure 43: Decision screen for positioning sliders — *Deliberation* treatment (*Effort Valuation*).

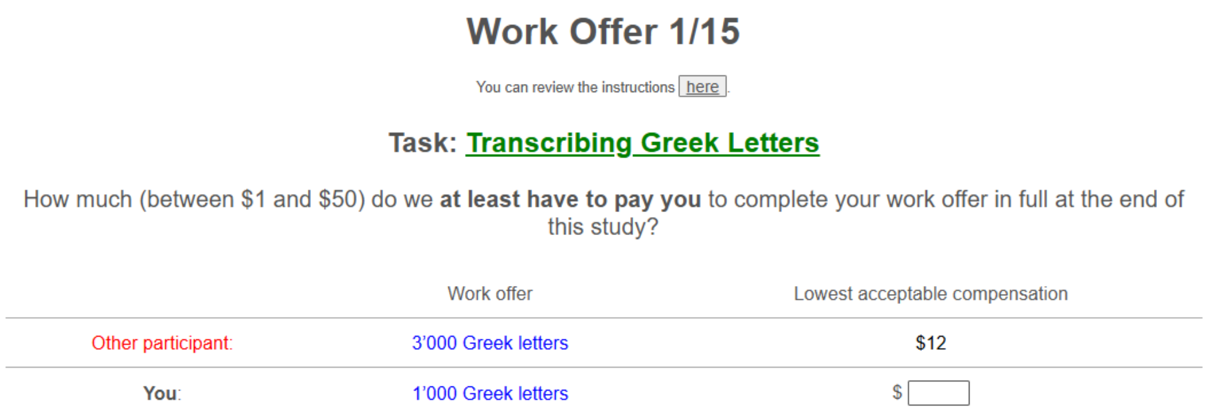


Figure 44: Decision screen for transcribing Greek letters — *Deliberation* treatment (*Effort Valuation*).

H.2.2 Effort Choice

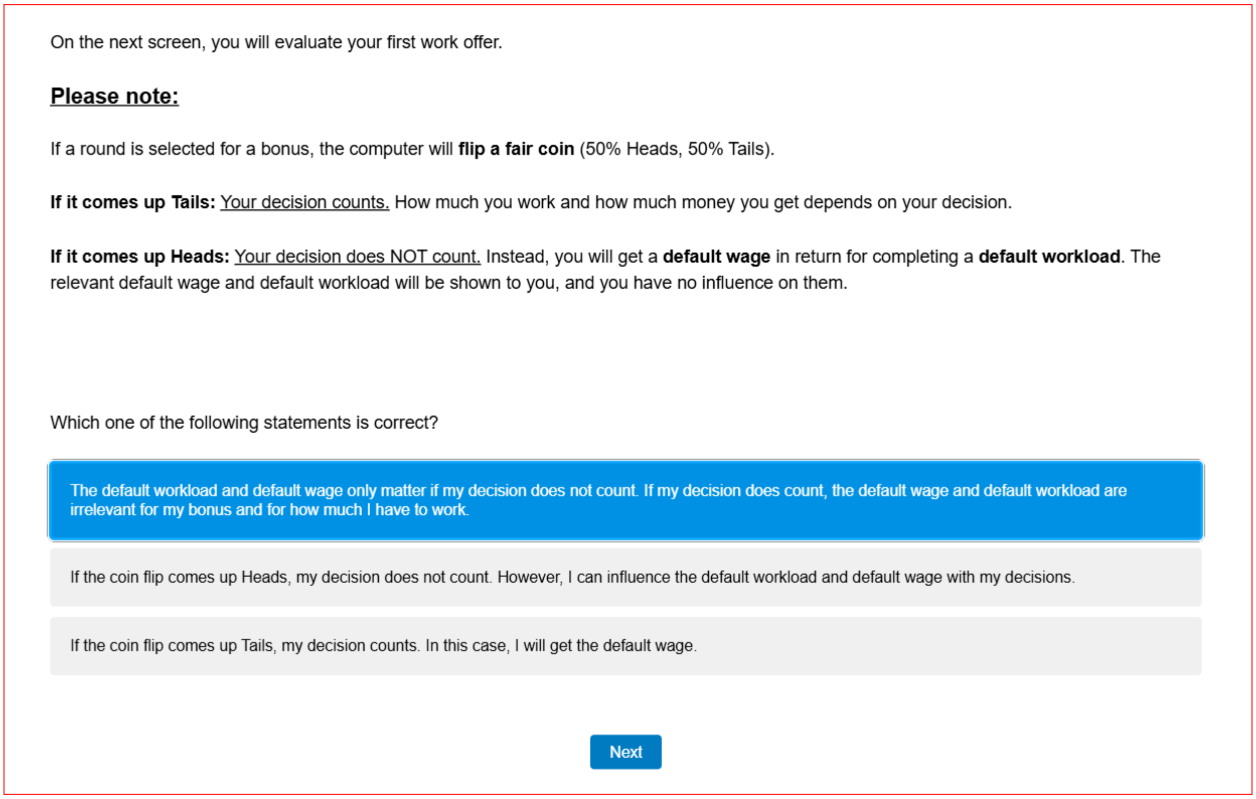


Figure 45: Treatment announcement screen (*Effort Choice*).

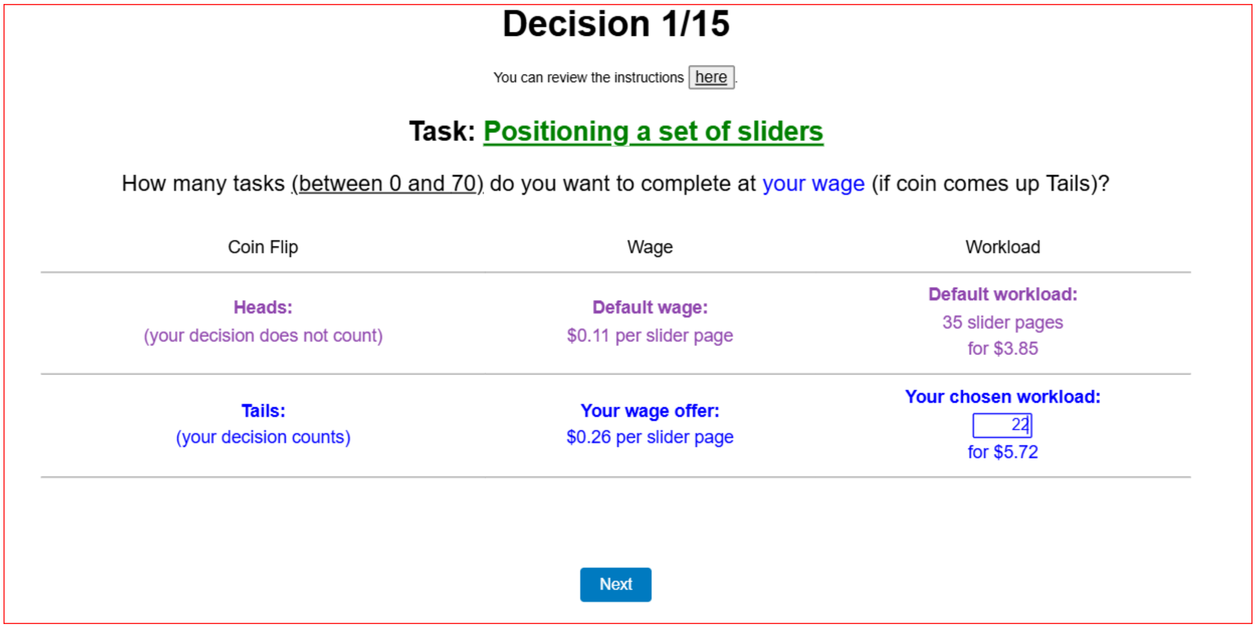


Figure 46: Decision screen (*Effort Choice*).

H.2.3 Investment

On the next screen, you will make your first decision.

Please note:

If a round is selected for a bonus, the computer will **flip a fair coin** (50% Heads, 50% Tails).

If it comes up Tails: Your decision counts. You decide how much to risk. Your wagered amount and your gamble – with the winning chance shown to you – will determine your bonus.

If Heads: Your decision is ignored. You have **no influence** on what happens in this case. The system will instead use a different, **imposed wagered amount** in an **imposed gamble** with a different winning chance.

If the coin lands **Heads**, which of the following is true?

- I still influence the outcome, since it's based on my wagered amount.
- **My decision is ignored, and the system imposes both the gamble and the wagered amount.**
- I see both the gamble and the wager that are imposed by the system.

If the coin lands **Tails**, which of the following is true?

- I get the imposed gamble, but I still control the wager.
- **My decision determines the amount wagered in my gamble.**
- I can revise my decision after seeing the coin flip.

Which one of the following statements is true?

- If the coin comes up Heads, the wager I choose also affects how much money is wagered in the computer-imposed gamble.
- **If the coin comes up Heads, the wagered amount I choose has no effect on how much money is wagered in the computer-imposed gamble. Instead, the computer-imposed wager, which is independent of and potentially different from my own wager, determines my bonus.**

Next

Figure 47: Treatment announcement screen (*Investment*).

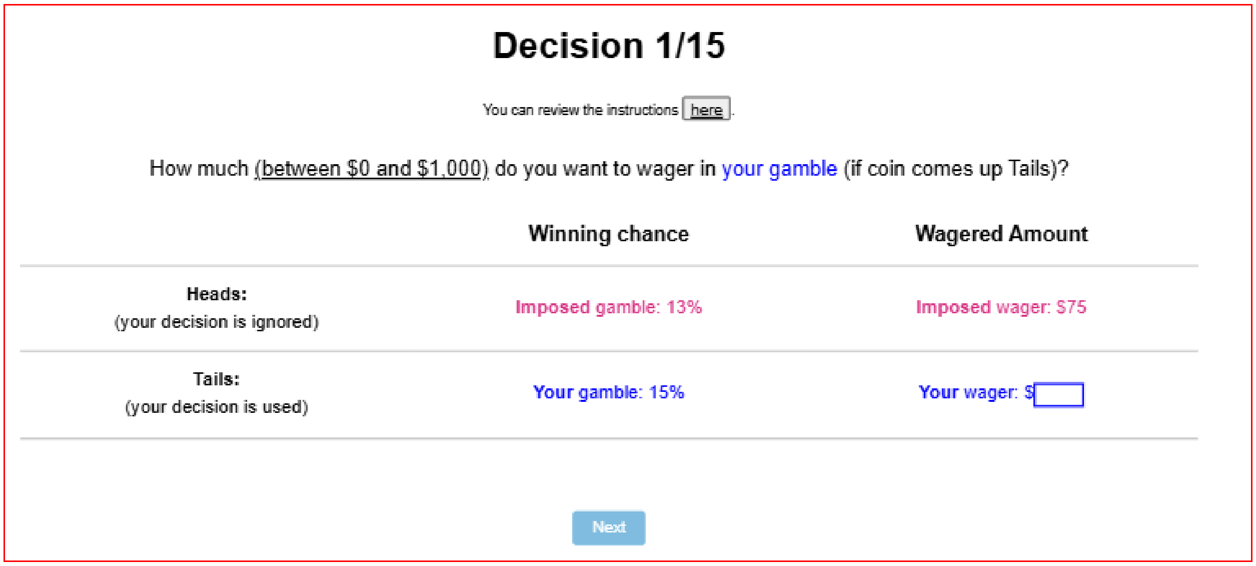


Figure 48: Decision screen — *Expectations CP* treatment (*Investment*).

H.2.4 Beliefs

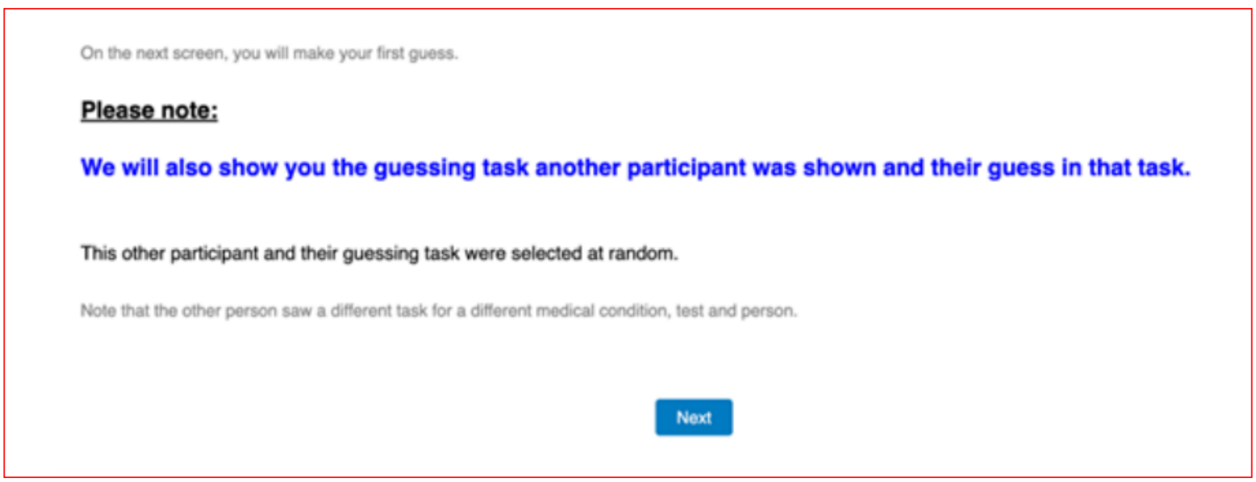


Figure 49: Treatment announcement screen (*Beliefs*).

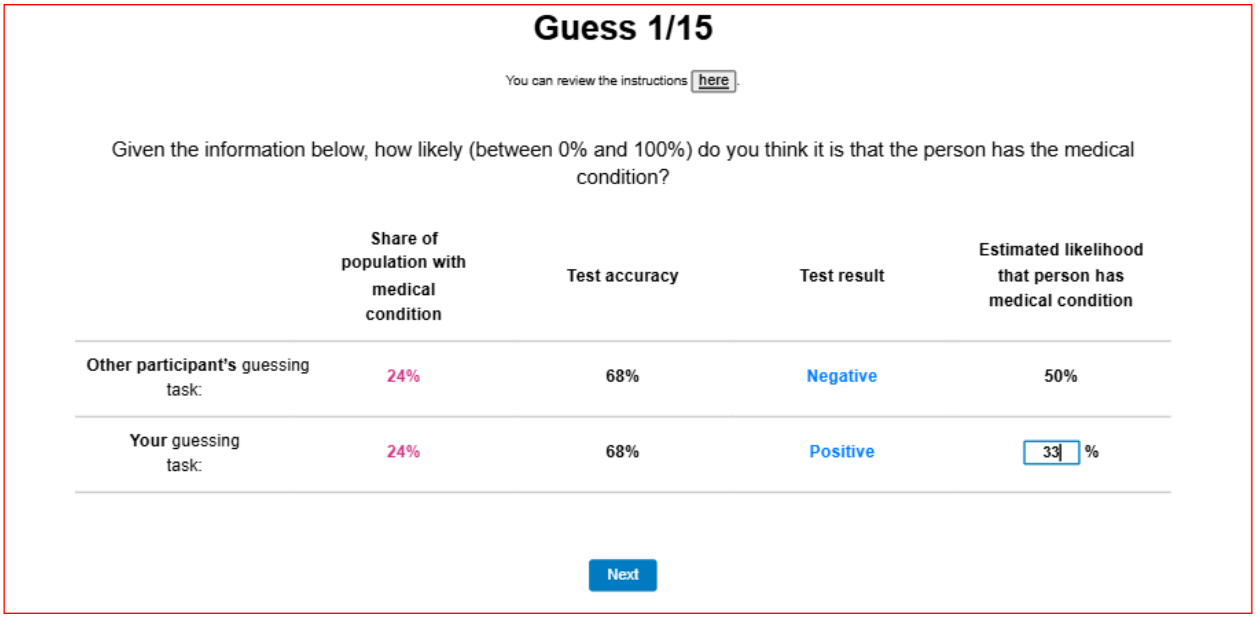


Figure 50: Decision screen — *Social CP* treatment (*Beliefs*).

H.3 Forecaster Survey (Social Science Prediction Platform)

Study: Understanding Comparison Effects

Introduction: Much research on decision-making studies how decisions depend on a decision parameter, such as: lottery valuations as a function of the winning chance, effort supply as a function of the wage, or belief updating as a function of signal accuracy, among many others. Assume that in a regression of decisions on such a decision parameter, the effect is significantly positive.

A large body of empirical research documents that decisions also respond to various types of **comparison points**, such as reference points, goals, anchors, norms, expectations, social comparisons etc.

Let's suppose we run a regression of decisions on the decision parameter (which has a positive effect) and on the comparison point provided in an experiment. The estimated coefficient on the comparison point may turn out to be **near zero, positive, or negative**.

Prediction 1: Think about all the different possible comparison points you can imagine. Do you have a **general rule for predicting the comparison effect** in a regression of decisions on the decision parameter and the comparison point?

Put differently, do you have a **general prediction for which types of comparison points create a positive, negative or no effect?**

Yes

No

What is your rule for predicting whether a comparison effect **creates a positive, negative or no effect?**

Figure 51: Forecaster survey introduction and general rule prompt (Prediction 1).

A payment decision

Participants in this experiment state the **minimum compensation** they require to **complete a given workload** of a tedious task (positioning sliders on a screen, counting numbers in a table, or transcribing blurry Greek letters).

In addition, subjects are shown a comparison point. Specifically, when making their decision they are **also shown the workload offered to another randomly-selected participant**. You can see an example screenshot here.

We run the following OLS regression (on N=3735 decisions from 249 participants):

$$\text{Required minimum compensation} = \alpha + \beta_1 \times \text{Offered workload} + \beta_2 \times \text{Workload offered to another participant} + \varepsilon$$

Prediction 2. Please predict the **coefficient of the comparison point:** β_2.

Regressors	Estimated coefficient
β_1 Workload:	Significantly positive
β_2 Another participant's workload:	**Your prediction:**

Figure 52: Effort Valuation: payment decision prompt – prediction for the input comparator.

A payment decision

Participants in this experiment state the **minimum compensation** they require to **complete a given workload** of a tedious task (positioning sliders on a screen, counting numbers in a table, or transcribing blurry Greek letters).

In addition, subjects are shown a comparison point. Specifically, when making their decision they are **also shown the minimum compensation another randomly-selected participant stated**. The other participant was potentially offered a different workload, and participants do not see which workload the other participant was offered. You can see an example screenshot here.

We run the following OLS regression (on N=3690 decisions from 246 participants):

$$\text{Required minimum compensation} = \alpha + \beta_1 \times \text{Offered workload} + \beta_2 \times \text{Minimum required compensation of other participant} + \varepsilon$$

Prediction 2. Please predict the **coefficient of the comparison point:** β_2.

Regressors	Estimated coefficient
β_1 Workload:	Significantly positive
β_2 Minimum required compensation of other participant:	**Your prediction:**

Not significantly different from zero
Significantly positive
Significantly negative
I don't know

Next

Figure 53: Effort Valuation: payment decision prompt – prediction for the output comparator.

A work decision

Participants in this experiment decide **how many rounds** of a tedious task they want to complete (positioning sliders on a screen, counting numbers in a table, or transcribing blurry Greek letters) for **a given piece-rate payment per round**.

In addition, subjects are shown comparison points. Specifically, when making their decision they are **also shown (i) a default piece-rate and (ii) a default number of rounds to be completed.** Participants are told that with 50% chance their own decision counts, meaning that they work the number of rounds they've chosen and receive the corresponding payment. With the remaining 50% chance, the participants' decision does not count. Instead, they have to complete a default number of rounds (that's known to them) in exchange for a default piece rate. You can see an example screenshot here.

We run the following OLS regression (on N=3690 decisions from 246 participants):

$$\text{Rounds choice} = \alpha + \beta_1 \times \text{Piece rate} + \beta_2 \times \text{Default piece rate} + \beta_3 \times \text{Default number of rounds} + \varepsilon$$

Prediction 3. Please predict the **coefficients of the comparison points**: β_2 and β_3.

Regressors	**Estimated coefficient**
β_1 Piece rate:	Significantly positive
β_2 Default piece rate:	**Your prediction:**
β_3 Default number of rounds:	**Your prediction:**

Not significantly different from zero
Significantly positive
Significantly negative
I don't know

Figure 54: Effort Choice: work decision prompt with default piece rate and default rounds (Prediction 3).

A guessing task

Participants are asked to **guess the likelihood that a hypothetical patient has a medical condition**, given the **prior likelihood of the condition**, the **accuracy of a test**, and the **test result**.

In addition, subjects are shown comparison points. Specifically, when making their decision they are **also shown (i) the task a randomly-selected other participant was shown** (prior likelihood of condition, test accuracy and test result) **and (ii) the other participant's guess**. The test accuracy in the other participant's problem was always the same as in their own guessing task, but the prior likelihood of the disease and the test result could differ. See an example screenshot here.

We run a regression on 3610 decisions from 246 participants where we transform the subjects' guesses into log-odds format (i.e., log(belief/(1−belief))), the prior into log-odds, and the test into a log likelihood ratio (canonical specification following Grether, 1980):

Log odds of stated guess = α + β_1 × Log prior likelihood of disease + β_2 × Log likelihood ratio of test + β_3 × Log prior likelihood of disease in other participant's task + β_4 × Log likelihood ratio of test in other participant's task + β_5 × Log odds of other participant's stated guess + ε

Prediction 4. Please predict the **coefficients of the comparison points**: β_3, β_4, and β_5.

Regressors	**Estimated coefficient**
β_1 Log prior likelihood of disease:	Significantly positive
β_2 Log likelihood ratio of test:	Significantly positive
β_3 Log prior likelihood of disease in other participant's task:	**Your prediction:** Not significantly different from zero Significantly positive Significantly negative I don't know
β_4 Log likelihood ratio of test in other participant's task:	**Your prediction:**
β_5 Log odds of other participant's stated guess:	**Your prediction:**

Figure 55: Beliefs: guessing task screen (Prediction 4).

An investment decision

Participants in this experiment decide **how much of a $1,000 budget to wager** on a risky gamble that **doubles the investment** with a certain **winning chance**, and **halves the investment** with the **remaining probability**.

In addition, subjects are shown comparison points. Specifically when making their decision they are also shown **(i) a different winning chance of another, "imposed" gamble, and (ii) an "imposed" amount of $100 to be invested in that imposed gamble**. Participants are told that with 50%, their own decision counts, meaning that their own investment decision determines their payoff. With the remaining 50% chance, the participant's decision does not count. Instead, the imposed amount will be wagered on the imposed gamble with the other winning chance on their behalf. You can see an example screenshot here.

We run the following OLS regression (2705 decisions from 247 participants):

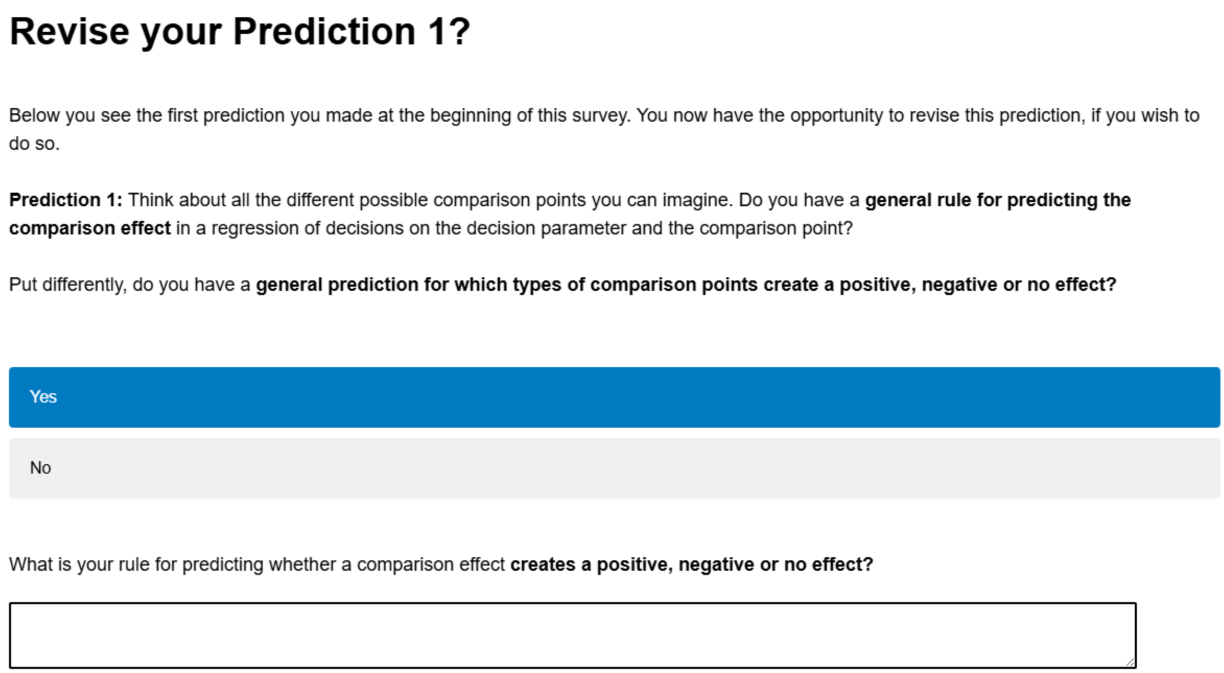
$$\text{Amount wagered} = \alpha + \beta_1 \times \text{Winning chance} + \beta_2 \times \text{Winning chance of imposed gamble} + \beta_3 \times \text{Imposed amount to be invested in imposed gamble} + \varepsilon$$

Prediction 5. Please predict the **coefficients of the comparison points**: of β_2 and β_3.

Regressors	Estimated coefficient
β_1 Winning chance:	Significantly positive
β_2 Winning chance of imposed gamble:	**Your prediction:**
β_3 Imposed amount to be invested in imposed gamble:	**Your prediction:**

Not significantly different from zero
Significantly positive
Significantly negative
I don't know

Figure 56: Investment: decision prompt with imposed gamble and imposed amount (Prediction 5).



Revise your Prediction 1?

Below you see the first prediction you made at the beginning of this survey. You now have the opportunity to revise this prediction, if you wish to do so.

Prediction 1: Think about all the different possible comparison points you can imagine. Do you have a **general rule for predicting the comparison effect** in a regression of decisions on the decision parameter and the comparison point?

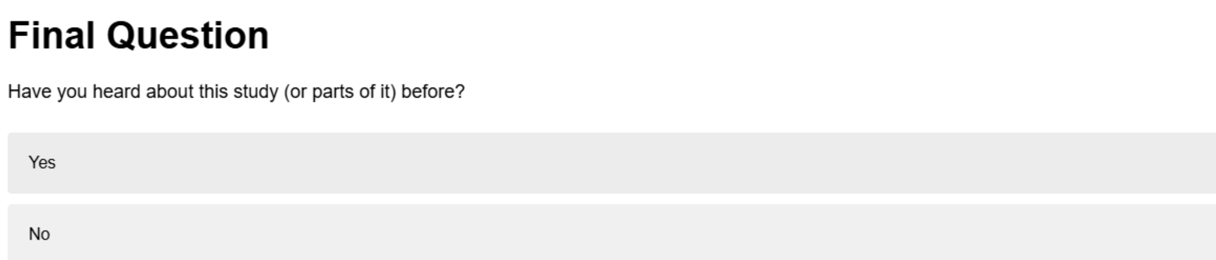
Put differently, do you have a **general prediction for which types of comparison points create a positive, negative or no effect?**

Yes

No

What is your rule for predicting whether a comparison effect **creates a positive, negative or no effect?**

Figure 57: Opportunity to revise the general rule prediction (Prediction 1).



Final Question

Have you heard about this study (or parts of it) before?

Yes

No

Figure 58: Final prior exposure question.