Comparisons*

Thomas Graeber ℗ Benjamin Enke

August 15, 2026

Abstract

Economic decisions are often shaped by comparison points such as reference points, 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 absolute evaluations—translating a decision input into an output—are difficult, people use comparison points as information about how to navigate tradeoffs. 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 experimentally reduced and when comparators are less informative, consistent with a mechanism of comparisons as information. Finally, a classification of the prior literature suggests that the input-vs.-output taxonomy organizes the sign of comparison effects across more than 100 experimental and field observational studies.

*Dustin Fichmann and Felix Mao provided outstanding research assistance. We thank Botond Köszegi, Nick Netzer, Alex Rees-Jones, Michael Woodford as well as seminar audiences in Berlin, Bergen, Berkeley, Bonn, Boston University, the CESifo Behavioral Economics Conference 2025, Chicago Booth, the Florence Economics Conference 2026, Harvard, MIT Sloan, Wharton, the Zurich Cognitive Economics Workshop 2026 as well as the Zurich Workshop in Economics and Psychology 2025 for useful feedback and discussions. The research described in this article was approved by the Institutional Review Board at Harvard University. 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: Massachusetts Institute of Technology, Department of Economics, and NBER, benjamin.enke@gmail.com.

1 Introduction

A large body of empirical work in economics and other fields shows that decisions are often influenced by comparisons with objects such as reference points, goals, targets or anchors: quantities that we refer to as *comparison points*. What is the 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 (Allen et al., 2017; Corgnet et al., 2015; 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 from a basic premise: that absolute judgments are often difficult and uncertain, and that people partly (at least implicitly) rely on comparison points to acquire additional information about how to navigate the tradeoffs involved. For example, it is arguably non-trivial to mentally map a known hourly wage into a utility-maximizing number of daily work hours, and people might partly rely on their typical daily earnings or on their peer's hourly wage to infer what a

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

reasonable decision might be.

To study these issues, it is useful to distinguish between comparators 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, a previous wage or a peer's wage both represent *input comparators*. Decision outputs are endogenous variables that are implicitly or explicitly controlled by the decision-maker, e.g., the number of hours worked or total earnings. Hence, the number of hours worked on a previous day or typical past earnings represent *output comparators*.[2]

The main hypothesis of this paper is that input comparators produce contrast effects, while output comparators produce assimilation. To illustrate the underlying logic, 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 $a^*(w) = \beta w$, yet I don't know β, for example because I'm uncertain how painful the task is or because I have uncertainty about my consumption-leisure preferences. 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 β. 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). Intuitively, output comparators provide information about a reasonable 'ballpark' of the decision (producing assimilation), while input comparators provide information about the direction in which one needs to adjust (producing contrast). This hypothesis does not require that comparison points are objectively informative or even 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.

We structure our investigation around six hypotheses. 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.

[2]In standard economics modeling parlor, inputs correspond to parameters (for the decision at hand), while outputs are either decision variables or outcomes that are influenced by the decision variable.

Baseline evidence. Our experimental strategy is to separately induce and vary input and output comparators. In our *Effort Valuation* paradigm, subjects 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 stated minimum compensation the output. In one condition, participants make their decisions while shown the workload offered to another participant along with that participant's WTA decision. These correspond to social variants of input and output comparators, respectively. In another condition, the comparison points are 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). In additional treatments, we find that the opposite effects of input and output comparators also arise when only one of the two comparators is displayed.

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 asset and a risky gamble. The gamble's winning chance is the decision input, and the amount wagered is the output. We again 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).

Finally, 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 base rate and a diagnostic test result. Again, we find that stated posteriors decrease in input comparators (e.g. the test result observed by another subject) and increase in the output comparator (another subject's stated posterior). This suggests that the input-output comparator distinction predicts the sign of comparison effects also in objective tasks in which preferences over comparisons (such as reference-dependent utility or social emotions like envy) are plausibly ruled out.

Turning input comparator into output 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 translate 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 leads to higher chosen workloads, in stark contrast to the *Effort Valuation* paradigm, in which it produces lower minimum required compensations. Thus, what matters for the sign of a comparison effect is not predominantly what the comparison variable is but whether it constitutes an input or output comparator for the decision at hand.

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

To test this idea, we increase subjects' deliberation about the input-decision map by asking them to state a full 'policy function' (i.e., to state their hypothetical WTA for a large and ordered set of workloads) prior to making their incentivized choices. 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.

Finally, we also implement a treatment in which we transparently reduce the informativeness of the comparison points. This substantially reduces the magnitude of both comparison effects (but does not eliminate them), again consistent with a mechanism of comparison points as implicit sources of information.

Classifying results of prior studies. To investigate the broader validity of our input-vs.-output-comparator hypothesis, 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 is inevitably an imperfect exercise given the heterogeneity of the studies involved, the presence of potential confounds, and the inherent subjectivity involved in classification exercises.

With those caveats in mind, we find that the results of a large 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 constitute an input comparator for actual wages, and—consistent with our hypothesis—consistently exert 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 likewise 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 generally found when the comparison point lives in the output domain, i.e., when it constitutes an endogenous 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 these contexts, decisions increase in the comparison point, consistent with our hypothesis of output comparators producing assimilation effects.

Overall, of the 109 papers we are able to classify, 102 (94%) produce results consistent with our taxonomy; among the 107 papers whose primary comparator test yields a directional effect, the share is 95%. Given the ambiguity and subjectivity involved in the classification, we view this number as an encouraging ballpark estimate rather than as a precise test statistic.

Forecaster survey. To get a sense of whether the input-output taxonomy reflects knowledge already held by the profession, we implement a forecaster survey on the *Social Science Prediction Platform* with 105 social science researchers. We ask our participants to predict the sign of comparison effects of different experiments, and to indicate whether they have a general rule for making such predictions. The main result is that the forecasters often make incorrect predictions, and none of the forecasters describes a rule resembling the input-output distinction we emphasize. These results suggest to us that our taxonomy was not previously recognized.

Limit(ation)s. As with any hypothesis, other things matter, too, and some effects go against our taxonomy. First, in field settings, input and output comparators are often correlated (e.g. hourly wage and total earnings), so that the estimated effect of an observed comparator can be spuriously positive, negative, or zero depending on the relative strength of the two channels.

Second, several other types of comparison effects are known to exist. For instance, our taxonomy is orthogonal to assimilation effects resulting from minimum detectable differences and stimulus uncertainty (e.g., Wu, 2024), or from memory- and categorization-based selective attention (e.g., Bordalo et al., 2020, 2025). Similarly, as discussed in greater detail below, our approach cannot explain comparison effects that reflect that people have incorrect mental models of the data-generating process, such as in the gambler's fallacy. Finally, our work is orthogonal to (and compatible with) people *also* exhibiting preferences over comparisons, such

as in models of reference-dependent preferences or of social emotions such as envy. Below we discuss under which assumptions some of our results can be explained by (unusual variants of) classic reference-dependence models, and why many others cannot.

Our account also struggles to explain the effects of uninformative comparison points, as in anchoring studies or when endowments are transparently randomized. One possibility is that people are accustomed to somewhat informative comparison points, and may thus heuristically use uninformative comparison points when caught in an unfamiliar experimental environment.

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 conflicting 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 predict the sign of comparison effects.[3] 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 a role for uncertainty in driving the 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; Dean et al., 2026).[4] Related, various papers have documented cognitive or framing effects in reference dependence (e.g. Gneezy and Potters, 1997; Ariely et al., 2003; Imas, 2016). For example, comparison effects sometimes weaken with expertise (List, 2003, 2004) and increase in decision uncertainty or complexity (Hsee, 1996; Arieli et al., 2011; Bondi et al., 2025). More generally, however, most of the recent literature on decision and value uncertainty is largely silent on comparison effects (e.g., Khaw et al., 2021; Gabaix, 2019; Frydman and Jin, 2021; Charles et al., 2024; Enke and Graeber, 2023; Bastianello and Imas, 2025; Augenblick et al., 2025).

Finally, our paper connects to work in perceptual psychology. In sequential magnitude estimation experiments, a recent result is that subjects' estimates decrease in the comparison

[3]While we emphasize that comparisons help in navigating tradeoffs and interpreting units, the literature on attention allocation posits that comparisons allow decision makers to learn which dimensions to attend to (e.g., Bordalo et al., 2012; Kőszegi and Szeidl, 2013).

[4]Broadly related are also models in which decision makers learn from other options (e.g., Kamenica, 2008; Bhui and Xiang, 2021; Shubatt and Yang, 2023).

stimulus but increase in the response to the comparison stimulus (e.g. Moon and Kwon, 2022; Imhoff and Barker, 2023; Sadil et al., 2024). Our paper documents analogous effects in economic choice. Related, Izakson et al. (2024) separate the influence of context into a starting-point channel and an evidence-accumulation channel, and find that contextual values pull valuations through the starting point, producing assimilation.

Section 2 lays out a stylized framework. Sections 3 and 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 them. Our exposition follows textbook definitions from social and perceptual psychology (e.g. Wedell et al., 2007).

Consider a problem of the form $\max_{a\in\mathbb{A}} u(a,x;\beta)$, 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 $a^*(x,\beta) = \arg\max_{a\in\mathbb{A}} u(a,x;\beta)$, where β captures all factors that influence how the DM translates a 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 $a^*(x,\beta)$ is increasing in x.

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

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

Equation (1) is an accounting identity, not a behavioral model: any observed decision can be written as the comparison-free benchmark $a^\star$, plus a term proportional to the gap between the comparison point and that benchmark, plus other biases. Under our normalization that a^* increases in x, assimilation means $w>0$. Reading w and u as invariant to c, this implies 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.

2.2 Decision Uncertainty and Comparison Points as Information

Underlying our approach is the premise that evaluating options in absolute terms is often cognitively demanding, and that comparisons may provide additional information when that is the case. 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. 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 more as an accounting identity than as a real behavioral model. We view our empirical investigation not as a test of a specific model but of the broad idea of learning from comparison points.

2.3 An Illustrative Model

We present a simple linear case for illustration. Appendix A.3 presents a non-linear version. Suppose the optimal action is given by

$$a^{\star}(x, \beta) = \beta x, \qquad \beta > 0.$$

where β does not depend on the comparison point.

The DM wishes to choose an action a that tracks $a^\star(x, \beta)$. However, he does not know the policy function parameter β (e.g., Yang, 2023; Ilut and Valchev, 2023). The DM plays

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

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 β with mean μ_β and variance σ^2_β. Moreover, he has access to comparison data, generated according to

$$a_c \;=\; \beta_c x_c.$$

The parameter β_c captures the mapping between inputs and optimal actions in the comparison context. It need not coincide with the DM's own mapping parameter β. 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, β), they do not affect the optimal action. 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 β_c is informative in the sense of posterior monotonicity:

$$\frac{\partial\, \mathbb{E}[\beta \mid \beta_c]}{\partial\, \beta_c} \;>\; 0.$$

A simple case satisfying this condition is $\beta_c = \beta + \nu$, where ν is independent mean-zero noise with a log-concave density (e.g., Gaussian noise). While there are many reasons to expect β and β_c to be different, we only require that β_c is treated as informative about β. 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 $\beta_c = \frac{a_c}{x_c}$ exactly. He then chooses

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

with $\frac{\partial\, \mathbb{E}[\beta|\beta_c]}{\partial \beta_c} > 0$ 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 $\beta_c = \frac{a_c}{x_c}$ and thus a higher $\mathbb{E}[\beta \mid \beta_c]$, which increases a. On the other hand, a higher comparison input x_c (holding a_c fixed) implies a lower β_c, decreasing the chosen action.[5]

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

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

with β and ϵ independent. Standard Bayesian updating yields

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

The formula illustrates two features. First, the DM's action is a precision-weighted average of the prior mean μ_β and the inferred mapping a_c/x_c from the comparison problem. Second, the expression 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.

We use this Gaussian setup to illustrate how comparator effects depend on the degree of mapping uncertainty and informativeness. The sign predictions of Prediction 1 extend to a general nonlinear model under mild regularity conditions, while the comparative statics below extend to conjugate families with linear posterior means (see Appendix A).

Prediction 2. *In the linear Gaussian benchmark:*

[5] The proportional form $a^\star = \beta x$ passes through the origin, which is what lets the DM recover the mapping from a single comparison observation. With an unknown intercept, $a^\star(x, \beta) = \alpha + \beta x$, a single comparison point (a_c, x_c) no longer identifies the slope, so the inference—and hence the sign of the input-comparator effect—becomes sensitive to beliefs about α and can attenuate or even reverse.

(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_β^2.

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

Only one comparator observed. What happens if the DM observes only one comparator, or if only one is made salient by an experimenter? In such cases, the same objective decision problem can produce very different behavior purely depending on whether the input or output comparator is present or salient. The comparison context (a_c, x_c, β_c) is still generated by $a_c = \beta_c x_c$, but the DM either sees only x_c or a_c. Appendix A.3 shows that, under mild assumptions, the qualitative predictions stated above still hold.[6] However, as we show formally in the same Appendix, the quantitative magnitude of both comparison effects decreases relative to the case in which both comparators are observed. To illustrate, when a Bayesian DM only observes the output comparator a_c, he attributes a high a_c not only to a high β_c but also to a high x_c. Due to this cross-inference, comparison effects are weaker when only one comparator is involved.

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. We normalize $g(\cdot)$ to increase in a. Based on the distinction between actions and outcomes, we also distinguish between two types of output comparators: *action comparators*, a_c, and *outcome comparators*, $y_c = g(a_c, x_c, \eta)$.

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 could be viewed as an extreme form of assimilation that arises when the DM places nearly all weight on comparison-based information. 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 ($\sigma_\beta^2 \to \infty$). Then, if $x \approx x_c$ (or if the DM is inattentive to the precise value of x and thus relies on x_c), a DM who targets the comparison outcome chooses an action that replicates it. Intuitively, when the DM has little knowledge about how to solve the problem, he treats the observed outcome in the comparison problem as

[6]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. 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.

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 outcome-based comparison points (Dean et al., 2026).

Interpretation of transparently uninformative comparison points. The model assumes that the comparison point is objectively informative, or at least perceived as such. In practice, people may also respond to comparison points that are not (perceived as) informative. 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 implication for us (in particular for our classification of the prior literature) is that even in those contexts in which comparison points are uninformative, we expect to see assimilation to output comparators and contrast to input comparators.

3 Experiment Design

Table 1 presents an overview. 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). We also list potential sources of decision uncertainty in these paradigms, such as preference uncertainty, uncertainty about the statistically-optimal updating rule or computational uncertainty.

The *Effort Valuation* paradigm is our workhorse application, in which we study (i) both social and expectations-based comparison points, (ii) situations in which both input and output comparators are shown, and (iii) situations in which only one of the two is displayed. Moreover, we also use this workhorse paradigm (iv) 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*). Finally, we also use this paradigm (v) 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 in other contexts, including in a setting with an objectively correct solution.

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

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
	Random (expectations)	Mechanism: reduce informativeness	300
	Potential sources of uncertainty: Task difficulty / completion time; consumption-leisure prefs; computational; optimization		
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
	Potential sources of uncertainty: Risk prefs; computational; optimization		
Beliefs Decision: posterior belief Inputs: prior, accuracy, result	No CP	Baseline; no comparison points	150
	Both CP (social)	Social input and output comparators; document same effects in obj. task	250
	Potential sources of uncertainty: Optimal updating rule; computational		
Effort Choice Decision: workload Input: piece rate	No CP	Baseline; no comparison points	150
	Both CP (expectations)	Comparing with *Effort Valuation*, flip workload from input to output	250
	Potential sources of uncertainty: As in *Effort Valuation*		

Notes. For the mechanism conditions, the "# subjects" column reports the total pre-registered collection for the corresponding wave: *Deliberation* comprises 300 treated subjects plus a re-collection of 150 *Both CP (social)* subjects; *Random* comprises 150 treated subjects plus a re-collection of 150 *Both CP (expectations)* subjects; *Time* comprises 350 treated subjects, who are compared against the original *Both CP (social)* collection. All other rows report single-arm target quotas.

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). A subject completes 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 I.3–I.5 in Appendix I.1.1 for details on each task.

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 I.1). If the randomly-drawn compensation was at least as high as the participant's stated lowest acceptable compensation, the participant had to complete the full workload of the randomly-selected round at the end of the study and received the drawn compensation as a bonus; otherwise, the participant did not work and received no bonus. We study five between-subject conditions.

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, see Figure 1. These correspond to social input and output comparators, respectively. Appendix Figure I.24 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."

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 I.22, 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

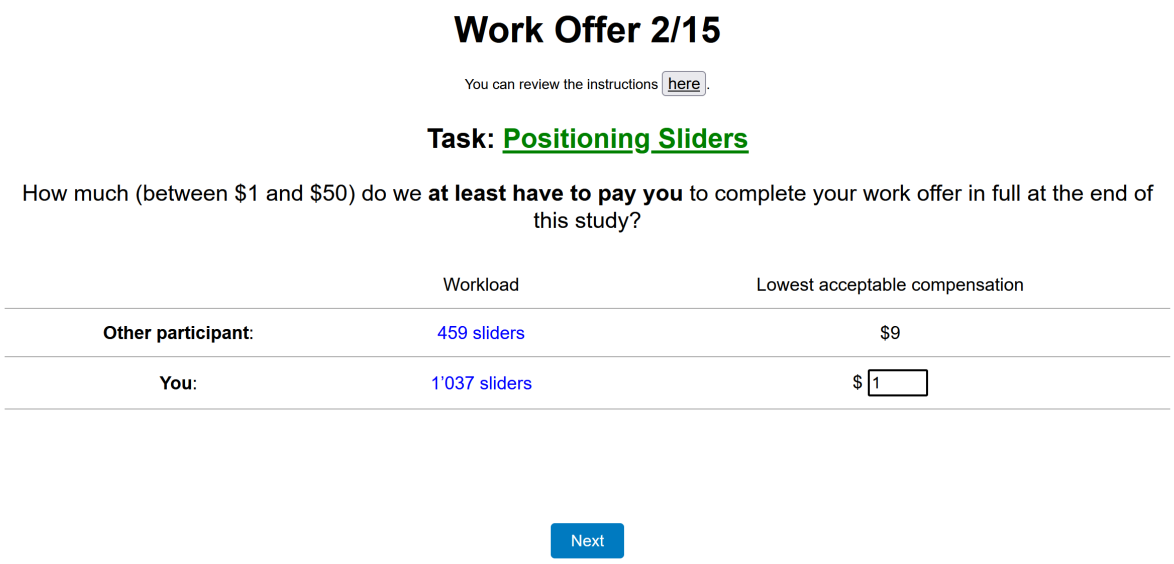

Figure 1: Decision screen in the *Social Both CP* treatment in the *Effort Valuation* paradigm.

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 I.23 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. 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 real numbers. Rather, arguably, the experimenter sets default compensations that are "reasonable."

Input CP only / Output CP only (social). 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 prob-

ability. 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 are unlikely to play a role.[7] 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.

[7] A residual possibility is a preference to conform with others' *reports*. Such conformity could in principle contribute to the positive effect of the comparison posterior, but it cannot generate the *negative* effects of the comparison prior and likelihood ratio—the distinctive prediction of our taxonomy in this paradigm.

3.4 Potential Sources of Uncertainty

As noted in Table 1, different sources of uncertainty about the best ex-ante decision may be at play across our different experimental paradigms. In the effort supply experiments, subjects may have (relatively traditional) uncertainty about how much time a given workload requires to complete (or how annoying it is). In addition, subjects may be uncertain about how to trade off consumption and leisure, or about how to identify a maximum. The investment experiment, on the other hand, is fully specified in the sense that all outcomes are known to subjects (unlike in the effort experiments, where the precise completion time may be unknown). Thus, if decision uncertainty is present, it likely reflects preference uncertainty or maximization uncertainty about how to trade off risk and return. Finally, in the belief updating experiment, the predominant source of uncertainty is plausibly lack of knowledge of the correct statistical rule, or computational uncertainty about how to implement it.

3.5 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 preregistered; Appendix Table B.1 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 *Random* treatments were collected later. However, no commercially available LLM agent could pass our video recording test at the time of data collection. For each treatment arm, we recruited participants until a preregistered quota of complete responses was reached (e.g., 150 in the *No CP* arms and 250 in

most comparison-point arms); responses collected beyond an arm's quota are not analyzed. The bot exclusions remove between one and six participants per arm, yielding the analysis samples reported in the tables (e.g., 145 of 150 participants in each of the three *No CP* arms).

4 Results

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.

Figure 2 summarizes the results on how decisions vary with the input and output comparators. 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 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' WTAs strongly decrease 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 2. 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 preregistered, we include fixed effects for 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.

Columns (4) and (5) show that these patterns are qualitatively identical when only either

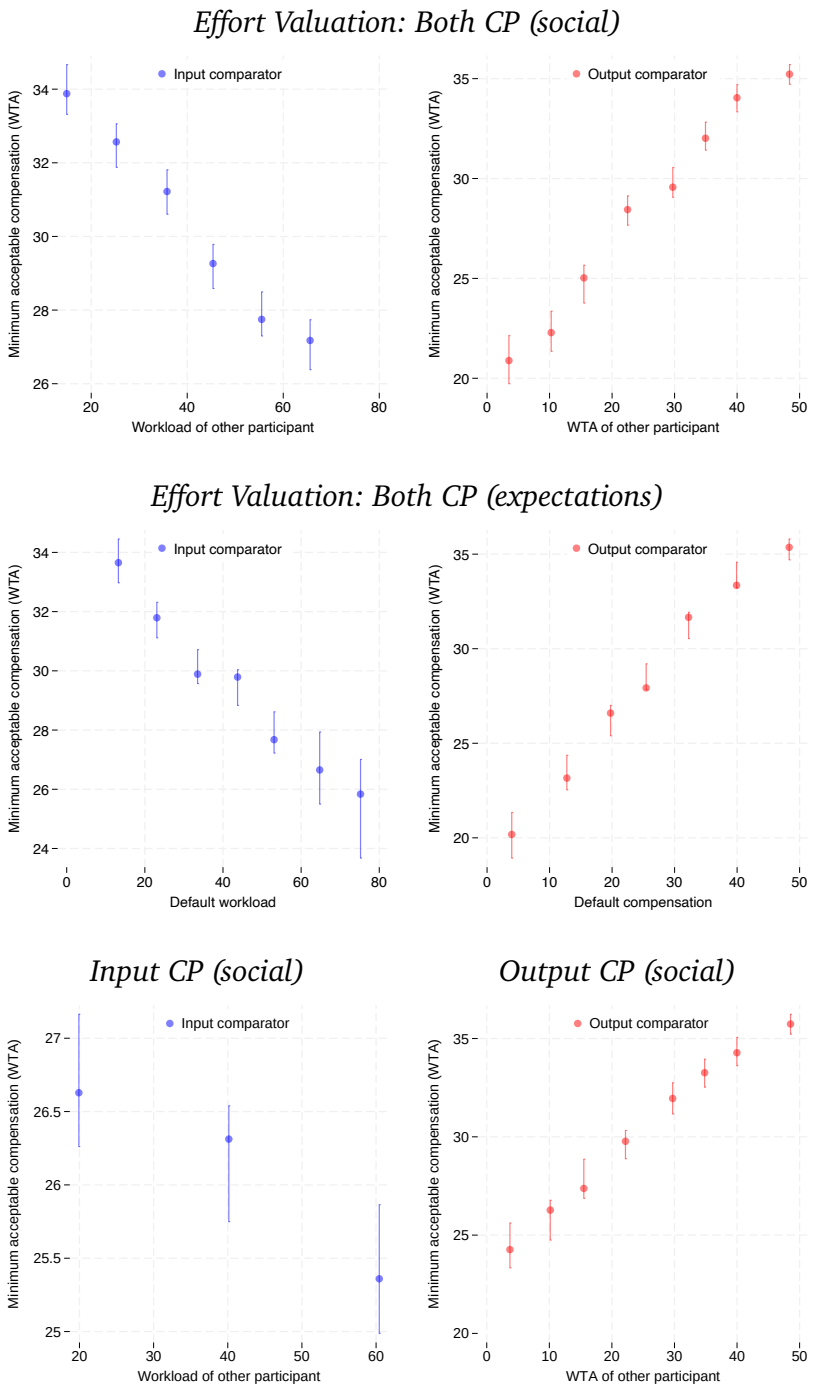

Figure 2: Results of treatments in *Effort Valuation* paradigm. All left-hand side panels show average WTA as a function of the input comparator, which is given by other subjects' workload in the social treatments and by the default workload in the expectations treatment. All right-hand side panels show 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)–(5) of Table 2. 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.

the input comparator or the output comparator are shown to subjects. At the same time, the quantitative magnitude of both coefficients decreases (compare with column (2)). As noted in Section 2, this is consistent with an informational mechanism. For instance, when only the output comparator is shown, then a high comparison point could in principle be due to either a higher mapping parameter (β in the model) or a high input comparator, and a Bayesian would partially attribute to both. This partial attribution means that any given variation in the output comparator affects decisions less than in the case in which both comparators are presented. A similar logic holds when only the input comparator is shown.

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 imposed asset's winning chance. The right-hand panel shows average investments as a function of the output comparator, i.e., the imposed investment amount. 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 2 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.[8]

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 2 report the associated Grether regressions.

This final application to beliefs suggests that the empirical regularity of a negative effect of

[8]Guesses of 0% or 100%—whether the subject's own or, in *Both CP (social)*, the comparison participant's—have undefined log odds and mechanically drop out of the estimation sample: 40 of 2,220 rounds in *No CP* and 80 of 3,690 rounds in *Both CP (social)*, leaving the 2,180 and 3,610 observations reported in Table 2.

Table 2: 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)*	*Input CP only (social)*	*Output 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.

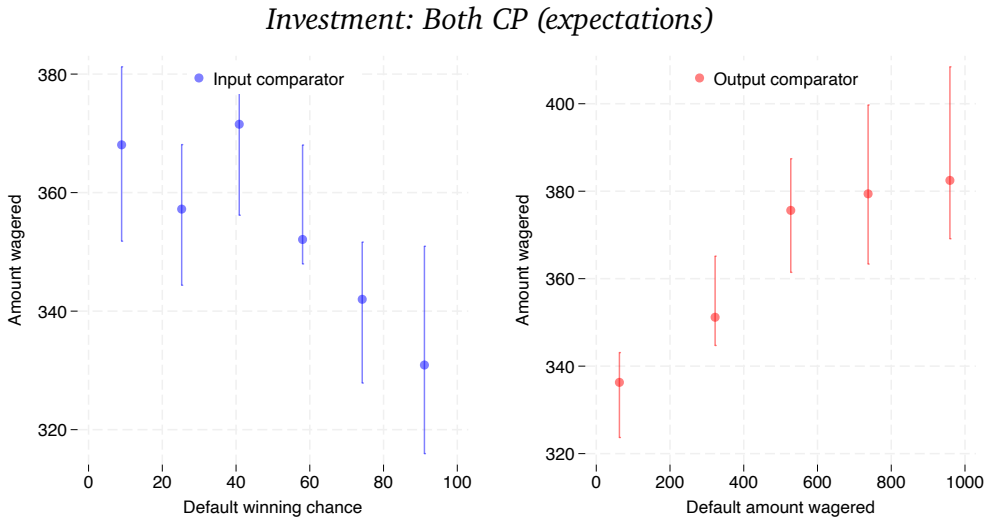

Figure 3: Results of *Investment* experiment. The left panel shows average investments as a function of the comparison winning chance. The right panel shows average investments as a function of the comparison amount wagered. All comparators are based on expectations. Both panels are constructed from multiple regressions as in column (7) of Table 2. Points show binned mean estimates, with 95% confidence intervals based on robust standard errors.

input comparators and of a positive effect of output comparators similarly emerges in objective problems in which, for example, reference-dependent preferences should not play a role.

4.4 Robustness and Secondary Analyses

Appendix F reports pre-registered robustness checks and secondary analyses across all paradigms and treatments (Tables F.1–F.14). These include alternative sample restrictions (strict human verification, response-time validity, protocol compliance), median regressions, round interactions, and sample splits by whether the *input* comparison is favorable or unfavorable. The results are broadly stable: across all specifications, the pattern of input contrast and output assimilation effects is preserved with generally similar magnitudes. A few estimates in smaller subsamples lose statistical significance (e.g. some input-comparator effects in the median regressions and in the favorable/unfavorable splits); see Appendix F for details.

5 Manipulating the Type of Comparison Point

A key implication of our account 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 required workload, which subjects map into a WTA. We complement this paradigm with the *Effort Choice* paradigm, in which

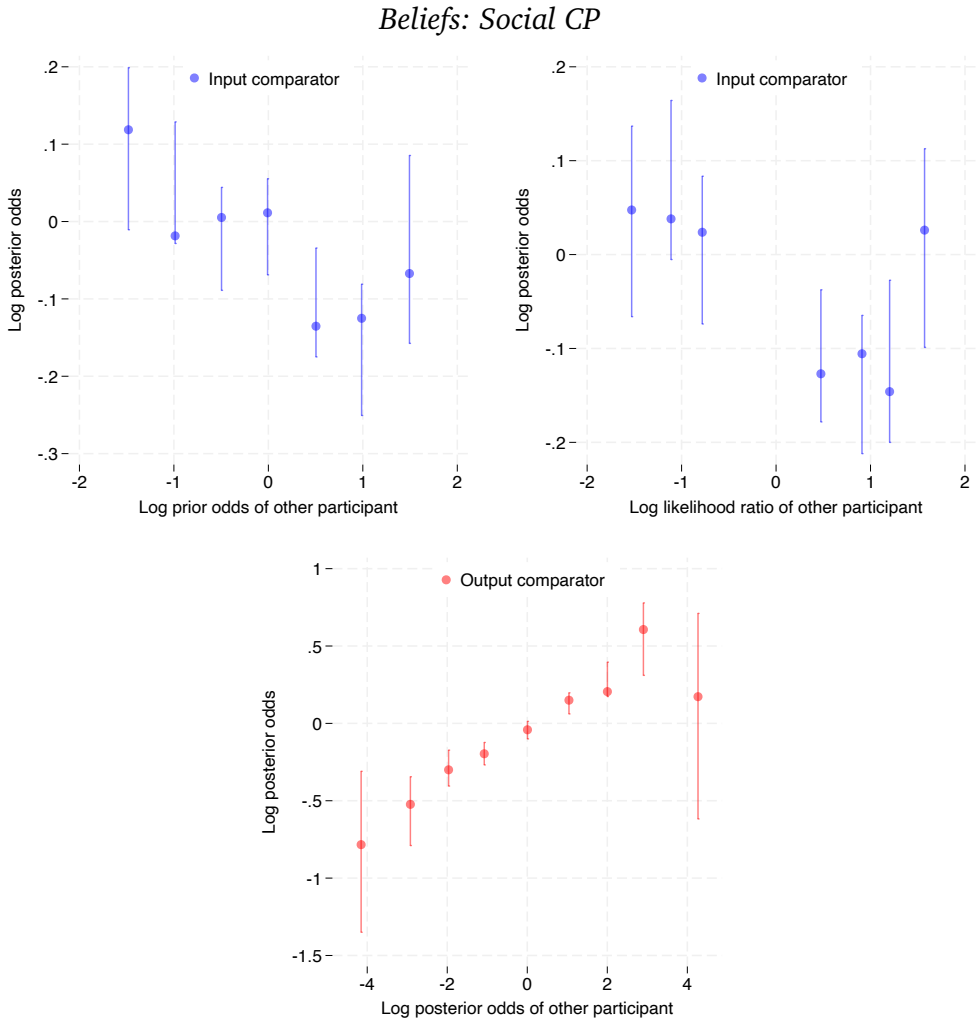

Figure 4: Results of *Beliefs* experiment. The top panels show average log posterior odds as a function of the comparison log prior odds and the comparison log likelihood ratio. The bottom panel shows the results for the comparison log posterior odds. All comparators are from a social comparison. The panels are constructed from multiple regressions as in column (9) of Table 2. Points show binned mean estimates, with 95% confidence intervals based on robust standard errors.

subjects 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 is either an input comparator or an output comparator. In both paradigms, we use the expectations-based variant of comparison points.

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 I for screenshots of instructions, comprehension quizzes and decision screens.

Results. Figure 5 compares how average decisions in *Effort Valuation* and *Effort Choice* vary as a function of the comparison workload. The left y-axis indicates decisions in *Effort Valuation* (where subjects stated a WTA) and the right axis 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 B.3 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.[9]

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.

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)).

[9]The two paradigms of course differ in more than the role of the comparison workload—the decision object, incentives, and response scales all change. The comparison therefore should not be read as a ceteris-paribus manipulation of a single factor.

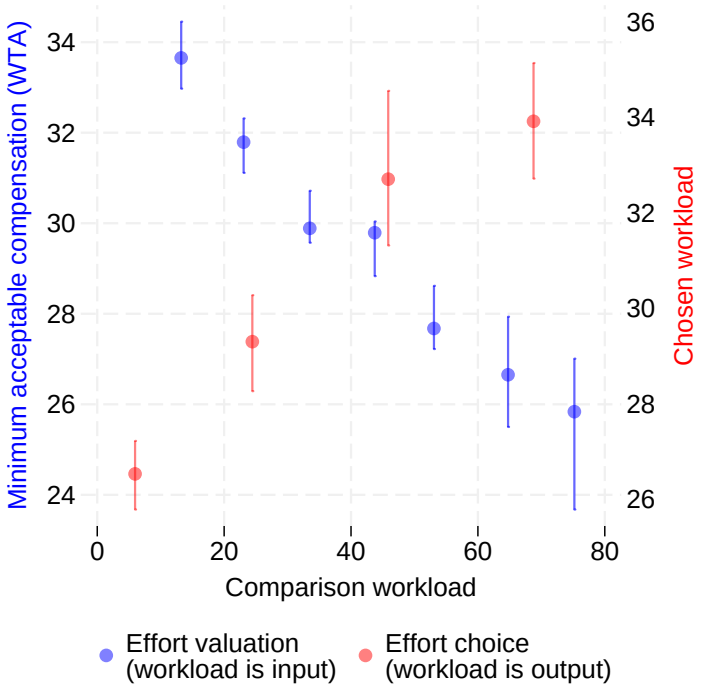

Figure 5: Results in treatments *Effort Valuation: Both CP (expectations)* and *Effort Choice: Both CP (expectations)*. The x-axis shows the comparison workload supplied to subjects and the y-axis average decisions. In *Effort Valuation*, shown on the left axis, subjects decide their minimum required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

6.1 Reducing Input-Output Mapping Uncertainty

We return to our main application, *Effort 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.

Design. The *Deliberation* condition is a variant of the *Social Both CP* condition. Before participants make their incentivized decisions, they encounter a 'training round': a screen that asks them to state—in purely hypothetical terms—their WTA for each of up to 13 different possible workloads in the form of a list (see Appendix Figure I.8). Participants only filled out this list for one randomly selected real effort task. Subsequently, they made their main incentivized decisions also for this type of 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. We do not analyze the data from the 'training round.' Rather, our hypothesis

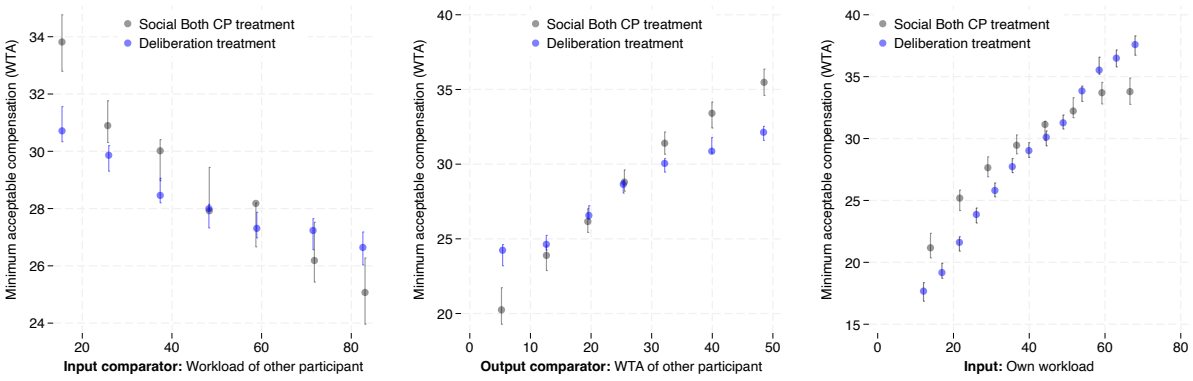

Figure 6: Effect of deliberation on sensitivity to own input and comparators. Results of conditions *Deliberation* and *Social Both CP* in the effort valuation paradigm. The y-axis shows minimum required compensations. From left to right, the x-axis shows the input comparator, the output comparator,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

pertains to the main incentivized decisions subjects take (in the same format as in the treatments discussed earlier).[10]

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 within-session treatment comparison. 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 3, 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 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.[11] The results of this treatment,

[10]By design, the *Deliberation* condition bundles the hypothetical policy-function elicitation with the fact that all main incentivized decisions then concern the single task type covered in the training round, whereas decisions in *Social Both CP* span three task types. Both components plausibly reduce mapping uncertainty—through explicit deliberation and through task-specific familiarity—so we interpret this treatment as manipulating mapping uncertainty broadly rather than as isolating the deliberation channel specifically.

[11]Like *Deliberation*, the *Time* treatment bundles two components: in addition to the duration information displayed on the decision screens, participants completed—before their incentivized decisions—a purely hypotheti-

summarized in column (2) of Table 3, mirror those of the *Deliberation* treatment: expressing workloads in a second, more familiar currency substantially reduces the effects of the input and output comparators (by 44% and 42%, respectively). 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 and thus engage in comparative thinking even when it is not justified. Still, on the margin, the combination of these two considerations predicts that comparison points should matter less when they are transparently uninformative for the decision at hand. This is not the case in our treatments involving expectations-based comparison points reported above. 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 outcomes with 50% chance) are determined. In practice, these comparison points are drawn from other subjects' decisions.

In the *Random* 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."[12]

Results. We pre-registered a collection of 150 subjects for the *Random* treatment alongside a recollection of 150 subjects for the *Expectations Both CP* condition. Column (3) of Table 3 and Appendix Figure B.1 summarize the results. We find that the informativeness manipulation

cal elicitation of their WTA for working 10, 20, . . . , 70 minutes (Appendix Figure I.6). Both components plausibly reduce mapping uncertainty by familiarizing subjects with valuing work durations, so we again interpret this treatment as manipulating mapping uncertainty broadly rather than as isolating the effect of the duration display alone. See Appendix D for details.

[12]The manipulation changes only the described meaning of the comparison points, not their statistical properties: the marginal distributions of the displayed workloads and compensations are statistically indistinguishable across the *Random* and *Expectations Both CP* arms of this collection (Kolmogorov–Smirnov $p = 0.95$ and $p = 0.12$, respectively), and the displayed workload–compensation pairs are essentially uncorrelated within both arms ($|r| < 0.02$).

Table 3: 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. Linear-combination tests of the implied total effects in the *Random* treatment (column (3), baseline plus interaction) yield $p = 0.011$ for the input comparator and $p < 0.001$ for the output comparator.

reduced sensitivity to input comparators by 69% and to output comparators by 33%. At the same time, comparison effects do not disappear in the *Random* treatment: the implied input-comparator effect, though attenuated, remains negative and significant ($-0.127 + 0.088 = -0.039$, $p = 0.011$), and the implied output-comparator effect remains sizable ($0.325 - 0.106 = 0.219$, $p < 0.001$). There are at least two potential explanations for this.

First, 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 make use of them. In this case, what matters 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.

Second, uninformative comparison points may matter for unrelated reasons such as experimenter demand effects. In keeping with the experimental economics tradition, we thus view the comparative statics result (that the reliance on comparison points increases in informativeness) as more informative than the level result (that even uninformative comparison points matter).

7 Classifying Prior Studies

To investigate the broader explanatory power of the input-output taxonomy, we revisit prior field observational and experimental studies. This classification exercise is inherently imperfect and partly subjective. We discuss relevant caveats in Section 7.2 and probe the robustness of our manual coding with an independent large-language-model classifier in Section 7.4.

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.[13] 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', 'anchoring', 'goal', 'target', 'hot hand' and 'gambler's fallacy' in the paper's title and abstract.[14] This search returned 613 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.[15]

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.

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 manipulation varies the choice options themselves—i.e. what we would characterize as decision inputs—rather than a comparison point; and *(iv)* the variation is not in the magnitude of the comparison point but

[13] 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*.

[14] 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.

[15] 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). The same logic extends to fixed non-zero thresholds when agents' positions relative to the threshold vary, as in bunching designs (e.g., par in golf).

in its type or existence. The latter includes designs that vary how frequently comparisons are observed (as in myopic loss aversion studies) and designs in which a comparison point is present in some conditions but entirely absent in others (e.g., a candidate's disclosed prior salary vs. no disclosed salary), since neither varies the magnitude of an existing comparison point.

This leaves us with a final sample of 109 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 identifying 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 typically 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, and in which direction it moves the decision (according to the analysis of the authors).

In classifying whether a variable is an input or output, the key criterion 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). While these definitions may appear relatively loose, we note that the vast majority of economics models take a stand on what an input or an output is (quantities that are chosen by the decision maker vs. taken as given).

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. Several papers in our sample report null results for secondary treatments or outcomes alongside a significant primary comparator effect; in these cases, we classify the primary effect and note the associated null results in Appendix Table G.1. When, instead, a paper's primary comparator test itself yields no statistically significant effect, we record the paper as "Null effect" (two papers only). When a paper reports a comparator effect that is non-monotonic in the comparator's level (e.g., goal effects that van-

ish for unattainable goals), we classify the paper's main directional effect; only when no main directional effect can be identified (e.g., a symmetric U-shape) would a paper be unclassifiable on this ground.

7.2 Caveats and Limitations

Our classification is inherently subjective. Papers are rarely designed with our taxonomy in mind; the boundary between a comparator and a decision input is sometimes ambiguous; and coding a study sometimes requires a judgment call about what the decision maker treats as given or chooses. We probe this concern directly in Section 7.4 by re-running the classification with an independent large-language-model coder.

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, although we retain and classify papers whose primary comparator tests yield null results, null findings are underrepresented in published work, so the share of significant findings consistent with our taxonomy should be interpreted with publication bias in mind.

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. 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.

Finally, the literature classification only tests the input-vs.-output-comparator taxonomy as such, rather than the idea of comparison points as sources of information.

Given these caveats, we view the results reported below not as a precise test statistic but, rather, as a broad ballpark estimate.

7.3 Results

Appendix Table G.1 reports the full annotation. According to our classification, 109 papers yield 111 classifiable comparator effects. Of these, 27 involve an input comparator: 26 produce a contrast effect and one yields a null result. 84 comparison effects involve an output comparator: 78 give rise to an assimilation effect, five to a contrast effect, and one yields a null result.

Table 4 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.

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 admissions decisions increase 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.

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. Consistent with our taxonomy, the endowment effect is an assimilation effect in output (out-

come) space.[16]

Five papers in our sample yield findings inconsistent with our predictions, all of which are related to the gambler's fallacy. The first is 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).

Three further inconsistent findings come from studies of lottery play that document that players tend to avoid recently-drawn numbers, a contrast effect in outcome space (Clotfelter and Cook, 1993; Suetens et al., 2016; Kong et al., 2020). The fifth is Loh and Warachka (2012), who document that post-earnings-announcement drift is strong when a surprise extends a streak of same-signed surprises and negligible when a streak ends—again a contrast effect in outcome space. Intuitively, it is clear that our input-vs.-output taxonomy predicts hot-hand rather than gambler's fallacy effects (decisions or outcomes increase in previous decisions or outcomes). Thus these gambler's fallacy papers go against our taxonomy. That said, one may be skeptical that hot-hand and gambler's-fallacy effects are comparison effects at all: people in these contexts may instead misperceive the data-generating process, such that previous realizations act as objective inputs into the decision ('How biased is this coin?') rather than as comparators. While we classify these cases as going against our taxonomy, they illustrate (i) the inherent ambiguity involved in subjective classification exercises and (ii) that our information-based approach is not the only mechanism at play in the literature.

7.4 Classification by a Large Language Model

A natural concern with the preceding exercise is that it is subjective and was carried out by us, the authors. To probe its robustness, we re-ran the entire classification using a large language model as an additional classifier. In designing the model's instructions, our goal was to make them as non-leading as possible. For instance, "not classifiable" (N/A), "no effect" (Null), "consistent" and "inconsistent" are presented as equally legitimate outcomes. The full instructions are reproduced in Appendix H. While we view the additional classification by the LLM as valuable robustness check, it is worth keeping in mind that it does not constitute an entirely independent second classification because the LLM operates based on our own classification instructions, including the examples we supply. In addition, the instructions must state our classification rules in a form that the model can verify mechanically, which leads to two differences relative to our own coding. First, to decide whether a comparator relates to a decision input

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

Table 4: 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. sell 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)

or a decision output, the instructions primarily ask whether the comparator is measured in the same units as the subject's decision (together with a list of exceptions), whereas our manual coding applies the input-vs.-output criterion of Section 7 directly. Second, the instructions classify a comparator relative to the decision as the subject reports it, whereas our manual coding classifies it relative to the underlying economic decision. For example, when subjects accept or reject an offer, the instructions treat a monetary benchmark as background information and hence as an input comparator; our own coding instead asks what role the benchmark plays in the underlying valuation problem, which can make it an output comparator. Some disagreements documented below therefore reflect these differences in operational rules rather than differences in judgment.

We use the Fable 5 model by Anthropic. For each paper, the model looks up the study, decides whether it contains a comparison point whose magnitude varies—and is therefore classifiable at all—and, if so, identifies the comparator, classifies it as relating to a decision input or output, and records the direction of its estimated effect. According to the model's classification, 134 of the screened papers contain a classifiable directional comparator effect. Of these, 33 involve an input comparator, 30 of which produce a contrast effect. The remaining 101 involve an output comparator, 92 of which give rise to an assimilation effect. In the intersection of studies that we also code as classifiable (Table G.1), the model codes 83 directional comparator effects: 18 input comparators, of which 17 produce contrast, and 65 output comparators, of which 59 produce assimilation.

Comparing the model's classification to ours over the full set of 613 papers that entered our screening—including papers we ultimately excluded—the two assign the same four-way label (consistent / inconsistent / null / not classifiable) in 86% of cases (525 of 613). Much of this agreement reflects papers that both we and the model deem out of scope. Within the 110 entries in Appendix Table G.1, the model assigns the same label in 80 cases (73%), with the modal disagreement being that the model codes a paper as not classifiable even when we do (21 cases). Conditional on both us and the model viewing a paper as classifiable, the two reach the same conclusion in 90% of cases (80 of 89). Restricting further to papers where both code a directional effect, we agree on whether it is consistent or inconsistent with the taxonomy in 95% of cases (79 of 83). If anything, the model applies a stricter bar within the papers it classifies: it codes more effects than we do as inconsistent with the taxonomy (12 versus 5) and codes as null several effects that we, following the papers' own framing, record as directional.

In summary, the LLM classifier recovers the same central regularity—input comparators producing contrast and output comparators producing assimilation. This tentatively suggests that our headline result is not merely an artifact of subjective coding.

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 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 I.43.

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 I.3.

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 our concrete experiment, 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 what would be expected from random guessing over the three substantive response options (33%; 25% when also including the "I don't know" option).[17] Appendix Table E.1 reports the full breakdown by paradigm and comparator type. The pattern holds across all four paradigms.[18]

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.

9 Discussion

Reference-dependent preferences. Our account does not stand in contradiction to the existence of reference-dependent preferences.[19] To take a trivial example, adaptation effects such as that lukewarm water feels more pleasant after cold than after hot water represent clear cases in which sensations and utility depend on reference points. At the same time, our classification of prior studies suggests that some of the evidence that is frequently interpreted through the

[17]Because the response options bundle sign and statistical significance, "correct" here means selecting the sign-and-significance category matching our estimate (all featured comparator effects are statistically significant). A forecaster who anticipated the correct sign but expected an insignificant effect would be coded as incorrect, so these rates are conservative measures of pure sign knowledge. Note, however, that misprediction of input comparators is predominantly sign error: the modal response for four of the five input comparators is *significantly positive* (Appendix Table E.1).

[18]One caveat: as documented in the note to Appendix Figure I.48, the *Investment* prompt as fielded contained two typos, including describing the imposed amount as a fixed "$100" although it varied across rounds. Because forecasters were asked to predict the coefficient on the imposed amount, the *Investment* output-comparator forecasts should be interpreted with this caveat in mind.

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

lens of models of reference-dependent preferences may be driven by various forms of decision uncertainty.

In this section, we discuss which of our experimental results are potentially compatible with canonical models of reference-dependent preferences, and which ones are more naturally explained by 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).[20] 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. People may experience higher effort costs when the piece rate at which they work is below a comparison piece rate, or the utility they derive from an investment return may be lower if the comparison return is higher. We have not attempted to write down and solve such unusual 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. We reiterate that inputs are variables that are taken as given by the decision maker, while outputs are variables that are implicitly or explicitly controlled by the decision maker in the context that 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 typically declares inputs (parameters) and outputs (decision variables and associated outcomes).

[20]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 suggests, reference earnings and reference wages have psychologically different effects.

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. The common thread is that absolute evaluations are hard, so that people lean on comparison points to simplify the problem and (at least implicitly) acquire information about how to navigate the tradeoffs involved. Crucially, the informational content of input and output comparators differs in a way that determines the sign of comparison effects.

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). Teasing apart which reference effects reflect preferences and which ones informational mechanisms is a promising direction for future work.

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.

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. 2012. "Salience theory of choice under risk." *Quarterly Journal of Economics* 127 (3): 1243–1285.

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.

Clotfelter, Charles T., and Philip J. Cook. 1993. "The Gambler's Fallacy in Lottery Play." *Management Science* 39 (12): 1521–1525.

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.

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.

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.

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.

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.

Ilut, Cosmin, and Rosen Valchev. 2023. "Economic agents as imperfect problem solvers." *The Quarterly Journal of Economics* 138 (1): 313–362.

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.

Izakson, Liz, Minhee Yoo, Adam Hakim, Ian Krajbich, Ryan Webb, and Dino J Levy. 2024. "Valuations of target items are drawn towards unavailable decoy items due to prior expectations." *PNAS nexus* 3 (7): pgae232.

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.

Kong, Qingxia, Georg D. Granic, Nicolas S. Lambert, and Chung Piaw Teo. 2020. "Judgment Error in Lottery Play: When the Hot Hand Meets the Gambler's Fallacy." *Management Science* 66 (2): 844–862.

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

Kőszegi, Botond, and Adam Szeidl. 2013. "A model of focusing in economic choice." *Quarterly Journal of Economics* 128 (1): 53–104.

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.

Loh, Roger K., and Mitch Warachka. 2012. "Streaks in Earnings Surprises and the Cross-Section of Stock Returns." *Management Science* 58 (7): 1305–1321.

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.

Suetens, Sigrid, Claus B. Galbo-Jørgensen, and Jean-Robert Tyran. 2016. "Predicting Lotto Numbers: A Natural Experiment on the Gambler's Fallacy and the Hot-Hand Fallacy." *Journal of the European Economic Association* 14 (3): 584–607.

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.

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

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.

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, $a^\star(x,\beta) = \beta x$, so observing (a_c, x_c) reveals $\beta_c = a_c/x_c$ exactly. The DM chooses

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

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

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

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

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

since $a_c, x_c > 0$ and $m'(\beta_c) > 0$. □

A.2 Proof of Prediction 2

Suppose $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ and $\beta_c = \beta + \epsilon$ with $\epsilon \sim N(0, \sigma_\epsilon^2)$ independent of β. Since the maintained assumptions require $\beta, \beta_c > 0$ (and hence $a_c > 0$), we take $(\mu_\beta, \sigma_\beta, \sigma_\epsilon)$ such that negative realizations occur with negligible probability, so that the unrestricted Gaussian model serves as an accurate approximation to a positive-support specification; the comparative statics below are evaluated at positive comparator realizations. Standard Bayesian updating yields:

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

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

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_β^2:

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

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

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

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

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

Consider the linear-Gaussian benchmark of Prediction 2. The comparison context (x_c, a_c, β_c) is still generated by $a_c = \beta_c x_c$, with (x_c, β_c) drawn from some joint distribution, but the DM now observes only one of the two comparators. Let $\varepsilon_{z,w} := \frac{\partial \log z}{\partial \log w}$ denote the elasticity of z with respect to w, and define

$$\varepsilon_x := \varepsilon_{\mathbb{E}[a_c|x_c],\, x_c} \qquad \text{and} \qquad \varepsilon_a := \varepsilon_{\mathbb{E}[1/x_c|a_c],\, a_c}.$$

Because the DM can no longer recover β_c exactly, updating proceeds via the law of iterated expectations. We assume that, conditional on β_c, the comparators contain no additional information about β (formally, β and (x_c, a_c) are independent given β_c), so that $\mathbb{E}[\beta \mid \beta_c, s] = \mathbb{E}[\beta \mid \beta_c]$. Then, for the observed comparator $s \in \{x_c, a_c\}$,

$$\mathbb{E}[\beta \mid s] \;=\; \mathbb{E}\big[\mathbb{E}[\beta \mid \beta_c] \mid s\big] \;=\; \mu_\beta + \lambda\big(\mathbb{E}[\beta_c \mid s] - \mu_\beta\big),$$

where the second equality uses the linear posterior mean from the proof of Prediction 2. Note that the weight λ is the same as in the two-comparator case: it governs the informativeness of β_c about β, not how much of the comparison context is directly observed. The decision rule is $a(x; s) = \mathbb{E}[\beta \mid s]\, x$.

Remark (Signs when only one comparator is observed)**.** *(a) (*Input comparator only*) If the DM observes only x_c, then input comparators create contrast, $\frac{\partial a}{\partial x_c} < 0$, if and only if $\varepsilon_x < 1$.*

(b) (Output comparator only) *If the DM observes only a_c, then output comparators create assimilation, $\frac{\partial a}{\partial a_c} > 0$, if and only if $\varepsilon_a > -1$.*

Proof. (a) Since $\beta_c = a_c/x_c$ and x_c is observed, $\mathbb{E}[\beta_c \mid x_c] = \frac{1}{x_c}\mathbb{E}[a_c \mid x_c]$. Differentiating the decision rule,

$$\frac{\partial a}{\partial x_c} = \lambda x \frac{\partial}{\partial x_c}\left(\frac{\mathbb{E}[a_c \mid x_c]}{x_c}\right) = \lambda x \frac{x_c \frac{\partial \mathbb{E}[a_c|x_c]}{\partial x_c} - \mathbb{E}[a_c \mid x_c]}{x_c^2} = \lambda x \frac{\mathbb{E}[a_c \mid x_c]}{x_c^2}\big(\varepsilon_x - 1\big),$$

which is negative if and only if $\varepsilon_x < 1$.

(b) Since $\beta_c = a_c/x_c$ and a_c is observed, $\mathbb{E}[\beta_c \mid a_c] = a_c\,\mathbb{E}[1/x_c \mid a_c]$. Differentiating,

$$\frac{\partial a}{\partial a_c} = \lambda x \frac{\partial}{\partial a_c}\Big(a_c\,\mathbb{E}[1/x_c \mid a_c]\Big) = \lambda x\,\mathbb{E}[1/x_c \mid a_c]\big(1 + \varepsilon_a\big),$$

which is positive if and only if $\varepsilon_a > -1$. □

Part (a) requires that the expected comparator output does not increase more than proportionally in x_c; part (b) requires that the expected inverse input does not decline more than proportionally in a_c. Both conditions hold in the natural case in which comparator inputs and outputs are positively but less than proportionally related, i.e., $\varepsilon_x \in (0,1)$ and $\varepsilon_a \in (-1,0)$. Under exactly this condition, both comparison effects are also quantitatively attenuated relative to the case in which the DM observes both comparators:

Remark (Attenuation when only one comparator is observed)**.** *Suppose $\varepsilon_x \in (0,1)$ and $\varepsilon_a \in (-1,0)$. Then the magnitude of each one-comparator effect is strictly smaller than the average magnitude of the corresponding two-comparator effect:*

$$\left|\frac{\partial a}{\partial x_c}\right|_{only\ x_c} = (1-\varepsilon_x)\,\mathbb{E}\left[\left|\frac{\partial a}{\partial x_c}\right|_{both} \,\middle|\, x_c\right] \qquad and \qquad \left|\frac{\partial a}{\partial a_c}\right|_{only\ a_c} = (1+\varepsilon_a)\,\mathbb{E}\left[\left|\frac{\partial a}{\partial a_c}\right|_{both} \,\middle|\, a_c\right],$$

with attenuation factors $(1-\varepsilon_x) \in (0,1)$ and $(1+\varepsilon_a) \in (0,1)$.

Proof. From the proof of Prediction 2, the two-comparator effects are $\frac{\partial a}{\partial x_c}\big|_{\text{both}} = -\lambda \frac{a_c}{x_c^2}\,x$ and $\frac{\partial a}{\partial a_c}\big|_{\text{both}} = \lambda \frac{x}{x_c}$. The first is linear in a_c and the second is linear in $1/x_c$, so averaging each over the comparator not observed by the DM yields

$$\mathbb{E}\left[\left|\frac{\partial a}{\partial x_c}\right|_{\text{both}} \,\middle|\, x_c\right] = \lambda x \frac{\mathbb{E}[a_c \mid x_c]}{x_c^2} \qquad \text{and} \qquad \mathbb{E}\left[\left|\frac{\partial a}{\partial a_c}\right|_{\text{both}} \,\middle|\, a_c\right] = \lambda x\,\mathbb{E}[1/x_c \mid a_c].$$

By the proof of the preceding remark, the one-comparator effects are

$$\left|\frac{\partial a}{\partial x_c}\right|_{\text{only } x_c} = \lambda x \, \frac{\mathbb{E}[a_c \mid x_c]}{x_c^2}(1-\varepsilon_x) \qquad \text{and} \qquad \left|\frac{\partial a}{\partial a_c}\right|_{\text{only } a_c} = \lambda x \, \mathbb{E}[1/x_c \mid a_c](1+\varepsilon_a),$$

which delivers the stated identities. Since $\varepsilon_x \in (0,1)$ and $\varepsilon_a \in (-1,0)$, both attenuation factors lie strictly in $(0,1)$. □

The intuition is cross-inference. When the DM sees only x_c, a higher input also signals a higher unobserved output a_c ($\varepsilon_x > 0$), which partially offsets the contrast channel. Conversely, when the DM sees only a_c, he attributes a higher output partly to a higher unobserved input x_c rather than to a higher β_c ($\varepsilon_a < 0$), which mutes the update about β and hence the assimilation effect.

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).[21]

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 draws on two additional ingredients: outcome targeting, which by itself generates bunching in outcome space, and inattention to input differences, which is additionally required for bunching in action space.

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:

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

[21] 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.

generating bunching at the action comparator a_c.

For intermediate values, the action is

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

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

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

This represents *partial bunching*: the action is a weighted average of the prior-based action $\mu_\beta x_c$ 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

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

The realized outcome is then $g\big(g^{-1}(y_c, x_c), x\big)$, which coincides with y_c when $x = x_c$; bunching at y_c therefore obtains approximately when true inputs are close to the comparison input x_c.

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

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

where the payoff $u(\cdot\,; y_c)$ is single-peaked at the target y_c; for example, under quadratic loss around the target,

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

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*: relative to the dispersed outcomes that would arise in the no-targeting benchmark in which a does not condition on y_c, realized outcomes concentrate around y_c, with residual dispersion 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 $a^\star(x, \beta) = f(x, \beta)$, where $f : \mathbb{R}_{++} \times B \to \mathbb{R}_{++}$ is twice continuously differentiable with $f_x > 0$ and $f_\beta > 0$. The DM is uncertain about $\beta \in B$, holds a prior over it, and chooses

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

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

Since $f_\beta(x_c, \beta_c) > 0$, for each fixed x_c the map $\beta_c \mapsto f(x_c, \beta_c)$ 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

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

The regularity conditions referenced in the main text are:

1. **Smoothness:** $f(x, \beta)$ is twice continuously differentiable with $f_x > 0$ and $f_\beta > 0$.
2. **Invertibility:** For each x_c, the map $\beta_c \mapsto f(x_c, \beta_c)$ is invertible (guaranteed by $f_\beta > 0$ and appropriate domain restrictions).
3. **Informativeness:** β_c is informative about β in the sense that $E[\beta \mid \beta_c]$ is strictly increasing

in β_c. This is satisfied, for example, when $\beta_c = \beta + \epsilon$ with ϵ independent noise with a log-concave density, or more generally when (β, β_c) are affiliated.

4. **Differentiability of posterior:** The posterior expectation $m(\beta_c) = E[\beta \mid \beta_c]$ is differentiable with $m'(\beta_c) > 0$. (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, β_c), stated there.

A.5.1 Proof of Prediction 1

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

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

Assimilation in output comparator. By the chain rule:

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

Since $f_\beta > 0$ (by smoothness), $m'(\beta_c) > 0$ (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:

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

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

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

A.5.2 Prediction 2 in the General Model

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

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

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

Holding the evaluation point of f_β fixed, both effects are amplified whenever $m'(\beta_c)$ increases; in the nonlinear model, a change in the prior or signal structure also moves the posterior mean $m(\beta_c)$ and hence the factor $f_\beta\big(x, m(\beta_c)\big)$, so the comparative statics below isolate the weighting channel. The two parts of Prediction 2 correspond to conditions under which $m'(\beta_c)$ increases.

(Part a) Mapping uncertainty. Compare two priors F_β and $\tilde{F}_\beta$ from the same conjugate family with the same mean but $\tilde{\sigma}_\beta^2 > \sigma_\beta^2$. In conjugate families with linear posterior means, a higher prior variance increases the weight on the signal β_c in the posterior mean, raising $m'(\beta_c)$ and thus amplifying the weighting channel of both comparison effects; outside this class, a higher prior variance need not raise $m'(\beta_c)$ pointwise. In the Gaussian case this is immediate: $m'(\beta_c) = \lambda = \sigma_\beta^2/(\sigma_\beta^2 + \sigma_\epsilon^2)$, which is strictly increasing in σ_β^2.

(Part b) Informativeness. Compare two signal structures H and $\tilde{H}$ within a conjugate family with linear posterior means, where $\tilde{H}$ has lower noise variance. Higher precision then increases the weight on β_c in the posterior mean, raising $m'(\beta_c)$ and thus amplifying the weighting channel of both comparison effects; Blackwell informativeness alone does not guarantee a pointwise increase in $m'(\beta_c)$. 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 β_c exactly. Maintaining the assumption that β and (x_c, a_c) are independent conditional on β_c (as in Appendix A.3), the law of iterated expectations gives:

$$E[\beta \mid x_c] = E\big[E[\beta \mid \beta_c] \mid x_c\big] = E\big[m(\beta_c) \mid x_c\big].$$

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

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

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

Sufficient condition: Suppose the joint distribution of (x_c, β_c) is such that $\beta_c \mid x_c$ 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 $E[\beta_c \mid x_c]$ is decreasing in x_c, and by informativeness, $E[\beta \mid x_c]$ is also decreasing in x_c.

In the linear case, this reduces to the elasticity condition $\varepsilon_x < 1$ from Appendix A.3.

(Part b) Only output comparator a_c observed. By the same argument:

$$E[\beta \mid a_c] = E\big[m(\beta_c) \mid a_c\big].$$

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

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

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

Sufficient condition: Suppose β_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 $E[\beta_c \mid a_c]$ is increasing in a_c, and by informativeness, $E[\beta \mid a_c]$ is also increasing in a_c.

In the linear case, this reduces to the elasticity condition $\varepsilon_a > -1$ from Appendix A.3. □

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, interpreting the linearized equation as the population linear projection of a on $(1, x, a_c, x_c)$, so that ν is uncorrelated with the regressors by construction. For any variable z, let $z^{\perp}$ denote the residual from the linear projection of z on $(1, x)$; the residualized moments below coincide with the conditional-on-x moments when conditional expectations are linear in x and conditional second moments do not vary with x (e.g., under joint normality).

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

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

Since $\beta_{x_c} < 0$ and $\mathrm{Cov}(x_c^{\perp}, a_c^{\perp}) > 0$, the bias term is negative. If $|\beta_{x_c}| \cdot \mathrm{Cov}(x_c^{\perp}, a_c^{\perp})/\mathrm{Var}(a_c^{\perp})$ 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:

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

Since $\beta_{a_c} > 0$ and $\mathrm{Cov}(a_c^{\perp}, x_c^{\perp}) > 0$, 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 B.1: 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
	Random (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 B.3: 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.

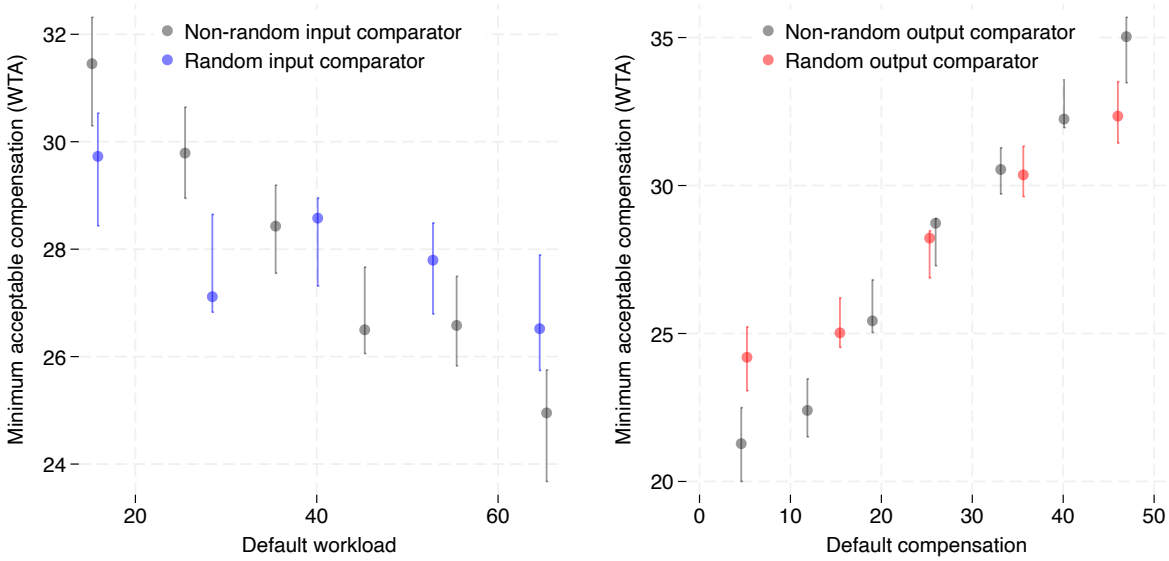

Figure B.1: Results for treatments *Random* and *Expectations Both CP* in the *Effort Valuation* paradigm. The left-hand side panel shows average minimum acceptable compensations as a function of the input comparator (the default workload). The right-hand side panel shows average minimum acceptable compensations as a function of the output comparator (the default compensation).

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 C.1, 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). For reference, columns (1), (4), (6), (8), and (11) of Table C.1 reproduce the correctly-specified benchmark regressions that include exactly the comparison points subjects observed; the remaining columns are the deliberately misspecified variants.

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 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 C.1: 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.[22] 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 I.6 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 3 and Appendix Figure D.1 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.[23] The uncertainty manipulation decreases the effect of the input comparator by 44% and that of the output comparator by 42%.

[22]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 I.31 for a screenshot.

[23]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).

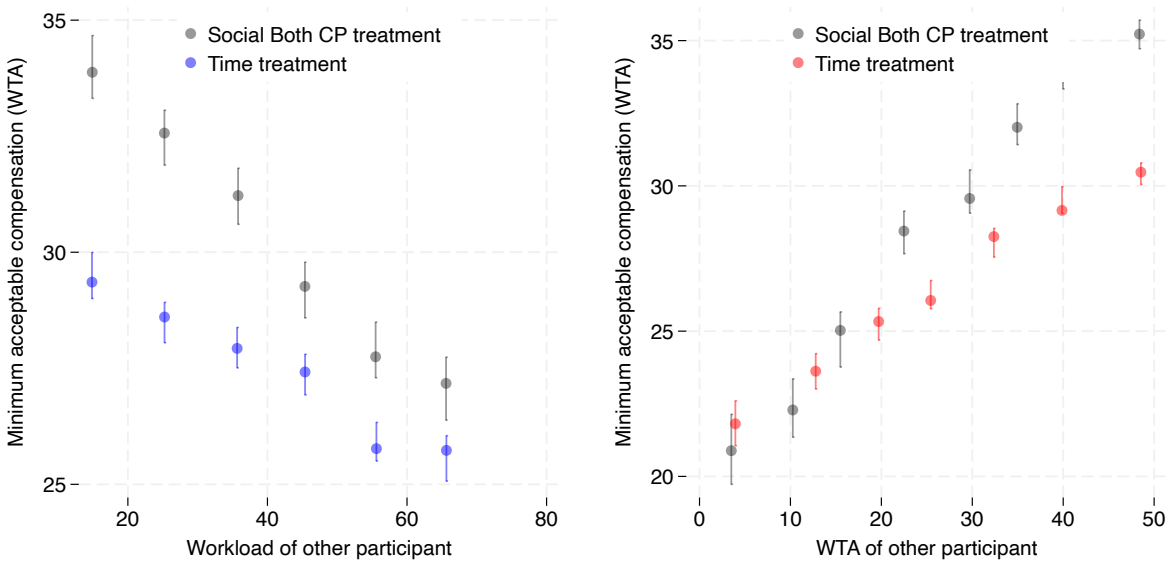

Figure D.1: Results for treatment *Time* in the *Effort Valuation* paradigm. The left-hand side panels show average minimum acceptable compensations as a function of the input comparator (other subjects' workload). The right-hand side panel shows average minimum acceptable compensations as a function of the output comparator (other subjects' WTA).

E Results of Forecaster Survey

Table E.1: Forecaster prediction accuracy by paradigm and comparator type

			Share predicting (%)				
Paradigm	Comparator	*N*	Sig. pos.	Sig. neg.	Not sig.	Don't know	Correct (%)
Panel A: Output comparators							
Effort Valuation	WTA of other (output)	49	**77.6**	4.1	14.3	4.1	**77.6**
Effort Choice	Default workload (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. The *Investment* prompt contained two typos as fielded (see the note to Figure I.48); the Investment rows, particularly the output-comparator row, should be read with this caveat.

F Robustness Checks and Sensitivity Analyses

All experiments reported in this paper were pre-registered prior to data collection; Appendix Table B.1 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 F.1, Tables F.1–F.5). Second, we assess robustness to the estimation method by replacing OLS with median (quantile) regressions, which are less sensitive to outliers (Section F.2, Tables F.6–F.7). 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 F.3, Tables F.8–F.9), whether comparison effects differ depending on whether the comparison is "favorable" or "unfavorable" (Section F.4, Table F.10), and additional specifications for the effort choice paradigm and mechanism treatments (Sections F.5–F.6, Tables F.11–F.14).

The conclusions from our main specifications are broadly robust: output comparators produce assimilation effects and input comparators produce contrast effects, with coefficients that are generally quantitatively similar to those reported in the main text. A few estimates in the smaller subsamples lose statistical significance—for example, the input-comparator effects in the median regressions for the *Input CP only (social)* and *Beliefs* treatments (Table F.6) and in several of the favorable/unfavorable splits (Table F.10)—but the sign pattern and magnitudes remain consistent with the main results across paradigms, mechanism treatments, and specifications. 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 F.8–F.9)—and that the direction of the effects does not depend on whether the *input* comparison is "favorable" or "unfavorable": output assimilation effects appear in both input-favorability subsamples with similar magnitudes, and input contrast effects point in the same direction in most subsamples, though some split estimates are individually insignificant and the favorable *Investment* split yields a small, insignificant positive input-comparator coefficient (Table F.10).

F.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 F.1: 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)*	*Input CP only (social)*	*Output 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 F.2: 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)*	*Input CP only (social)*	*Output 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 F.3: 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)*	*Input CP only (social)*	*Output 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 F.4: 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 F.5: Determinants of Effort Valuation

Dependent Variable:	Minimum Acceptable Compensation (WTA)							
Paradigm:	**(1) Effort Valuation**			**(2) Effort Valuation**		**(3) Effort Valuation**		
Mechanism (M):	*Time (social)*			*Random (expectations)*		*Deliberation (social)*		
Sample / Spec:	Strict human verification	Response-time validity	Protocol compliance	Strict human verification	Protocol compliance	Strict human verification	Response-time validity	Protocol compliance
Main Effects								
Input (workload)	0.246***	0.243***	0.249***	0.199***	0.192***	0.231***	0.237***	0.233***
	(0.0133)	(0.0113)	(0.0113)	(0.0192)	(0.0169)	(0.0173)	(0.0163)	(0.0165)
Input comparator (workload)	-0.153***	-0.143***	-0.146***	-0.137***	-0.124***	-0.113***	-0.118***	-0.117***
	(0.0136)	(0.0121)	(0.0120)	(0.0168)	(0.0150)	(0.0128)	(0.0119)	(0.0119)
Output comparator (WTA)	0.338***	0.326***	0.333***	0.334***	0.321***	0.319***	0.340***	0.333***
	(0.0192)	(0.0170)	(0.0169)	(0.0279)	(0.0251)	(0.0272)	(0.0248)	(0.0251)
Interaction Effects (Mechanism × Variable)								
M × Input (workload)	0.162***	0.169***	0.163***	-0.0518**	-0.0393*	0.134***	0.135***	0.132***
	(0.0183)	(0.0158)	(0.0157)	(0.0251)	(0.0235)	(0.0227)	(0.0206)	(0.0209)
M × Input comparator (workload)	0.0722***	0.0661***	0.0685***	0.109***	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 × Output comparator (WTA)	-0.144***	-0.139***	-0.145***	-0.120***	-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	4,471	5,470	6,695	6,864
R-squared	0.735	0.730	0.730	0.676	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.

F.2 Median (Quantile) Regressions

Table F.6: 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)*	*Input CP only (social)*	*Output 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 $\tau = 0.5$. Bootstrap standard errors in parentheses are clustered at the subject level (200 replications). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Unlike the corresponding OLS specifications, the median regressions are estimated without subject or task fixed effects, as the fixed-effects quantile bootstrap does not reliably converge. In the mechanism columns, the regressors are interacted with the treatment indicator and the baseline coefficients refer to the comparison condition.

Table F.7: Median regression ($\tau = 0.5$): Effort Valuation Mechanisms and Effort Choice

Paradigm:	**Effort Valuation (WTA)**			**Effort Choice (Workload)**	
Treatment:	**Deliberation (social)**	**Time (social)**	**Random (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)	0.318***	0.367***	0.286***	27.78**	40.45***
	(0.0197)	(0.0163)	(0.0298)	(11.67)	(7.949)
Input comparator (Workload/Piece Rate)	-0.126***	-0.181***	-0.223***		-17.33***
	(0.0179)	(0.0165)	(0.0228)		(5.267)
Output comparator (WTA/Workload)	0.557***	0.565***	0.575***		0.247***
	(0.0319)	(0.0209)	(0.0372)		(0.0555)
Interactions					
Treatment × Input (Workload)	0.156***	0.117***	-0.0567		
	(0.0262)	(0.0231)	(0.0477)		
Treatment × Input comparator (Workload)	-0.00474	0.0494**	0.159***		
	(0.0241)	(0.0214)	(0.0469)		
Treatment × Output comparator (WTA)	-0.240***	-0.306***	-0.142**		
	(0.0406)	(0.0338)	(0.0568)		
Observations	6,744	8,850	4,456	2,175	3,690
Subjects	450	590	298	145	246

Notes. Median (quantile) regressions with $\tau = 0.5$. Bootstrap standard errors in parentheses are clustered at the subject level (200 replications). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Unlike the corresponding OLS specifications, the median regressions are estimated without subject or task fixed effects, as the fixed-effects quantile bootstrap does not reliably converge. In the mechanism columns, the regressors are interacted with the treatment indicator and the baseline coefficients refer to the comparison condition.

F.3 Round Interactions

Table F.8: 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	0.00107		
	(0.00238)		
Round × Input comparator (workload)	0.00104		
	(0.00192)		
Round × Output comparator (WTA)	-0.00345		
	(0.00295)		
Time × Round × workload	0.00397		
	(0.00281)		
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)		3.497***	0.946
		(0.660)	(0.641)
Round × Input comparator (piece rate)			0.583
			(0.576)
Round × Output comparator (workload)			-0.00637*
			(0.00260)
Observations	8,850	2,175	3,690
Subjects	590	145	246
R-squared	0.733	0.692	0.657

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 F.9: 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)	(0.0321)		
Round × Input comparator (winning chance)		0.0351		
		(0.0348)		
Round × Output comparator (amount wagered)		-0.00236		
		(0.00317)		
Panel B: Beliefs Regressors				
Round × Input (log prior)			0.00784	0.000789
			(0.00550)	(0.00544)
Round × Input (log likelihood ratio)			0.00528	0.00276
			(0.00428)	(0.00351)
Round × Input comparator (log prior)				0.000824
				(0.00568)
Round × Input comparator (log likelihood ratio)				0.00391
				(0.00517)
Round × Output comparator (log posterior)				-0.00310
				(0.00460)
Observations	2,175	3,705	2,180	3,610
Subjects	145	247	148	246

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.

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

Table F.10: 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)	0.241***	0.163***	0.212***	0.106***	0.215***	0.158***	28.34**	44.07***	8.315***	7.800***
	(0.020)	(0.018)	(0.024)	(0.016)	(0.019)	(0.018)	(8.70)	(6.09)	(0.36)	(0.40)
Input comparator (workload/piece rate/winning chance)	-0.085***	-0.118***	-0.114***	-0.096***	-0.022	-0.031	-3.894	-8.451*	0.235	-0.789**
	(0.019)	(0.019)	(0.026)	(0.021)	(0.019)	(0.019)	(7.12)	(4.12)	(0.33)	(0.28)
Output comparator (WTA/workload/amount wagered)	0.326***	0.305***	0.316***	0.363***			0.119***	0.112***	0.067**	0.087***
	(0.020)	(0.023)	(0.025)	(0.024)			(0.019)	(0.018)	(0.023)	(0.017)
Observations	1765	1894	1594	2094	1823	1912	1518	2172	1897	1808
Subjects	244	243	244	246	249	249	246	246	247	247
R-squared	0.723	0.720	0.698	0.692	0.760	0.757	0.730	0.678	0.736	0.705

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 subject compares favorably or unfavorably to the comparison point, based on the *input* comparator only; the output comparator plays no role in the split. A round is classified as favorable if the subject's own input is better than the comparator's: the comparator's workload exceeds the subject's own workload (*Effort Valuation*), the comparator's piece rate is below the subject's own (*Effort Choice*), or the comparator's winning chance is below the subject's own (*Investment*). Ties are classified as unfavorable. All regressions include subject fixed effects. Regressions for Effort Valuation and Effort Choice also include task-type fixed effects.

F.5 Effort Choice – Additional Specifications

Table F.11: 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 F.12: 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.

F.6 Additional secondary analysis – Random (expectations) and Deliberation (social)

Table F.13: Random (expectations) – decision-making self-report

Group	Subjects	Share considering default workload and/or compensation
Uninformative (random)	145	68.3%
Informative (control)	153	71.9%

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

Table F.14: Deliberation (social) – per task analysis

Effort Valuation only: split by task type (OLS)

	Counting	Sliders	Greek
	(1)	(2)	(3)
Input (workload)	0.309***	0.449***	0.331***
	(0.0205)	(0.0228)	(0.0204)
Input comparator (workload)	-0.0635***	-0.104***	-0.0492***
	(0.0111)	(0.0197)	(0.0103)
Output comparator (WTA)	0.233***	0.199***	0.178***
	(0.0261)	(0.0270)	(0.0229)
Observations	1497	1498	1500
Subjects	100	100	100
R-squared	0.793	0.803	0.832

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.

G Classification of Prior Studies

Table G.1: Classification of prior studies, with independent LLM coding

Paper	Relevant inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	Null (no effect)
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	
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) Assimi-lation	Consistent	
Andrikogiannopoulou and Papakon-stantinou (2018)	Bet odds and matchup (individual account-level data)	Whether/how much to bet	Whether team wins, bet payoff	Whether team recently won	Over-bet winners, under-bet losers; none profit	Output (outcome)	Assimilation	Consistent	

Continued on next page

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
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	Out of scope
Baquero and Verbeek (2022)	Fund performance (controlled)	Net flow into/out of fund	Whether fund subsequently wins, return	Whether fund recently won	Flows rise with streak length	Output (outcome)	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	

Continued on next page

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Bartling and Schmidt (2015)[24]	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	
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	
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	Input CP – Assimilation – Inconsistent
Beggs and Graddy (2009)[25]	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	

Continued on next page

[24]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.

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Bhargava and Fisman (2014)[26]	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	
Bindler and Hjalmarsson (2019)[27]	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	
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	
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	Output (decision) CP – Contrast – Inconsistent

Continued on next page

[26]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 the analysis; participants only see partner attributes.

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Bracha (2020)	Risk-free rate, lottery returns	Allocation between safe asset and risky lottery	Portfolio return	Zero nominal return (sign of the risk-free rate)	Risk-taking does not change significantly when the risk-free rate turns negative	Input	Null	Null effect	Out of scope
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	Output (outcome) CP – Contrast – Inconsistent
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	
Brown and Sauer (1993)	NBA team fundamentals (aggregate market only)	Market point spread on the game (implied win prob.)	Whether team wins, bet payoff	Whether team recently won	Spreads shift toward streak	Output (outcome)	Assimilation	Consistent	
Burson et al. (2013)[28]	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	

Continued on next page

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
Camerer (1989)	NBA team fundamentals (aggregate market only)	Market point spread on the game (implied win prob.)	Whether team wins, bet payoff	Whether team recently won	Spreads shift toward streak	Output (outcome)	Assimilation	Consistent	
Cao et al. (2020)[29]	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	Out of scope
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	

Continued on next page

[29]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 G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
Cerulli-Harms et al. (2019)[30]	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	Null (no effect)
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	
Chang et al. (2019)[31]	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	Out of scope
Chen et al. (2016)[32]	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	

Continued on next page

[30]We ignore the experiments with null results.
[31]Paper primarily studies an additional framing manipulation.
[32]Prior case's true quality (input comparator) is unobserved and plausibly correlated with the lagged decision.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
Clotfelter and Cook (1993)	Daily 3-digit lottery, equal odds	Which number, how much	Whether number wins, payout	Whether number recently won	Bets drop just after a number wins, recover over months	Output (outcome)	Contrast	Inconsistent	
Cohn et al. (2014)[33]	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. (2015)[34]	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	
Corgnet et al. (2018)	Task parameters, wage contract	Effort provision	Production	Goal (target production level)	Production increases in goal	Output (outcome)	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	

Continued on next page

[33]We ignore treatment with null effect.
[34]We ignore null result for non-reasonable goals.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Czibor et al. (2022)[35]	Task, opportunities for theft/helping	Effort allocation (task, theft, helping)	Final wealth	Initial wealth	More stealing when initial wealth higher	Output (outcome)	Assimilation	Consistent	Out of scope
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	
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	Out of scope
Dhar and Zhu (2006)[36]	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	Out of scope

Continued on next page

[35]We ignore the null results for other outcome variables.
[36]Paper focused on effect of sophistication on disposition effect.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
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	Out of scope
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)[37]	Trivia question	Belief report	Belief	Random anchor (20 or 80)	Beliefs increase in anchor	Output (decision)	Assimilation	Consistent	Out of scope
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	Out of scope

Continued on next page

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	Out of scope
Fehr et al. (2022)[38]	Attributes of goods	Keep vs. exchange	Ownership	Assigned initial ownership	Prob. keeping increases in assigned ownership	Output (outcome)	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	Out of scope
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	

Continued on next page

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Fudenberg et al. (2012)	Attributes of goods and lotteries	Stated valuation (WTP or WTA)	Good ownership and payment made	Randomized numeric anchor prompt (SSN-based number)	Valuations do not significantly increase in the numeric anchor (very weak effects for goods, none for lotteries)	Output (decision)	Null	Null effect	
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	
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	Out of scope
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	

Continued on next page

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
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	Out of scope
Heffetz and List (2014)[39]	Characteristics of goods	Keep vs. exchange	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent	Null (no effect)
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	

Continued on next page

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	Out of scope
Imas et al. (2017)[40]	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	Out of scope
Iturbe-Ormaetxe et al. (2011)[41]	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	Out of scope
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	
Kiessling et al. (2022)	Characteristics of task ("suicide run" in a PE class)	Effort (running)	Performance (time)	Peer performance (time)	Performance regresses towards peer performance	Output (outcome)	Assimilation	Consistent	Out of scope

Continued on next page

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

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Knetsch (1989)	Good characteristics	Keep vs. exchange	Final ownership	Initially assigned ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent	
Kong et al. (2020)[42]	Single-winner lottery, known odds	Which drawn number, bet share	Whether number wins, payout	Whether number recently won	Lower bet share on recent winners	Output (outcome)	Contrast	Inconsistent	Input CP – Contrast – Consistent
Kong et al. (2020)	2–3 winner lottery, known odds	Which drawn number, bet share	Whether number wins, payout	Whether number recently won	Higher bet share on recent winners	Output (outcome)	Assimilation	Consistent	
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)[43]	Property attributes	Counteroffer	Price paid	Asking price	Counteroffers gravitate toward asking price	Output (decision)	Assimilation	Consistent	Out of scope
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	

Continued on next page

[42]This paper reports two effects in different lottery designs: a gambler's-fallacy contrast effect (single-winner) and a hot-hand assimilation effect (multiple-winner). Both are classified here.

[43]We classify baseline effect; paper also studies effect of round vs. precise asking prices.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
List (2003)[44]	Attributes of goods	Keep vs. exchange	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent	
List (2004)[45]	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	
Loh and Warachka (2012)	Earnings-surprise size (controlled)	Market's pricing of the stock at the announcement	Subsequent abnormal return drift	Whether prior surprises have the same sign (streak)	Announcement prices underweight streak-continuing surprises, as if expecting reversal: drift is large when a streak extends, negligible when it ends	Output (outcome)	Contrast	Inconsistent	
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	

Continued on next page

[44]Baseline endowment effect for inexperienced traders. Paper shows effect diminishes with market experience.
[45]We ignore the null result in the experts sample.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	Null (no effect)
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)[46]	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	
Plott and Zeiler (2007)[47]	Attributes of goods	Keep vs. exchange	Final ownership	Initially assigned ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent	Null (no effect)

Continued on next page

[46]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.

[47]Baseline endowment effect with traditional procedures. Paper shows effect disappears when procedures are modified.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
Pope and Simonsohn (2011)[48]	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	
Pope and Schweitzer (2011)[49]	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	
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	
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	
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

[48]No variation in threshold

[49]Analysis does not exploit variation in par (uses par-fixed effects).

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	Out of scope
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)[50]	Characteristics of goods	Valuation (WTA or WTP)	Final ownership	Initial ownership	Final ownership increases in initial ownership	Output (outcome)	Assimilation	Consistent	
Schwerter (2024)[51]	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	

Continued on next page

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

[51]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 G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
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)[52]	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	Out of scope
Suetens et al. (2016)	Weekly lotto, equal odds	Which numbers, how much	Whether number wins, payout	Whether number recently won	Less staked on recently-winning numbers	Output (outcome)	Contrast	Inconsistent	

Continued on next page

[52]We treat the constant option's risk as a comparison level of zero when it is a certain amount, so certain vs. stochastic constitutes magnitude variation in the comparator's risk level rather than variation in its type.

Table G.1 continued

Paper	Inputs	Decision (output)	Outcomes (output)	Comparison point	Finding	Type of CP	CP effect	Classi-fication	LLM disagreement?
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	
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)[53]	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	
Yu et al. (2021)[54]	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	

[53]Goal classified as Input (not Output as in other goal papers) because the decision is a happiness judgment, not the goal-directed action itself.
[54]Classification of results from treatment with generally accurate delay information

Notes. The column "LLM disagreement?" reports an independent classification produced by a large language model (Fable 5; see Appendix H for the instructions). A blank cell indicates that the model agreed with our four-way classification (consistent / inconsistent / null / not classifiable). Where it did not, the cell reports the model's verdict: "Out of scope" where the model judged the study to fall outside the classification (e.g. pure framing, a fixed reference point, or no variation in the comparator's magnitude), "Null (no effect)" where it judged the study in scope but found no significant effect, and otherwise the model's own comparator type, direction, and consistency. Most filled cells reflect a differing scope judgment rather than an opposing finding: among the studies it places in scope the model agrees with our classification in the large majority of cases, reaching the opposite directional conclusion for only four (Bateman et al., 2009; Bordalo et al., 2016; Brandts and Solà, 2001; Kong et al., 2020)

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.

Andrikogiannopoulou, Angie, and Filippos Papakonstantinou. 2018. "Individual Reaction to Past Performance Sequences: Evidence from a Real Marketplace." *Management Science*.

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.

Baquero, Guillermo, and Marno Verbeek. 2022. "Hedge Fund Flows and Performance Streaks: How Investors Weigh Information." *Management Science*.

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.

Bracha, Anat. 2020. "Investment Decisions and Negative Interest Rates." *Management Science* 66 (11): 5316–5340.

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.

Brown, William O., and Raymond D. Sauer. 1993. "Does the Basketball Market Believe in the 'Hot Hand'? Comment." *American Economic Review* 83 (5): 1377–1386.

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.

Camerer, Colin F. 1989. "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review* 79 (5): 1257–1261.

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, 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.

Clotfelter, Charles T., and Philip J. Cook. 1993. "The Gambler's Fallacy in Lottery Play." *Management Science* 39 (12): 1521–1525.

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.

Fudenberg, Drew, David K. Levine, and Zacharias Maniadis. 2012. "On the Robustness of Anchoring Effects in WTP and WTA Experiments." *American Economic Journal: Microeconomics* 4 (2): 131–145. 10.1257/mic.4.2.131.

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.

Kong, Qingxia, Georg D. Granic, Nicolas S. Lambert, and Chung Piaw Teo. 2020. "Judgment Error in Lottery Play: When the Hot Hand Meets the Gambler's Fallacy." *Management Science* 66 (2): 844–862.

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.

Loh, Roger K., and Mitch Warachka. 2012. "Streaks in Earnings Surprises and the Cross-Section of Stock Returns." *Management Science* 58 (7): 1305–1321.

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.

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.

Suetens, Sigrid, Claus B. Galbo-Jørgensen, and Jean-Robert Tyran. 2016. "Predicting Lotto Numbers: A Natural Experiment on the Gambler's Fallacy and the Hot-Hand Fallacy." *Journal of the European Economic Association* 14 (3): 584–607.

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.

H LLM Classification Instructions

The following are the verbatim instructions provided to the language model classifier. They are reproduced exactly as given to the model, including one typo in worked example 9.6, whose "Input(s)" field reads "Own hours" where the input is the per-tool spend; the example's classification logic is unaffected.

0. Neutrality rules (read first)

These rules override any tension elsewhere in the document.

0.1 **No preferred answer.** Contrast and assimilation are equally expected. Consistent, Inconsistent, Null, and N/A are equally legitimate outcomes. Nothing in your reading should raise the prior on one conclusion.

0.2 **Classify what the paper shows, not what the taxonomy predicts.** First establish the paper's empirical finding and its direction. *Only then* check whether it is consistent or inconsistent with the prediction. Never reason backward from "this should be assimilation" to a reading of the evidence.

0.3 **Gate ≠ conclusion.** Deciding that a paper is *in scope* (it has a measured comparator whose level varies, §3–§7) is a separate, earlier step from deciding the *conclusion* (Consistent / Inconsistent / Null). Recognizing a familiar design — endowment, default, anchor, goal — tells you only that the paper is in scope. It does **not** tell you the answer. An in-scope endowment or default study may come out Null (no effect) or Inconsistent. Read type and direction from the data every time.

0.4 **N/A is not a failure.** Many papers will lack a measured comparator whose level varies. Do not stretch a paper to fit. But do not over-apply N/A either: apply the §7.3 level test exactly, and remember that zero, a probability, and a quantity are all levels.

0.5 **Classify the paper's primary, clearly-identified comparator effect** — the one the authors foreground. Do not hunt across subgroups, secondary outcomes, or interactions to manufacture a significant effect. When several comparator effects qualify, classify the standard baseline reference-dependence effect (§4, step 3a) and ignore the others.

0.6 **Report disconfirming and null evidence at face value.** If the cleanest reading of the primary result contradicts the prediction → **Inconsistent**. If the comparator has no significant effect → **Null**.

0.7 **No design gets a pre-assigned answer.** This document teaches you *how to decide*; it never tells you the answer for any particular method.

0.8 **Mechanism-agnostic scope.** Scope is determined by the *design and the measured effect* — never by the mechanism the **authors** advance to interpret it. That a paper frames its finding as social preferences, fairness views, reciprocity, norms, self-image, or motivated reasoning neither puts it in scope (if no level-varying comparator exists) nor takes it out of scope (if one does). Route by structure, not by the paper's interpretive label. This cuts both ways and raises the prior on no particular conclusion.

1. Task

Classify each empirical economics paper by whether its findings exhibit **contrast** or **assimilation** effects relative to a **comparator** (comparison point), and whether this is **consistent** or **inconsistent** with the Graeber–Enke input/output taxonomy — or whether the paper has no classifiable comparator effect (**N/A**) or a **Null** effect.

For each paper: look it up, identify the empirical design, and either fill in a classification or mark N/A with a one-line reason.

2. Framework (the prediction under test)

2.1 Intuition

Decision-makers are often uncertain about how to map problem inputs (e.g., the toughness of a piece of wood, the humidity of flour, a contest entry's craftsmanship) into optimal decisions (e.g., minutes spent carving, kneading time, a score to assign). Comparison points may reduce this uncertainty by signaling something about the input-to-output mapping. The taxonomy's central claim is that **input comparators and output comparators carry opposite informational content**:

- A high **output** comparator (e.g., a fellow hobbyist's high practice hours) suggests the mapping is "high output per unit of input," predicted to **pull the decision toward it → assimilation**.
- A high **input** comparator (e.g., a fellow hobbyist's high stipend) suggests "each unit of input warrants less output," predicted to **push the decision away from it → contrast**.

This is the hypothesis. The data may or may not bear it out for any given paper.

2.2 Predictions

Comparator type	Predicted effect	Direction relative to comparator
Input comparator	Contrast (predicted)	Decision moves *opposite* to comparator
Output comparator	Assimilation (predicted)	Decision moves *toward* comparator

A paper is **Consistent** if its measured direction matches the prediction for the comparator type, **Inconsistent** if it runs the other way. Bunching at a reference point is one possible form of assimilation, but treat it as evidence to be evaluated, not as an automatic Consistent.

2.3 Formal statement

Consider a decision problem `max_a u(a, x)` with optimal action `a*(x, θ) = θ · x`, where `x` is the decision input and θ is the unknown input-to-output mapping. The DM observes comparison data generated by `a_c = θ_c · x_c`, where `θ_c` is informative about θ.

- **Predicted direction:** Action assimilates in the output comparator `a_c` ($\partial a / \partial a_c > 0$) and contrasts in the input comparator `x_c` ($\partial a / \partial x_c < 0$).
- **Predicted magnitude:** Comparator effects are predicted to be larger when the DM's own mapping uncertainty is higher and when the comparison signal is more informative.

The classification records whether each paper's measured effect matches these predictions; it does not assume it does.

3. Core definitions

Term	Definition
Decision input	A variable taken as given by the decision-maker for the studied decision (e.g., today's flour humidity for a bake-or-wait decision).
Decision output	The decision itself or its outcome (e.g., "bake now or wait"; "loaf volume").
Input comparator	A comparison point in the same space as a decision input — context brought to the decision (e.g., a fellow baker's reported humidity, the toughness of a previously whittled piece).
Output comparator	A comparison point in the same space as the decision or its outcome — a benchmark in the choice/outcome space (e.g., a recipe yield, a default stitch count, a peer's pottery output).
Contrast	Decision moves *opposite* to the comparator. A higher comparator makes the same input feel worse/smaller, pushing behavior away.
Assimilation	Decision moves *toward* the comparator. A higher comparator pulls the decision toward it. Bunching is one form.
Comparator level	The value the comparator takes in a given condition, on its ordered dimension. **Zero/absent is a level. A probability of the reference state is a level. A quantity or degree is a level.** (See §7.3.)

The same variable can be either input or output. The classification depends on the decision the *analyst is studying*, not on the variable itself. This is a genuine ambiguity to be resolved by §5, not a license to pick whichever reading confirms the prediction.

Hypothetical — original yarn cost. For a binary **finish-or-frog** decision (continue knitting or unravel), original yarn cost is plausibly an **input** comparator. For a **gift-pricing** decision over the finished piece, it is plausibly an **output** comparator (both in dollar space). The classification follows the decision being studied; the directional effect is whatever the data show.

4. Classification procedure

For each paper, in order:

1. **Look it up.** Identify the empirical design — what varies, what is measured, what decision the subjects make. If the abstract is unclear, read the empirical specification; comparators sometimes only become visible there.

2. **Identify the decision and its inputs.** What does the subject choose? What information is given to them as fixed?

3. **Look for a comparator whose LEVEL varies** — a thing observed in the data that serves as a comparison standard for the decision and takes at least two distinct levels across conditions or observations. **A move between absent/zero and a positive level, a change in the probability of the reference state, and a change in quantity/degree all count as level variation** (see §7.3). If only the comparator's *label/frame/identity* or the *number/frequency of comparisons* differs while its level is fixed → **N/A** (see §7).

   Assess scope per foregrounded comparator effect, not per paper. A paper is NOT made N/A merely because it *also* contains a pure-framing arm, a moderator manipulation, or a higher-profile multidimensional headline — if it *separately foregrounds* a measured, level-varying comparator effect, classify that effect. Conversely, a quantity that varies only *incidentally* (e.g., an endowment varied 0–10 solely to span a give/take choice space) while the foregrounded result is framing does **not** put the paper in scope. The test is whether the paper documents a foregrounded reference-dependent effect of a level-varying comparator — not whether *any* number happens to vary. **"Foregrounded" means reported as a treatment-level result — not that it must be the paper's headline contribution.** A measured, level-varying comparator effect that the authors report as a finding is in scope even when the paper's main thesis concerns a different mechanism (e.g., the comparator arm is one of several treatments in a paper about motivated reasoning or social image). "Foregrounded" excludes subgroup-mined or buried-robustness results; it does not exclude results the authors report but interpret through another mechanism (§0.8).

   Step 3a — arm selection when several comparator effects qualify. A paper may contain more than one classifiable comparator effect (e.g., a standard anchoring arm on the central decision variable plus a second anchored quantity elsewhere in the design, or an arm interacting the comparator with a moderator). Classify the **standard baseline reference-dependence effect**: the arm that most closely replicates a canonical comparator design (an anchor on the decision variable itself, a default, an endowment, a goal, a

peer level) with the fewest added manipulations, whose comparator lives closest to the paper's **central decision variable**. If such a baseline arm exists, classify it and **ignore all other classifiable arms** — do not let a derivative, auxiliary, or moderator-interacted comparator effect substitute for it. (This rule selects *which* effect to classify; it never makes a paper N/A. A paper whose headline treatment is unclassifiable is still classified on its baseline arm.)

4. **Classify the comparator type** (Input vs. Output) using §5. If the type genuinely cannot be resolved, say so and mark N/A.

5. **Establish the primary measured effect and its direction** *before* comparing to the prediction.

6. **Apply the contrast/assimilation test** (§6) and record Consistent, Inconsistent, or Null based on whether the measured direction matches the prediction for that comparator type. (Recall §0.3: being in scope does not imply Consistent.)

7. **Nulls.** If the paper's primary comparator test is a null → **Null**. If a paper has a clear primary significant effect *and* a separately reported, substantively important null, you may add a **second Null entry** — but do not use secondary results to override the primary finding.

8. **Output** the classification in the format of §8.

Guiding heuristics (neutral)

- **Stay close to the paper's own framing** of the decision and the reference point. When multiple readings are defensible, follow the paper's own language — not the reading that best fits the prediction.
- **Focus on what is empirically measured**, not on theoretical interpretation.
- **Endogenous and observational comparators are fine** as long as the comparator is measured (see §7.2b on model-identified references) and its level varies.
- **Beliefs and judgments count as decision variables.**
- **Classify the primary comparator effect**, not a subgroup result selected for its direction.
- **Apply the same evidential bar in both directions.** Classify when there is a measured comparator whose level varies with directional evidence; mark N/A when there is not. Neither default.

5. Determining comparator type (Input vs. Output)

This is the hardest step. The questions below are **diagnostics, not look-up answers** — work through the specific paper, do not match it to a named method and read off a verdict. Apply in order; the first that resolves wins.

Rule A — Classify the decision as the subject actually performs it

Many papers reconstruct a continuous latent variable (WTP, reservation price) from a discrete task (accept/reject). Classify against the **actual task the subject performs**, not the analyst's reconstruction.

- *Is the decision the subject actually makes binary/discrete, or a genuine continuous choice?*
- *Is the numeric comparator in the same space as the subject's actual choice, or in the space of a quantity the analyst infers?* A numeric comparator not in the space of the actual (binary) choice is contextual information — consider Input. One in the space of a genuine numeric choice — consider Output.
- *Does a past value enter as a fixed fact that colors the current state (keep/discard), or as a benchmark the current choice is set against?* Former → Input, latter → Output.
- **Exception — optional stopping.** When the binary choice is a stop/quit/continue/exit decision that accumulates into (or terminates at) a quantity in the comparator's units, do **not** apply the binary-choice reading; classify per Rule B's stopping clause below.

Work out which case the paper is in, then read the measured direction.

Rule B — "Same units" test for output comparators

A comparator is **Output** when measured in the same units, on the same dimension, as the decision or its outcome. *Could you plot the comparator and the decision/outcome on the same axis?* If yes → Output.

- Prior loaf weight (g) and new loaf weight (g) → Output.
- Recommended practice minutes and actual practice minutes → Output.
- A performance benchmark in performance units and the corresponding performance → Output. Whether performance moves toward it (assimilation) or away (contrast) is empirical.

- When the decision is made *through* an instrument (choosing a plot, watering effort), classify on the outcome the DM effectively controls and that lives in the comparator's units.
- **Optional-stopping/threshold decisions.** A stop/quit/continue/exit decision that determines a cumulative or final quantity in the comparator's units (accumulated earnings against an earnings reference; final winnings against a winnings reference; an accepted price against a price reference) is classified against that quantity → **Output**. That the instantaneous decision is binary does not make the comparator an Input. Whether the quantity moves toward the comparator (assimilation) or away (contrast) is read from the data.

Rule C — "Different units" test for input comparators

A comparator is **Input** when it provides contextual information shaping how the problem is *perceived*, but is **not** in the same units as the decision.

- Neighbor's fertilizer cost ($) and own weeding effort (min). Different units → Input.
- Announced expected wait time and a stay/leave decision. Information vs. binary choice → Input.
- Prior reward level and effort on a later task with a different reward scheme → Input.

The *type* is set by units/role; the *direction* is whatever the paper measures.

Rule D — Tiebreakers (use sparingly)

1. **Mechanism check.** Does the comparator work by *setting a benchmark the decision is measured against* (Output) or by *changing how the DM perceives an otherwise-fixed situation* (Input)? Decide on the mechanism the paper evidences, not on which yields a cleaner prediction.
2. **If still genuinely unresolved, mark N/A** with a note that the type is indeterminate. Do **not** default to whichever type makes the result Consistent.

Special case — goals and targets

A goal/target is often an **Output** comparator (performance units). But it can be **Input** when the DV is a *subjective evaluation* of a fixed outcome — the goal re-colors how the same outcome feels. Decide from the DV (objective performance vs. subjective rating), then read direction from the data.

Special case — expectation-based disappointment/elation

When a reference lives in *outcome* space (an expected win, an expected payoff) but drives a **downstream behavioral decision** (exit, breach, quit) by *re-coloring how the decision-maker's own realized outcome feels*, classify the reference as an **Input** comparator, not Output (a higher expectation that makes the same own-outcome feel worse and pushes the decision away is **Contrast → Consistent**).

This rule fires ONLY if ALL THREE of the following hold — otherwise it does **not** apply and the paper is routed by the ordinary rules:

1. **Measured & anchored.** The reference is a *measured level that varies* and is anchored to observed data (passes §7.2b) — **not** a calibrated or latent belief object inferred only from the behavior it explains (e.g., a conjectured "daily income target" that is never a regressor → N/A under §7.2b).

2. **Scalar level, not risk-character.** The variation is in a scalar level, not in the *timing/uncertainty/risk-character* of the reference (§7.3).

3. **Cross-space only.** The reference is **not** in the units of any quantity the DM's decision determines. If the decision determines a cumulative or final quantity in the reference's units (accumulated earnings, final winnings, an accepted price or offer), this rule does **not** apply: Rule B types the comparator **Output**, and the direction — toward or away — is read from the data.

6. The contrast/assimilation test

6.1 Input comparators

1. Does the decision increase in the basic input?

2. Does the decision increase or decrease in the comparator?

3. **Opposite signs → Contrast. Same signs → Assimilation.**

Contrast on an input comparator is **Consistent**; assimilation on an input comparator is **Inconsistent**.

6.2 Output comparators

Does the decision/outcome move *toward* the comparator (assimilation) or *away* (contrast)? Assimilation on an output comparator is **Consistent**; contrast is **Inconsistent**.

6.3 Nulls

- **Primary null.** No significant effect on the main comparator test → **Null**. State the finding as "Higher [comparator] has no significant effect on [decision]."
- **Significant-plus-null.** A clear primary significant effect *and* a substantively important reported null → primary conclusion + a separate Null entry. Do not let secondary results flip the primary.
- Do not search exhaustively across subgroups to convert a null into an effect.
- **Fragile/non-robust effects default to Null.** If the authors themselves describe the primary comparator effect as *tentative, not robust, sensitive to small design changes, or insignificant under their preferred specification* — even if the point estimate runs in the predicted direction — classify it **Null**, not Consistent. A directionally-right but fragile result is not evidence for the prediction.

7. N/A criteria

Mark a paper **N/A** if **any** of the following hold. N/A is an expected outcome — but apply criteria 7.3 and 7.2b exactly; do not over-reject in-scope designs.

7.1 **Pure theory.** No empirical analysis or data.

7.2 **No measured comparator.** No comparison point is directly observed or measured in the data (see 7.2b for model-identified references).

7.2b. Model-identified reference — inclusive rule. A reference point counts as a **measured comparator (in scope)** when it is **constructed from quantities observed in the data** — realized past prices, realized recent income, the purchase/acquisition price, a lagged status quo, posted odds, an experimenter-set probability — *even if a model estimates how those observed quantities are weighted or how fast the reference adjusts.* The reference must move with observable data and be identified independently of the focal decision's own residual variation.

It is **N/A** only when the reference has **no observable anchor**: it is a free latent parameter chosen solely to best-fit the focal choices it is meant to explain, an assumed constant (e.g., zero) used in calibration, or a purely axiomatic/theoretical construct.

Calibrated/model-derived equilibrium objects are NOT measured comparator effects. A "comparator effect" that is actually a best-response slope, strategic-complementarity coefficient, or assimilation implication *derived from a calibrated or estimated structural/search model*

(rather than directly observed as a decision moving with a measured comparator) → **N/A**. The assimilation must be measured in the data, not produced by the model's equilibrium.

7.3 **No variation in the comparator's LEVEL.** N/A on this ground **only** when the comparator's level is identical across all conditions and only its *label, framing, salience, precision, or elicitation procedure* differs; OR when only the *number/frequency of comparisons* varies while each comparison's content is fixed (e.g., myopic-loss-aversion feedback-frequency designs). **The following DO count as level variation — classify them, do not N/A:**

- **Zero is a level.** Absent/zero vs. a positive level: no endowment vs. an endowment; no default vs. a default at rate r; no goal vs. a goal at level g; no anchor vs. an anchor of value v.
- **Probability is a level.** Varying the probability of the reference state (e.g., P(ownership) $= 0.1$ vs. 0.9; probability of forced exchange).
- **Quantity/degree is a level.** 1 unit vs. 5 units; full vs. partial endowment/depletion; a small vs. large endowed amount.
- **Does NOT count:** changing *which* good/category is the reference (identity, not level), or changing only how a fixed reference is described or elicited.
- **Does NOT count — proportional/ratio perception.** A fixed quantity judged by its *ratio to a varying base* (a percentage discount, a Weber-fraction, a "deeper discount," a share-of-spending) is magnitude/framing perception, not a shift in a comparator's level → N/A unless an absolute reference level itself varies across ≥ 2 values.
- **Does NOT count — risk-character change.** Changing only the reference's *risk character* (a certain vs. an uncertain/stochastic reference position) without moving a scalar level → N/A. It counts only if a scalar risk level (a variance, or a probability) is itself varied across ≥ 2 values. (This also covers a lottery-feature × gain/loss-domain interaction driven by probability-weighting / risk-seeking-in-losses — that is a risk-character effect, not a comparator-level shift → N/A.)
- **Does NOT count — a fixed distribution center.** A constant benchmark (a fixed mean/median demand, a fixed center of a single fixed distribution) toward which choices are pulled is **not** level-varying: the center never moves; only the optimum's *distance* from it changes. → N/A (the "pull-to-center" is anchoring to a fixed point, not a varying comparator).
- **Does NOT count — a strategic best-response variable.** A peer's action in a strategic-substitutes/complements game (own action moving with others' actions because

payoffs are interdependent) is a best-response effect, **not** a contrast/assimilation comparator → N/A (§7.6).

(§0.3 reminder: that a level varies puts the paper in scope; it does not set the conclusion. A 0-vs-positive endowment study may be Consistent, Null, or Inconsistent.)

7.4 **Pure framing/labeling.** The manipulation only changes how the *same* decision is described (relabeling, rhetoric, gain/loss wording, uniform payoff shifts) **with the comparator's level held identical across conditions**. This is the complement of criterion 7.3: if no measured quantity differs in level, only the frame, → N/A.

7.5 **Structural estimation / calibration only.** The reference point is a model assumption with no observable anchor (see 7.2b), not observed to vary.

7.6 **No reference-dependent mechanism — structural only.** The paper studies something else (market structure, time preferences, mental accounting, self-selection, strategic environment, etc.) **and its design contains no measured comparator whose level varies**. This criterion is structural, not interpretive: it must **never** be used to exclude a paper whose design does contain a measured, level-varying comparator with a directional effect on the DM's own decision — regardless of whether the authors interpret that effect as reciprocity, gift exchange, fairness, social preferences, trust/betrayal, self-image, or motivated reasoning (§0.8). (The §7.3 strategic best-response exclusion is unaffected: it is a payoff-structure test — the peer's action enters the DM's payoffs directly — not a mechanism-interpretation test.)

7.7 **History-dependence without a comparator.** A prior outcome affects the next decision as time-non-separable utility, not as evaluation against a reference.

7.8 **Choice-set manipulation.** The manipulation changes which options are *available* (decoy effects, adding/removing alternatives).

7.9 **Varying an option's payoff, not a comparator.** Changing the value/payoff of one choice is standard rational choice.

7.10 **Multidimensional distance comparator.** The reference is a location in a high-dimensional feature space; the key variable is *distance* from it, not a level on a single ordered dimension.

7.11 **Non-monotonic comparator effect.** A directional classification is impossible (e.g., U-shaped).

7.12 **Indeterminate comparator type.** Rules A–D cannot resolve input vs. output (§5 Rule D).

Candidate patterns to check (in both directions)

Before finalizing N/A *or* a classification, check whether any of these are present **and** whether the comparator's **level** actually varies (per §7.3 — including zero, probability, quantity) with directional evidence. The pattern label alone is not sufficient.

- Endowment effects — does the endowment level vary, **including 0 vs. owning**?
- Anchoring effects — does the anchor vary in value (**incl. anchor vs. none**)?
- Default effects — does the default level vary (**incl. default vs. none**)?
- Acquisition-cost-driven retention — does the past cost vary?
- Goal/target effects — does the target level vary (**incl. goal vs. none**)?
- Sequential evaluation — does the prior case vary; is it the prior *attribute* (input) or the prior *decision* (output)?
- Peer comparison — does the peer outcome vary?
- Status quo / habitual level — does the habitual level vary?
- Probability-of-reference-state designs — does P(reference) vary? **(In scope.)**
- Performance bunching at salient benchmarks.
- Belief formation against a reference group.

Governing question, applied symmetrically: **Is there a comparison standard observed in the data, does its LEVEL vary (zero, probability, and quantity all count), and does the paper show directional evidence of how the decision shifts relative to it?** If all three hold → classify (read direction honestly). If any fails → N/A.

Caveat — omitted variable bias

Most papers observe only *one* comparison point. Input and output comparators are often correlated, so observing one risks omitted-variable bias. Follow the paper's own framing and classify the comparison point the authors posit; note this limitation where it plausibly affects the direction (§9).

8. Output format

For each classifiable paper, produce a bulleted list:

- **Input(s):** factors taken as given by the decision-maker
- **Output: Decision:** the choice being made
- **Output: Outcome(s):** the resulting state (may overlap with decision)
- **Comparator:** the comparison point whose level varies
- **Finding:** "Higher [comparator] increases/decreases [decision/outcome]" — always state the measured direction
- **Comparator type:** Input / Output (decision) / Output (outcome)
- **CP Effect:** Contrast / Assimilation / Null
- **Conclusion:** Consistent / Inconsistent / Null

Use **Output (decision)** / **Output (outcome)** / plain **Output** per the space of the comparator. When a paper has a primary significant effect *and* a substantively important null, write two entries (§6.3).

For N/A papers: state N/A and give the reason in one line.

9. Worked examples

All examples are **hypothetical** recreational-hobby scenarios (baking, knitting, gardening, music, hiking, hobby clubs), invented so they do **not** correspond to any economics paper in the dataset. They illustrate the *logic*, including N/A, Null, and Inconsistent outcomes. **Do not** treat the distribution of outcomes here as the expected distribution of your classifications.

9.1 Output comparator → Assimilation → Consistent

Default — houseplant app frequencies. The pre-filled default watering frequency varies across cohorts; chosen frequency rises with the default.

- **Comparator:** Default watering frequency · **Finding:** Higher default → higher chosen frequency · **Type:** Output · **CP Effect:** Assimilation · **Conclusion:** Consistent

9.2 Input comparator → Contrast → Consistent

Peer reward — community garden volunteers. Higher fellow-volunteer stipend lowers own weeding effort.

- **Input(s):** Own stipend · **Comparator:** Fellow's stipend · **Finding:** Higher fellow stipend → less effort · **Type:** Input · **CP Effect:** Contrast · **Conclusion:** Consistent

9.3 Zero is a level — endowment 0-vs-1, GATE ≠ CONCLUSION

Two hobby-club studies share the *same* in-scope design — members are randomly endowed with a tool (level 1) or not (level 0), then choose to keep or trade — but reach **different conclusions**. Being in scope does not fix the answer.

- **9.3a (Consistent).** Endowed members keep at a far higher rate than non-endowed members swap in.
  - **Comparator:** Endowment (0 vs. 1 tool) · **Finding:** Holding the tool raises keep rate · **Type:** Output (outcome) · **CP Effect:** Assimilation · **Conclusion:** Consistent
- **9.3b (Null).** A near-identical study that controls for unfamiliarity with the trading procedure finds **no** keep-vs-swap asymmetry.
  - **Comparator:** Endowment (0 vs. 1 tool) · **Finding:** Endowment has no significant effect on keep rate once procedure is controlled · **Type:** Output (outcome) · **CP Effect:** Null · **Conclusion:** Null

Why both are in scope. The endowment level moves between 0 and 1 — a level variation (§7.3), not "mere existence." The data then decide Consistent vs. Null. Never N/A a 0-vs-1 endowment design for "no magnitude variation."

9.4 Probability is a level → Consistent

Seed-raffle ownership odds. Gardeners are told their probability of winning a rare seed packet (10% vs. 90%); higher win probability raises stated valuation of the packet.

- **Comparator:** Probability of ownership (0.1 vs. 0.9) · **Finding:** Higher ownership probability → higher valuation · **Type:** Output (outcome) · **CP Effect:** Assimilation · **Conclusion:** Consistent

9.5 Output comparator → Contrast → Inconsistent

Garden club bed-of-the-month reviewers. A member's own previous awarding decision (output comparator) predicts the next with **negative** autocorrelation.

- **Comparator:** Own previous decision · **Finding:** Current decision moves *away from* the previous · **Type:** Output (decision) · **CP Effect:** Contrast · **Conclusion: Inconsistent**

9.6 Input comparator → Assimilation → Inconsistent

Whittling-club tool budgets. Hearing a *higher* peer per-tool spend (input comparator) *raises* own carving hours.

- **Input(s):** Own hours · **Comparator:** Peer spend · **Finding:** Higher peer spend → more own hours · **Type:** Input · **CP Effect:** Assimilation · **Conclusion: Inconsistent**

9.7 Model-identified reference — in scope vs. N/A

- **9.7a (in scope, §7.2b).** A study models a baker's "expected loaf quality" reference as a weighted average of her *observed* last five bakes, estimating the decay weight; satisfaction is then measured against it. The reference is anchored to observed past bakes → **in scope**; classify direction from data.
- **9.7b (N/A, §7.2b/5).** A study fits a free reference-quality parameter chosen solely to best-explain the same satisfaction ratings, with no independent observable anchor → **N/A** (reference has no observable anchor).

9.8 Pure null → Null

A study tests whether a fellow hobbyist's stipend (input comparator) affects volunteer effort and finds nothing on its primary margin.

- **Comparator:** Fellow's stipend · **Finding:** No significant effect on effort · **Type:** Input · **CP Effect:** Null · **Conclusion:** Null

9.9 N/A examples

Reason	Hypothetical
No measured comparator	Cyclists' stopping is "consistent with daily mileage targets," but the target is never measured — only inferred, with no observable anchor.
Level identical, only label differs (framing)	A craft fair shifts the *same* payoff matrix into a loss domain via a uniform show-up gift card; every subject's reference (zero) is identical — only the wording differs. *(Contrast with §7.3: nothing's level changes.)*
Only frequency of comparisons varies	Knitters see their project valuation *more vs. fewer times*; each valuation's content is fixed. (MLA-style — N/A.)
Identity, not level	An exchange study varies *which* craft good subjects are endowed with (yarn vs. clay), not how much; the reference identity changes, not its level.
Pure theory	An axiomatic model of stochastic reference-dependent choice, no data.
Choice-set manipulation	A craft store adds a dominated yarn labeled "default"; the choice set changes, not a comparator level.
Varying an option's payoff	An experiment varies the perks of an "easy" membership tier — more uptake when it pays more is standard optimization.
Multidimensional distance	Cookbook reissues scored by similarity to the original across many dimensions; the comparator is a high-dimensional location, not a level on one axis.
Indeterminate type	A single dollar-valued comparator equally defensibly the input (a given cost) and the output (a price), with no design feature breaking the tie.

9.10 Authors' mechanism label does not gate scope

Suggested-contribution jar at a hobby fair. A study's headline contribution concerns *norm compliance among club members*, and the authors interpret their results throughout as fairness/norm preferences. One treatment varies the posted suggested contribution at the entry jar (2 vs. 8 tokens); members' own contributions are measured.

- **Comparator:** Posted suggested contribution (2 vs. 8) · **Type:** Output (decision) · **Finding & Conclusion:** read from the data — if own contributions rise in the posted level → Assimilation → Consistent; if they fall → Contrast → Inconsistent; if unaffected → Null.

Why in scope: the design contains a measured comparator whose level varies and the paper reports the directional response of the DM's own decision. The authors' interpretation (norms/fairness) is not a scope criterion (§0.8), and being in scope does not fix the conclusion (§0.3).

I Experimental Instructions, Comparison Point Announcements, and Decision Screens

I.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.

I.1.1 Effort Valuation

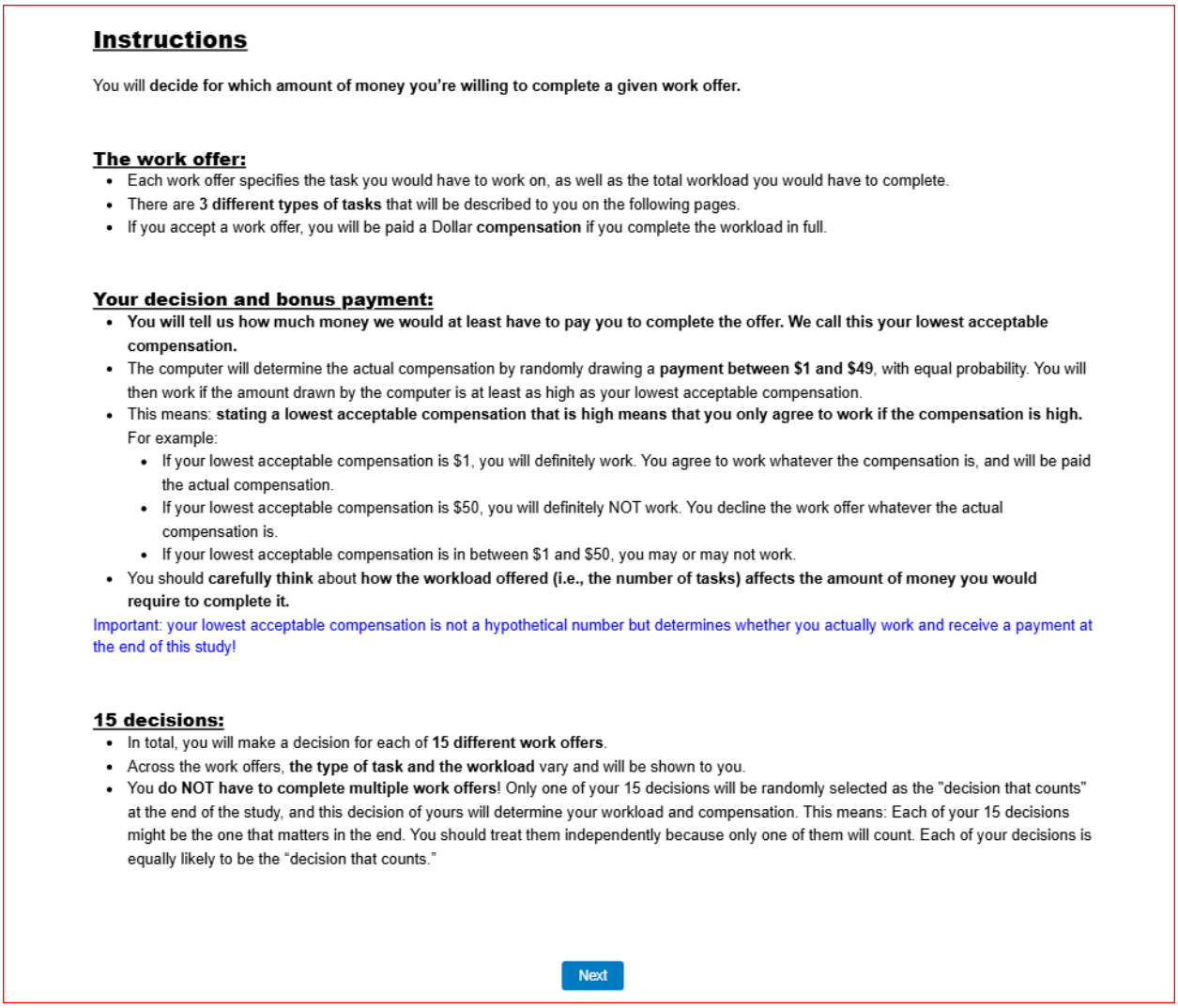

Figure I.1: Instruction screen (2) for the *Effort Valuation* paradigm.

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

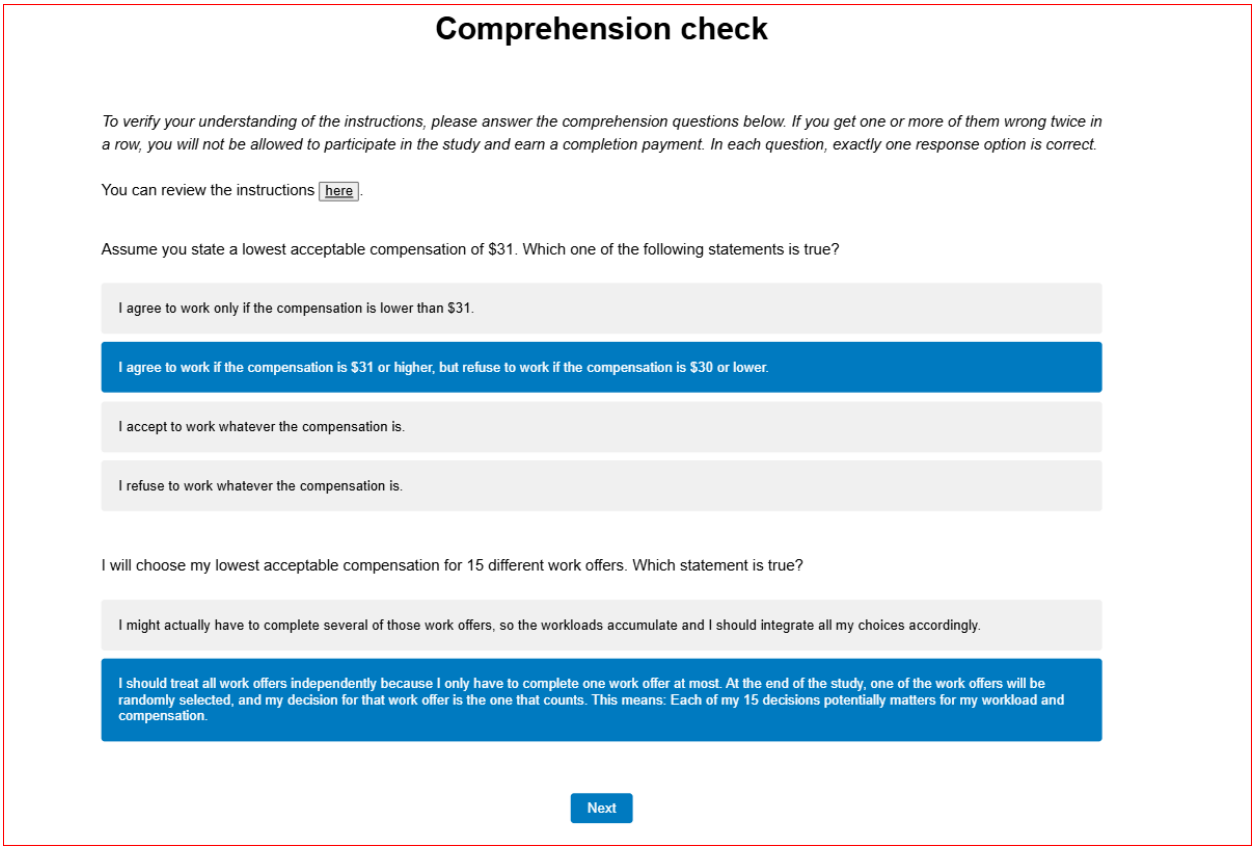

Figure I.2: Comprehension-check screen for the *Effort Valuation* paradigm.

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

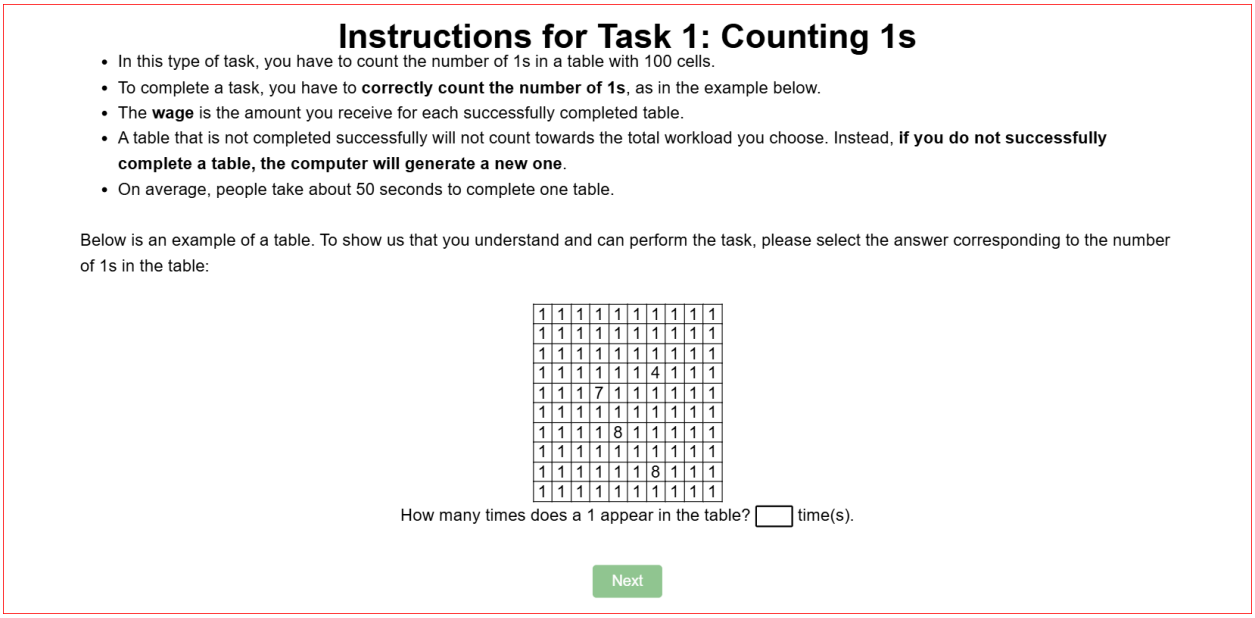

Figure I.3: Counting-task instruction screen for the *Effort Valuation* paradigm.

Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Random*, *Deliberation*); the decision screen that followed varied. This and the following two task-description screens carry over two phrases from the task instructions of the *Effort Choice* paradigm (a per-unit "wage" and "the total workload you choose"). In *Effort Valuation*, the workload was exogenously assigned and compensation was determined by the Becker–DeGroot–Marschak mechanism, as stated on the general instructions (Figure I.1) and on every decision screen.

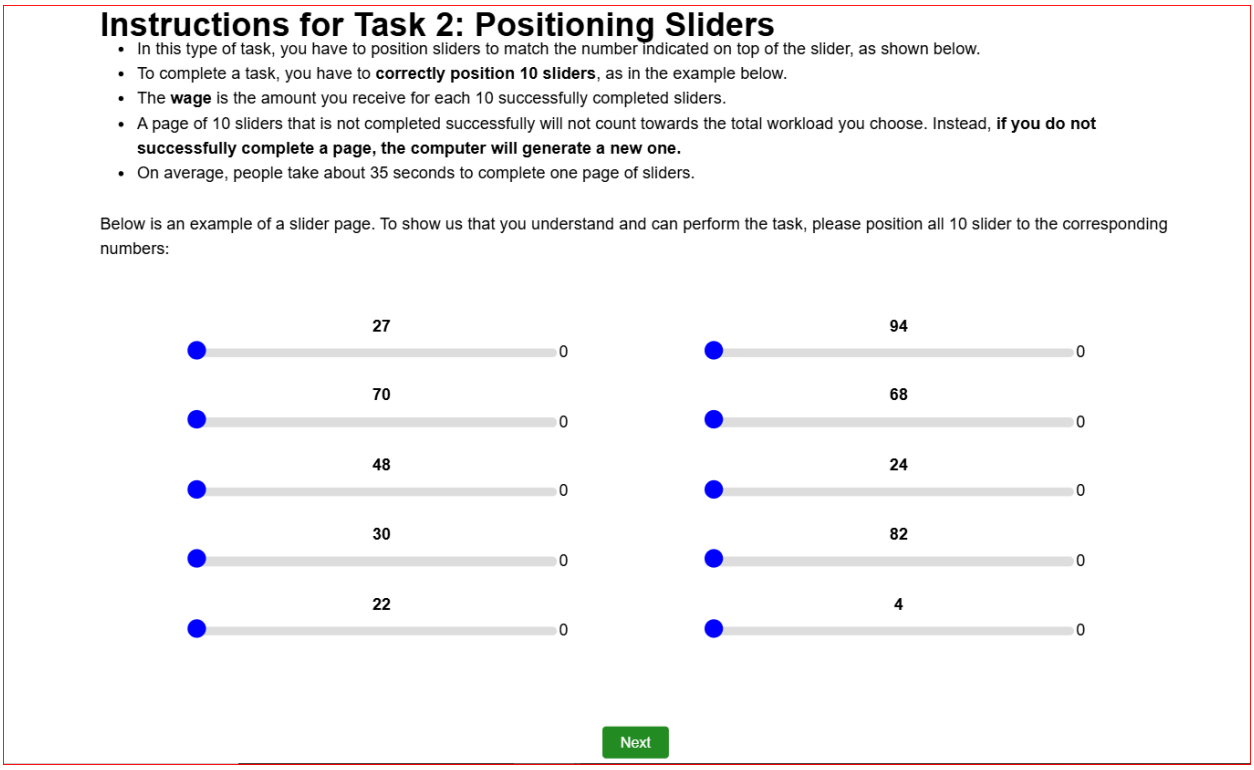

Figure I.4: Slider-task instruction screen for the *Effort Valuation* paradigm.

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

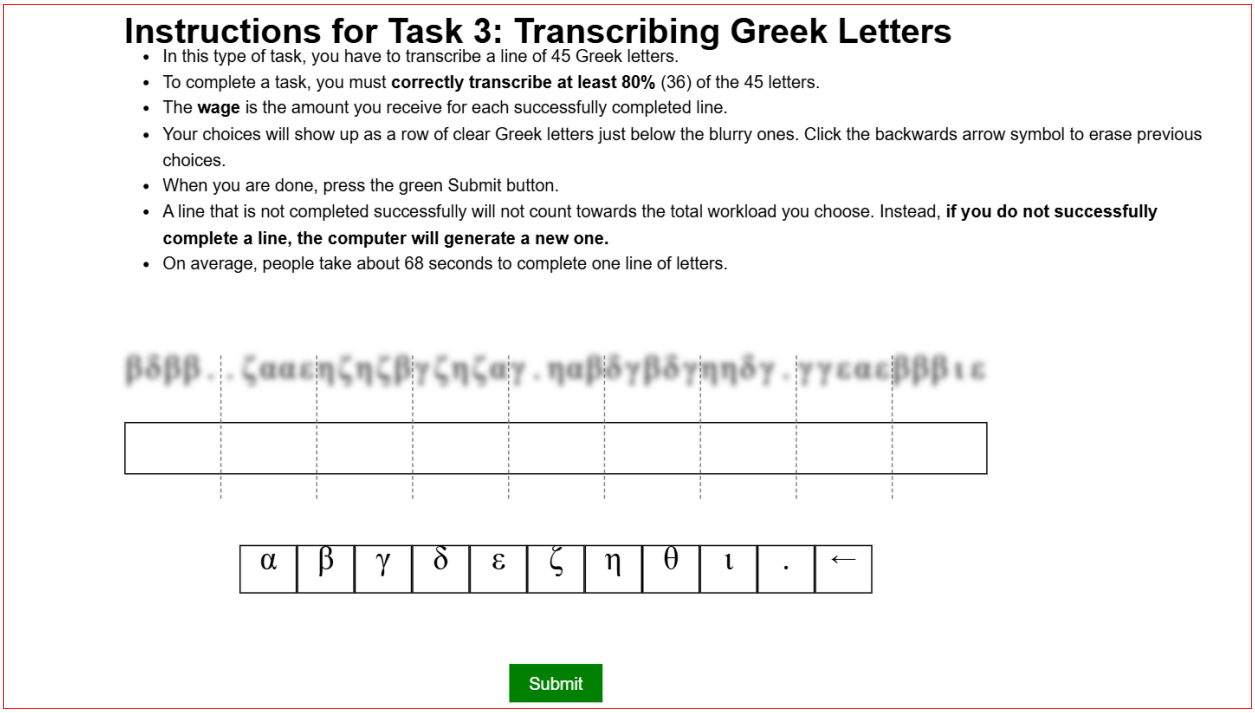

Figure I.5: Greek-letters typing-task instruction screen for the *Effort Valuation* paradigm.
Note: The screen was identical across the different treatment conditions (*Expectations CP*, *Social Both CP*, *Social Input CP*, *Social Output CP*, *Time*, *Random*, *Deliberation*); the decision screen that followed varied.

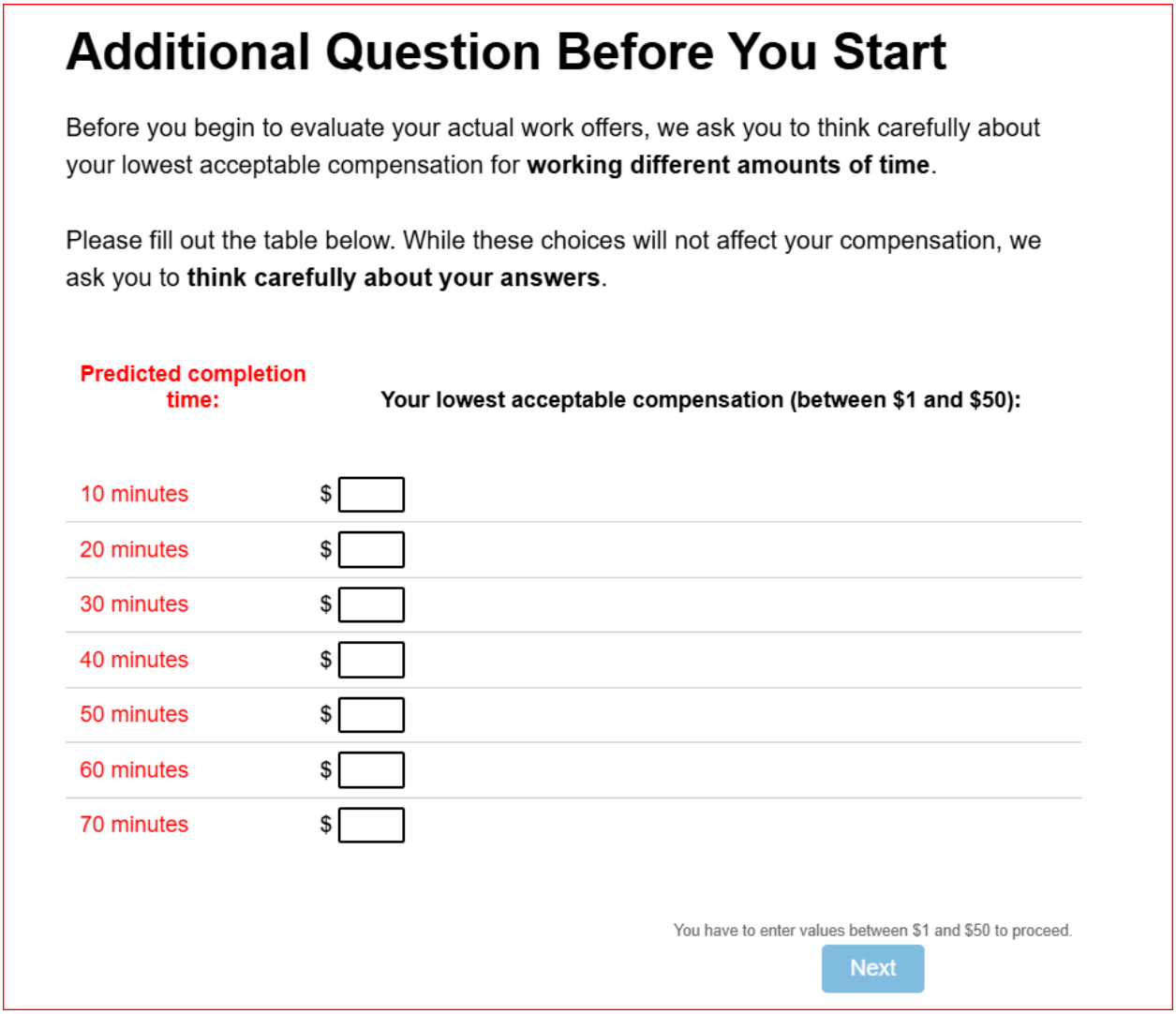

Figure I.6: Additional question shown only in the *Time* treatment.

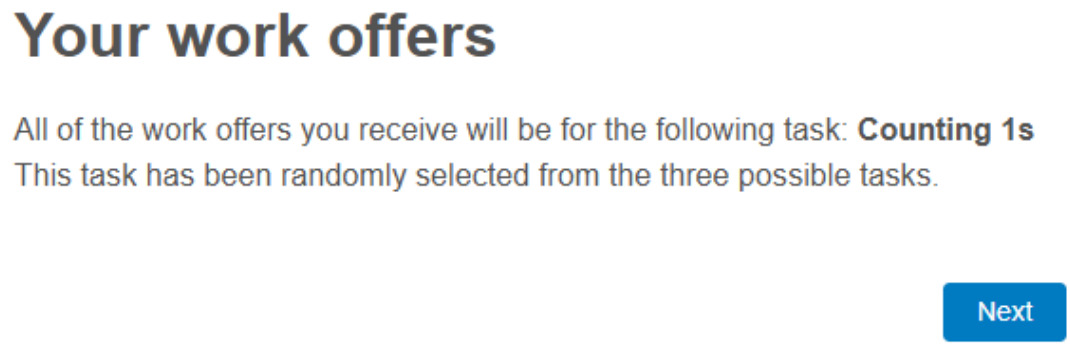

Figure I.7: Counting ones announcement screen for the *Deliberation* treatment.

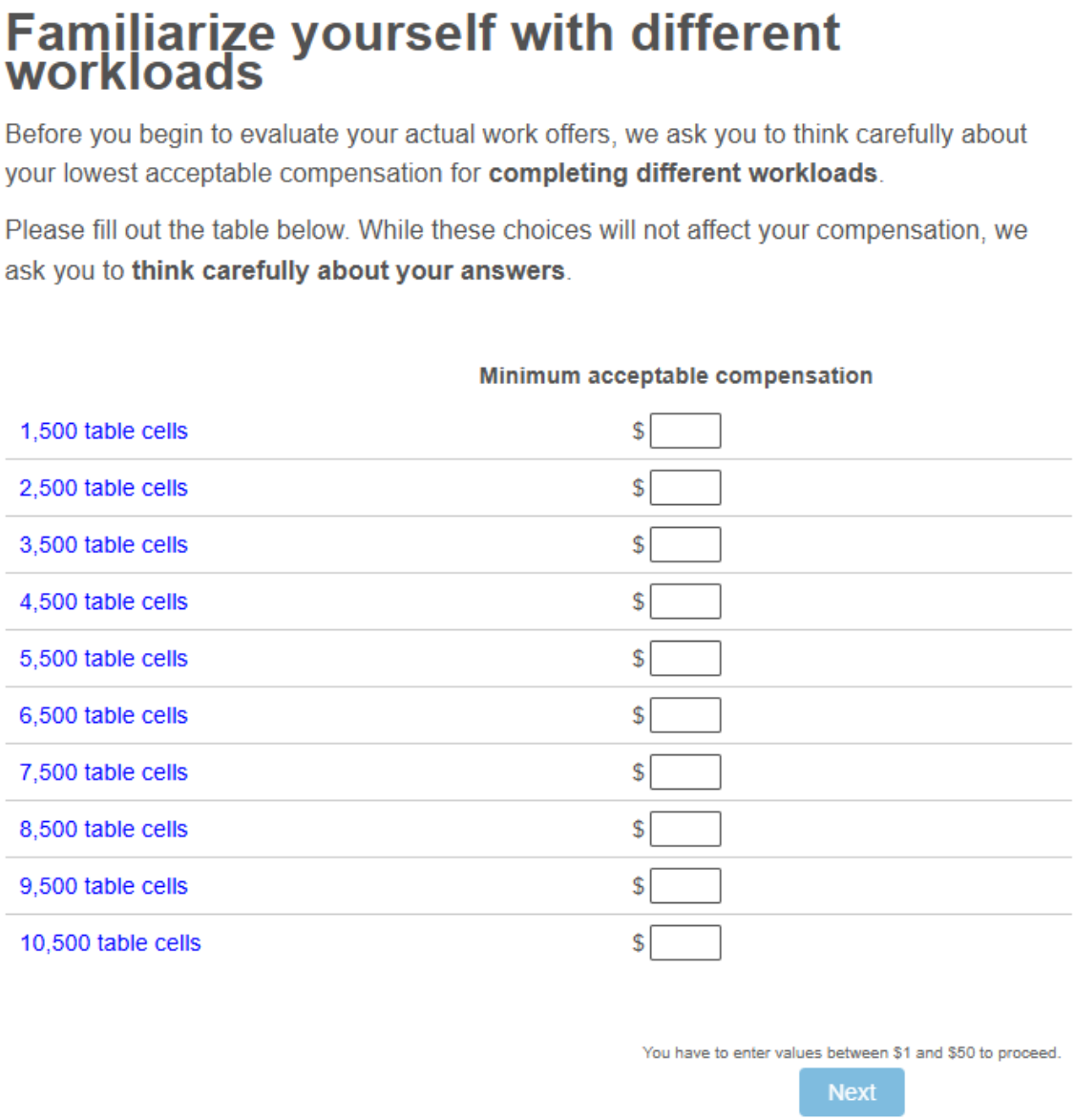

Figure I.8: Counting ones training screen for the *Deliberation* treatment.

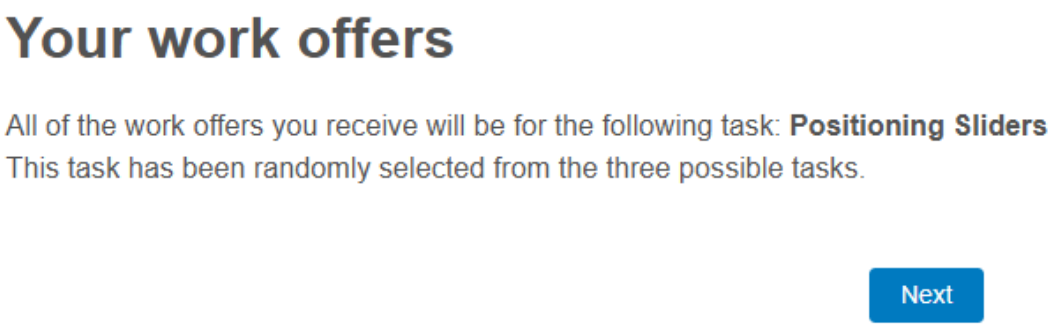

Figure I.9: Positioning sliders announcement screen for the *Deliberation* treatment.

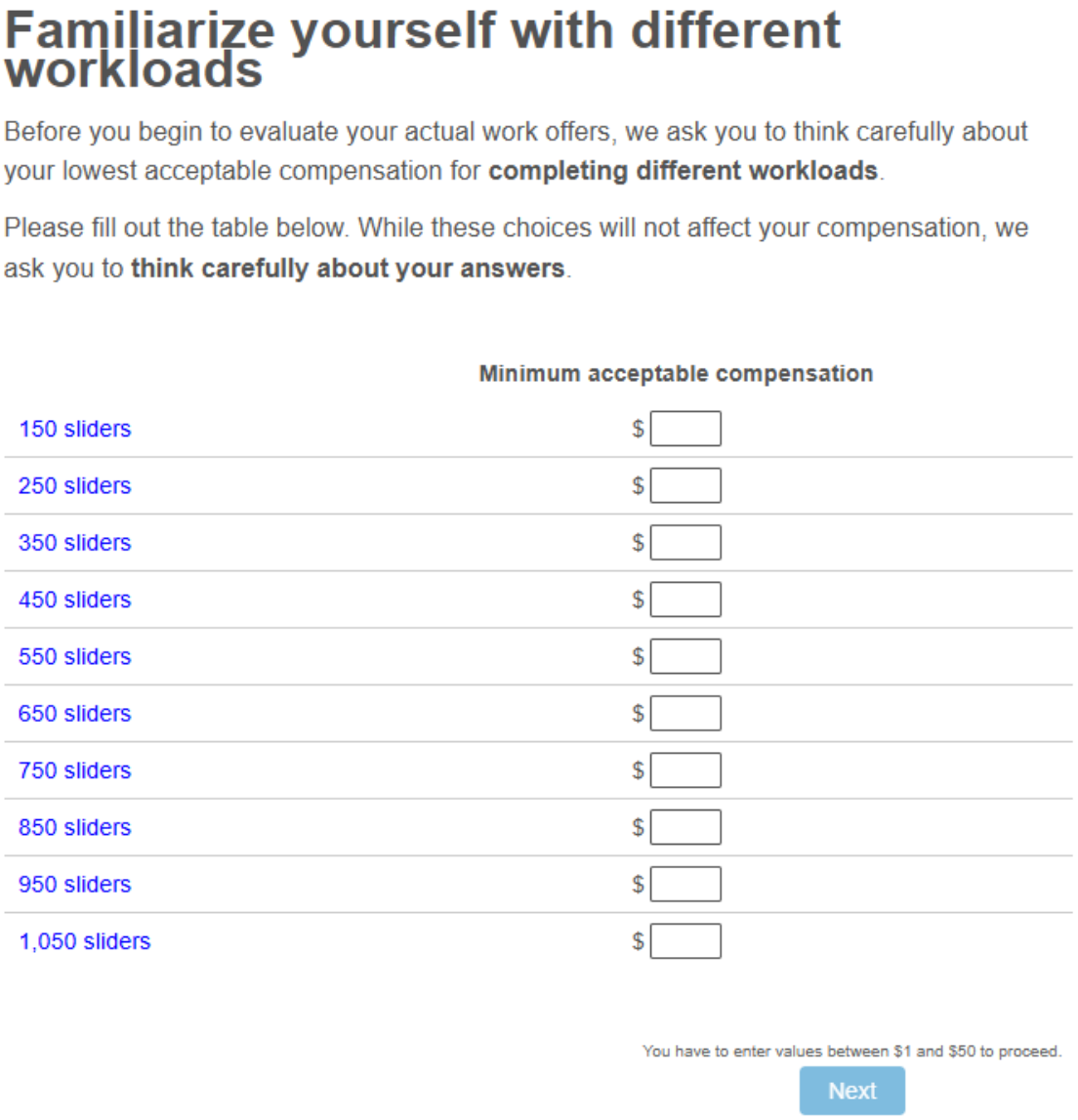

Figure I.10: Positioning sliders training screen for the *Deliberation* treatment.

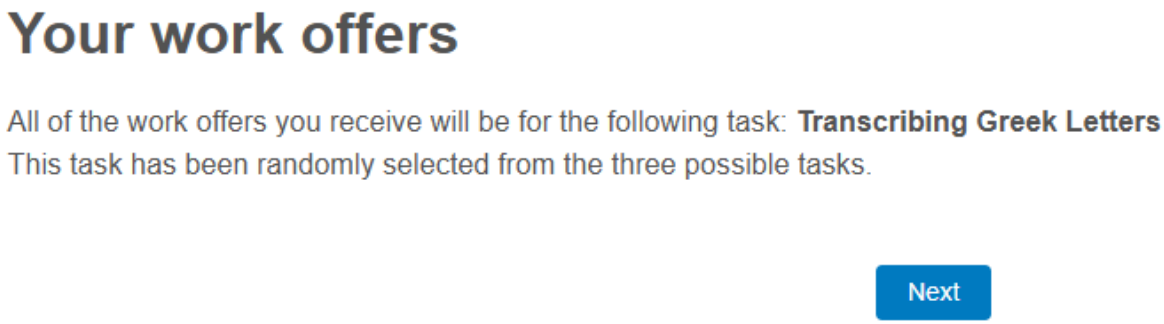

Figure I.11: Greek-letters typing-task training screen for the *Deliberation* treatment.

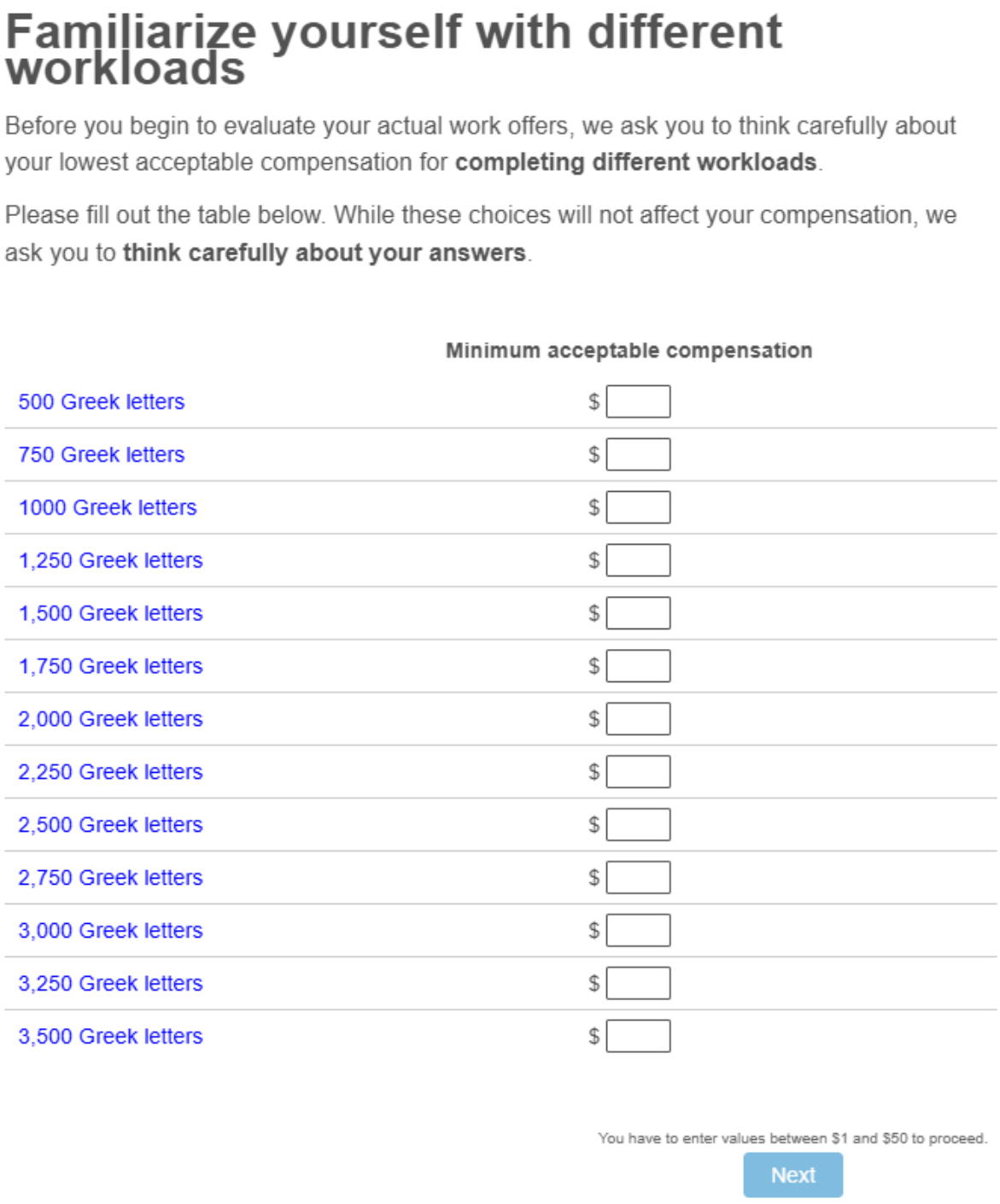

Figure I.12: Greek-letters typing-task training screen for the *Deliberation* treatment.

I.1.2 Effort Choice

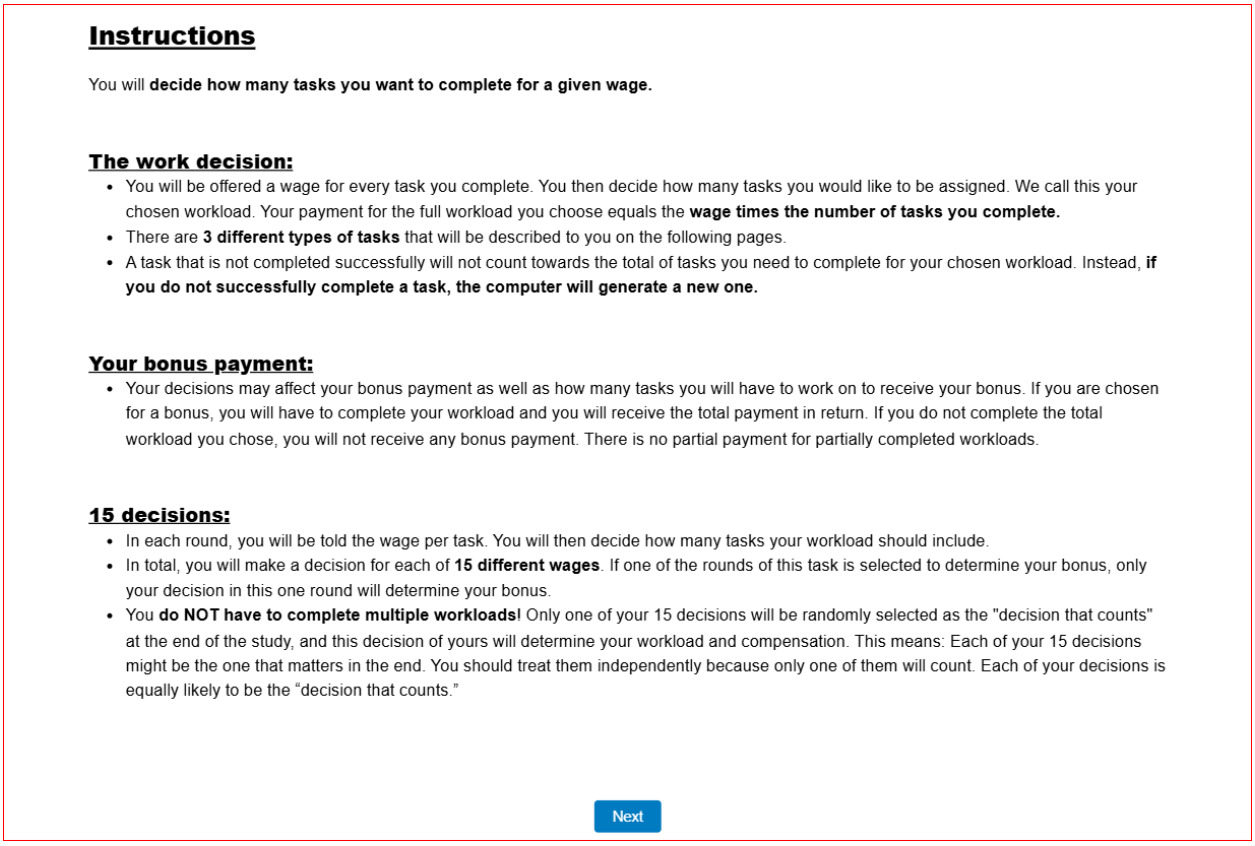

Figure I.13: Instruction screen for the *Effort Choice* paradigm.

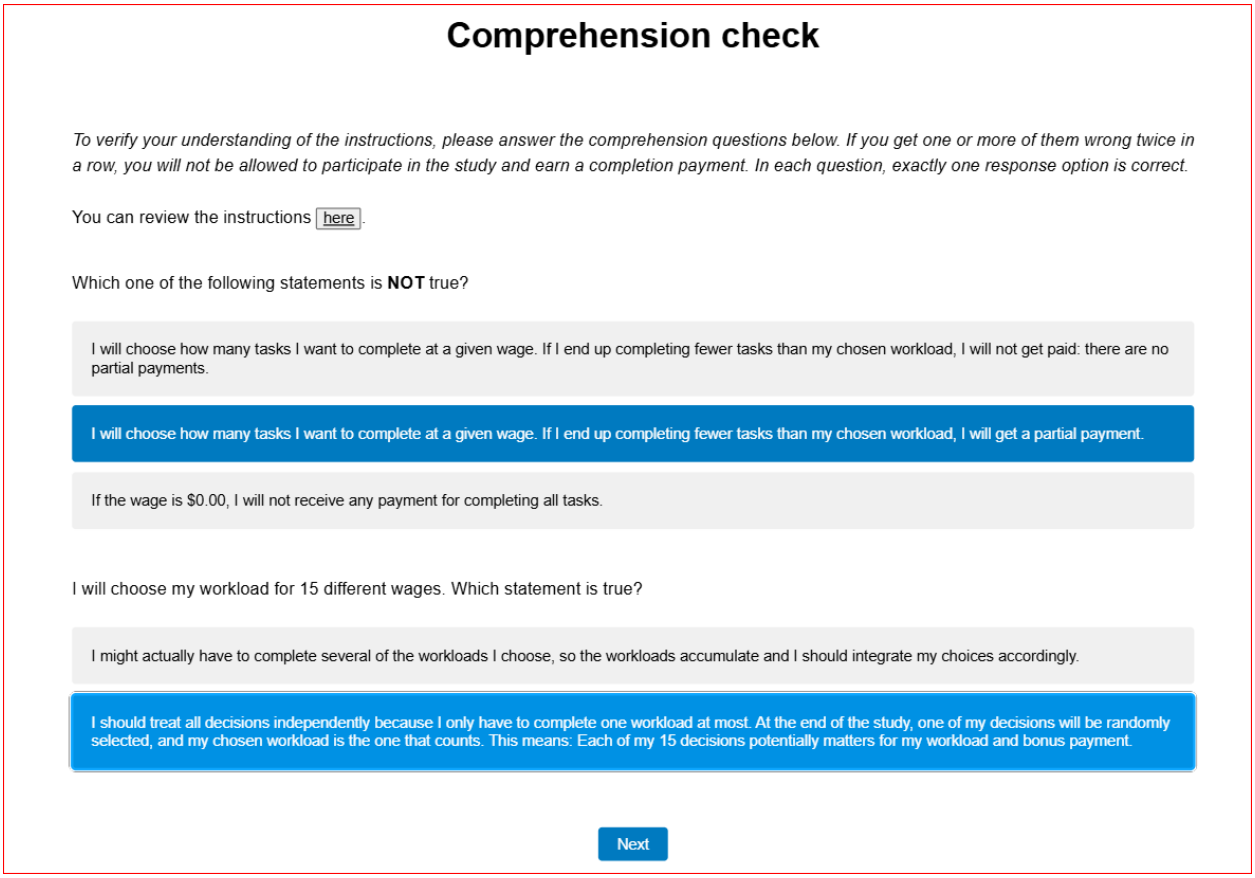

Figure I.14: Comprehension-check screen for the *Effort Choice* paradigm.

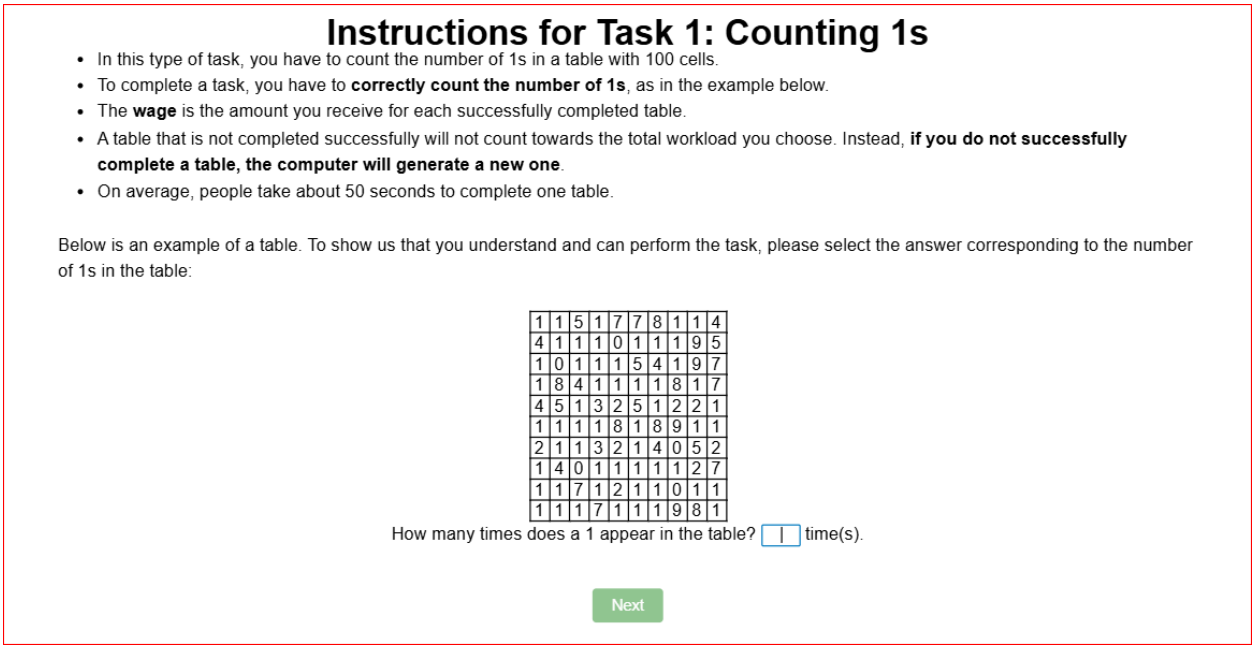

Figure I.15: Counting-task instruction screen for the *Effort Choice* paradigm.

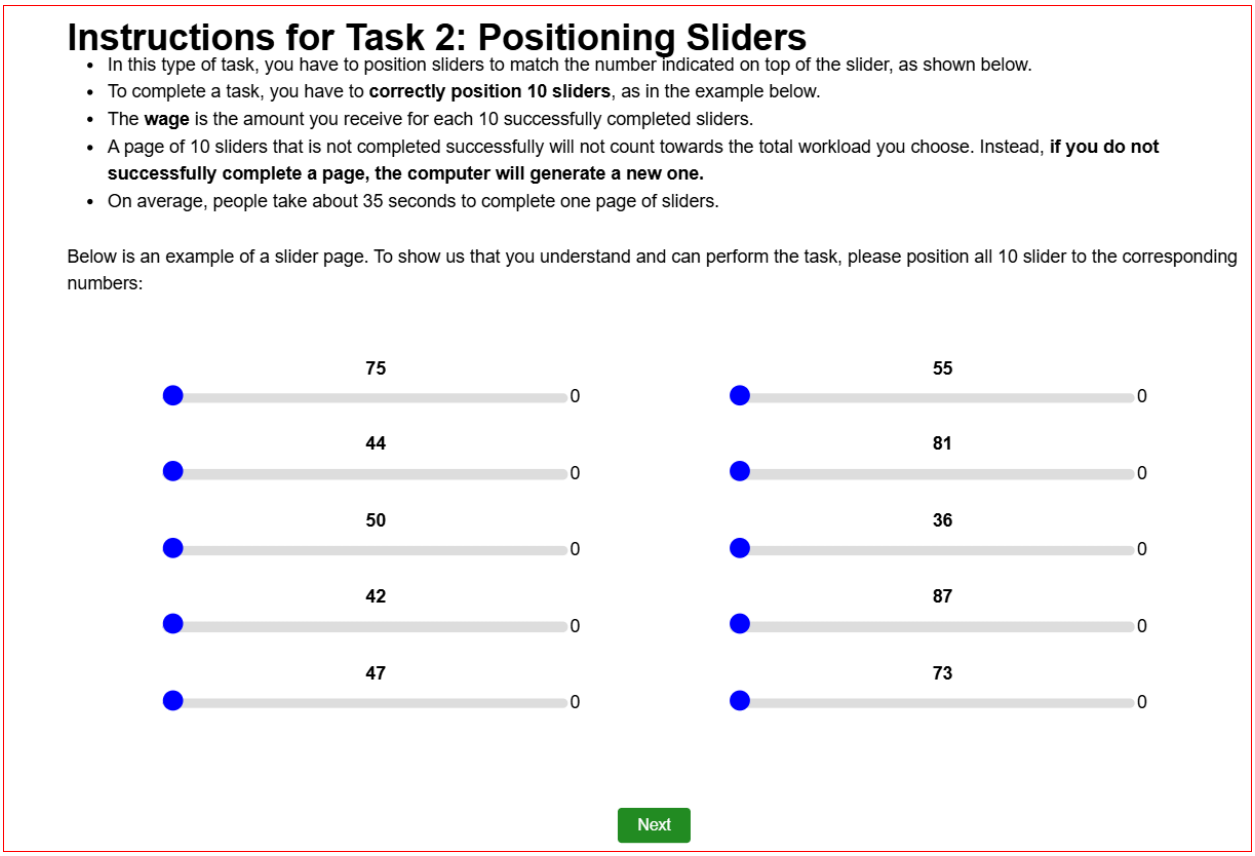

Figure I.16: Slider-task instruction screen for the *Effort Choice* paradigm.

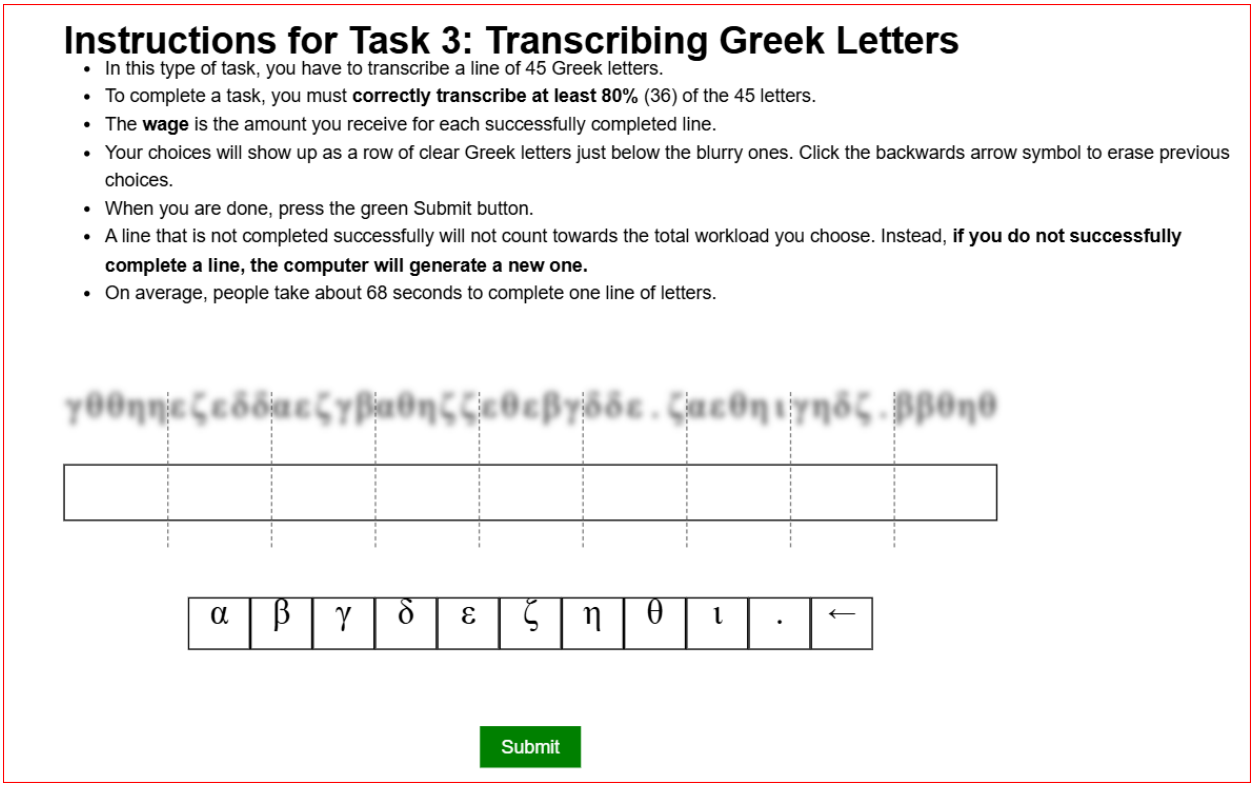

Figure I.17: Greek-letters typing-task instruction screen for the *Effort Choice* paradigm.

I.1.3 Investment

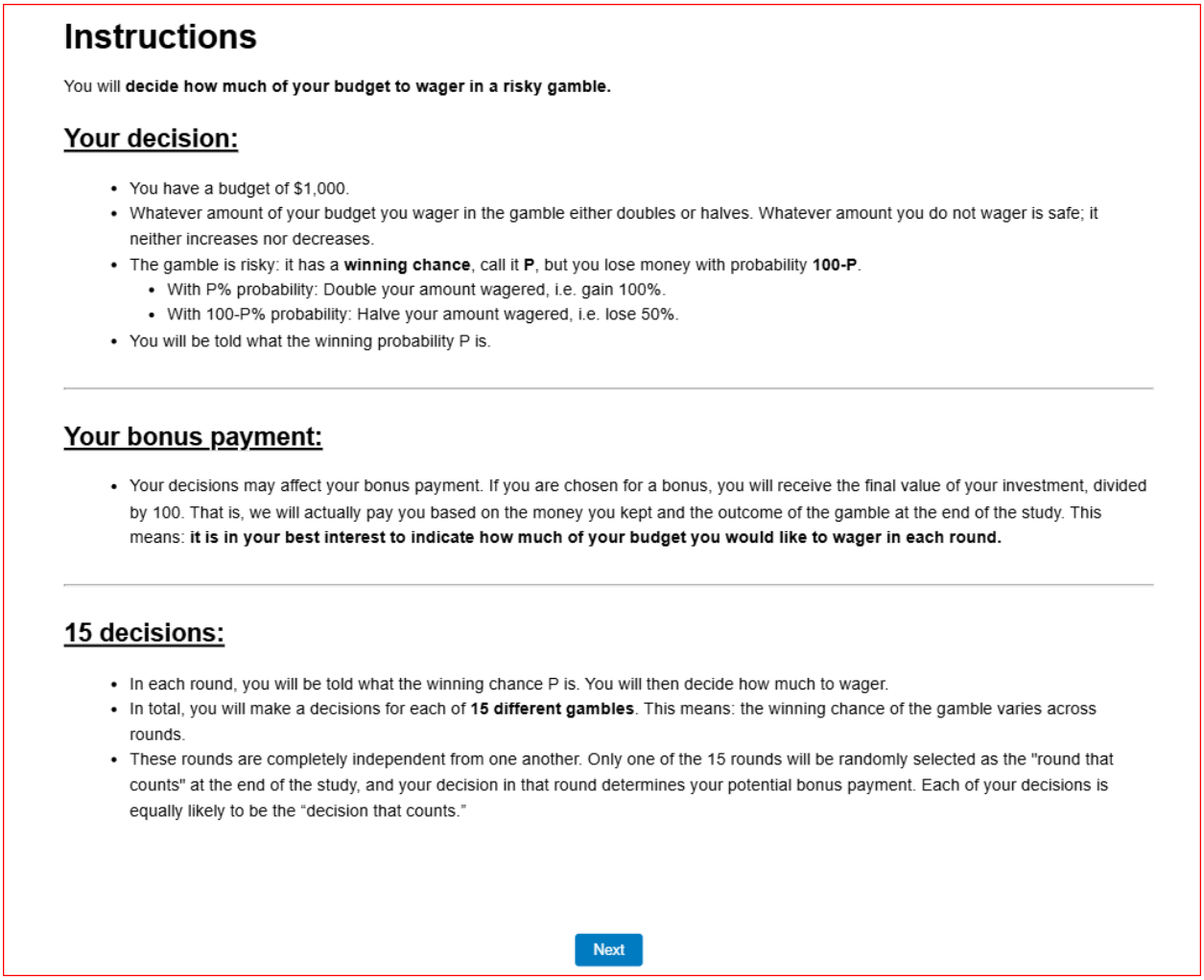

Figure I.18: Instruction screen for the *Investment* paradigm.

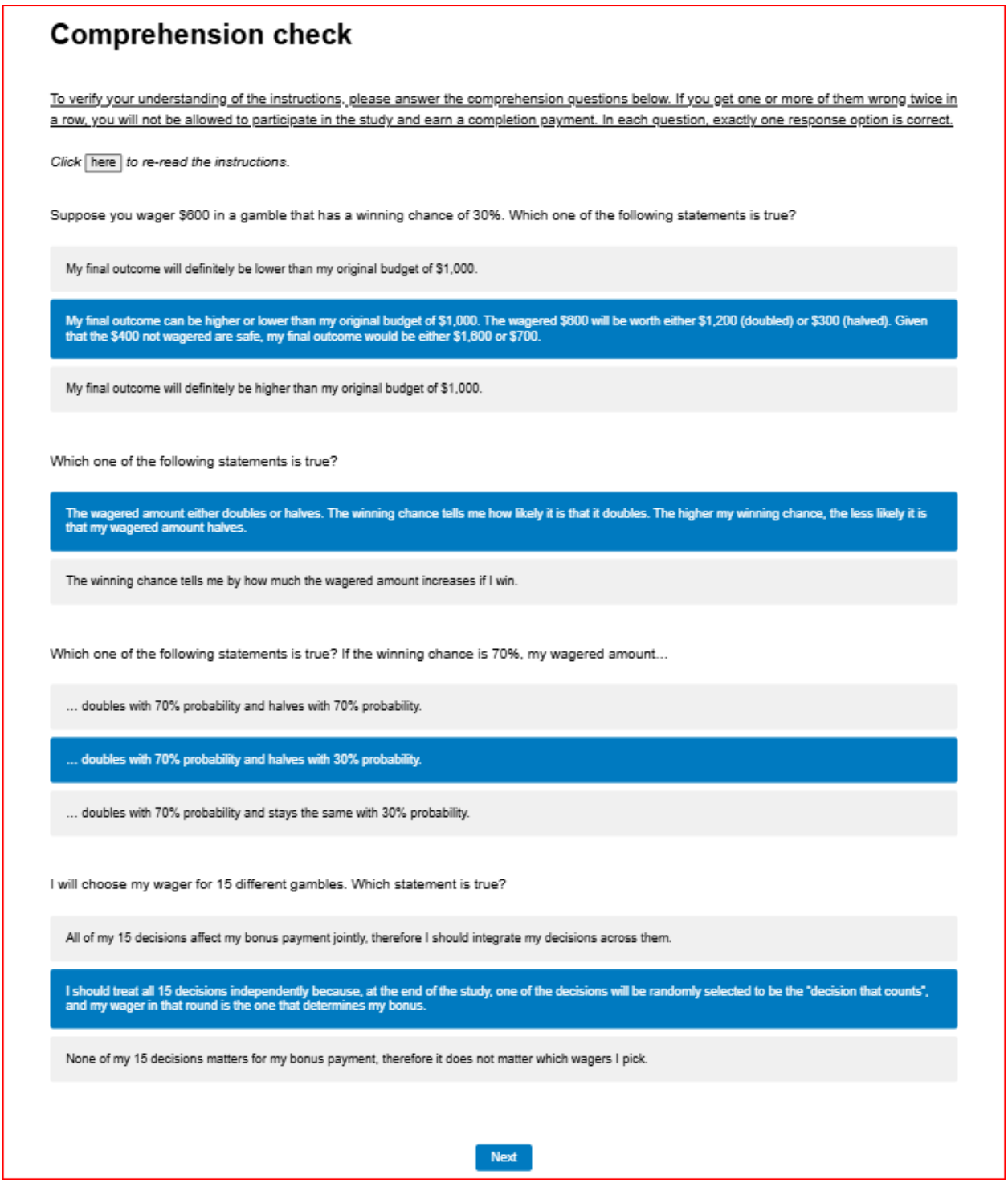

Figure I.19: Comprehension-check screen for the *Investment* paradigm.

I.1.4 Beliefs

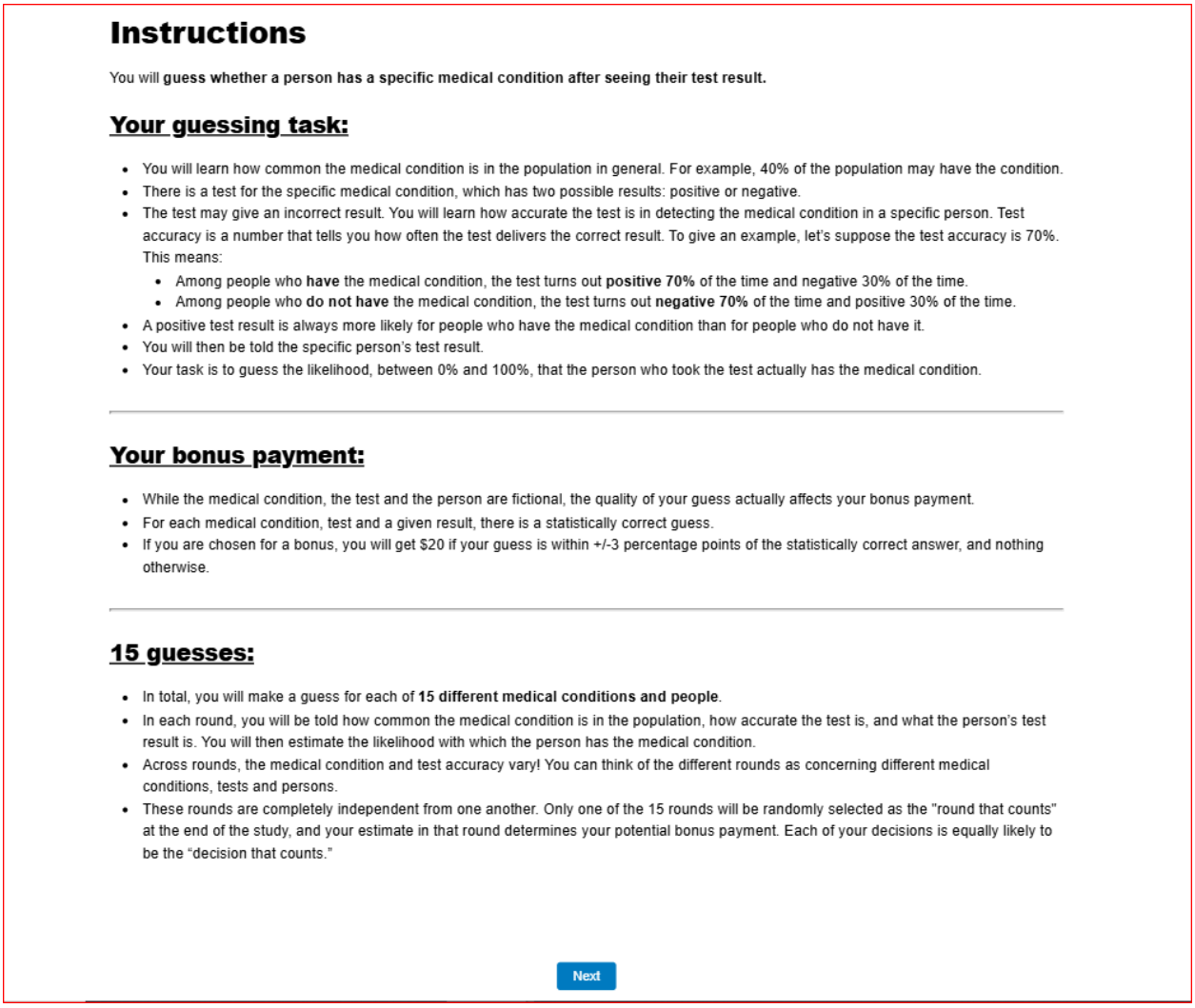

Figure I.20: Instruction screen for the *Beliefs* paradigm.

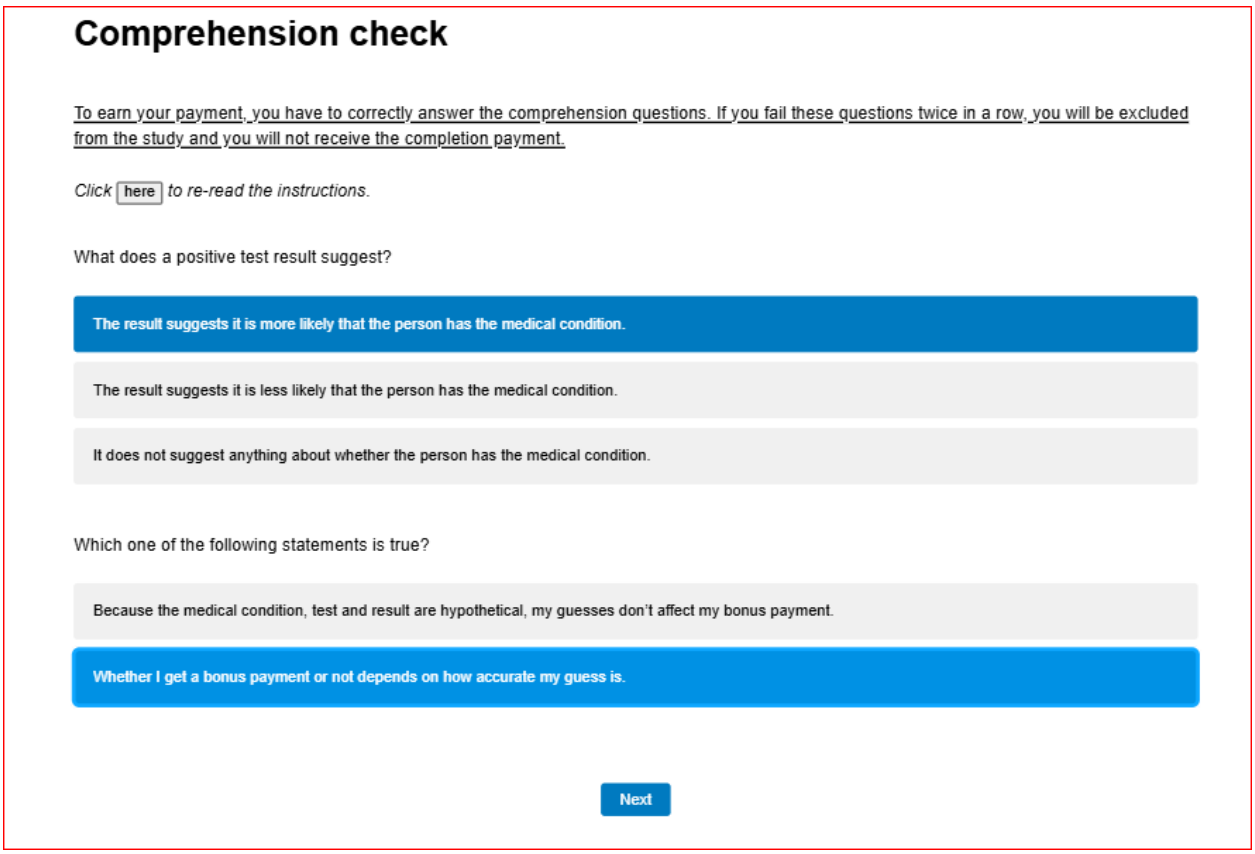

Figure I.21: Comprehension-check screen for the *Beliefs* paradigm.

I.2 Comparison Point Announcement and Decision Screens

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

I.2.1 Effort Valuation

Treatment: Expectations CP

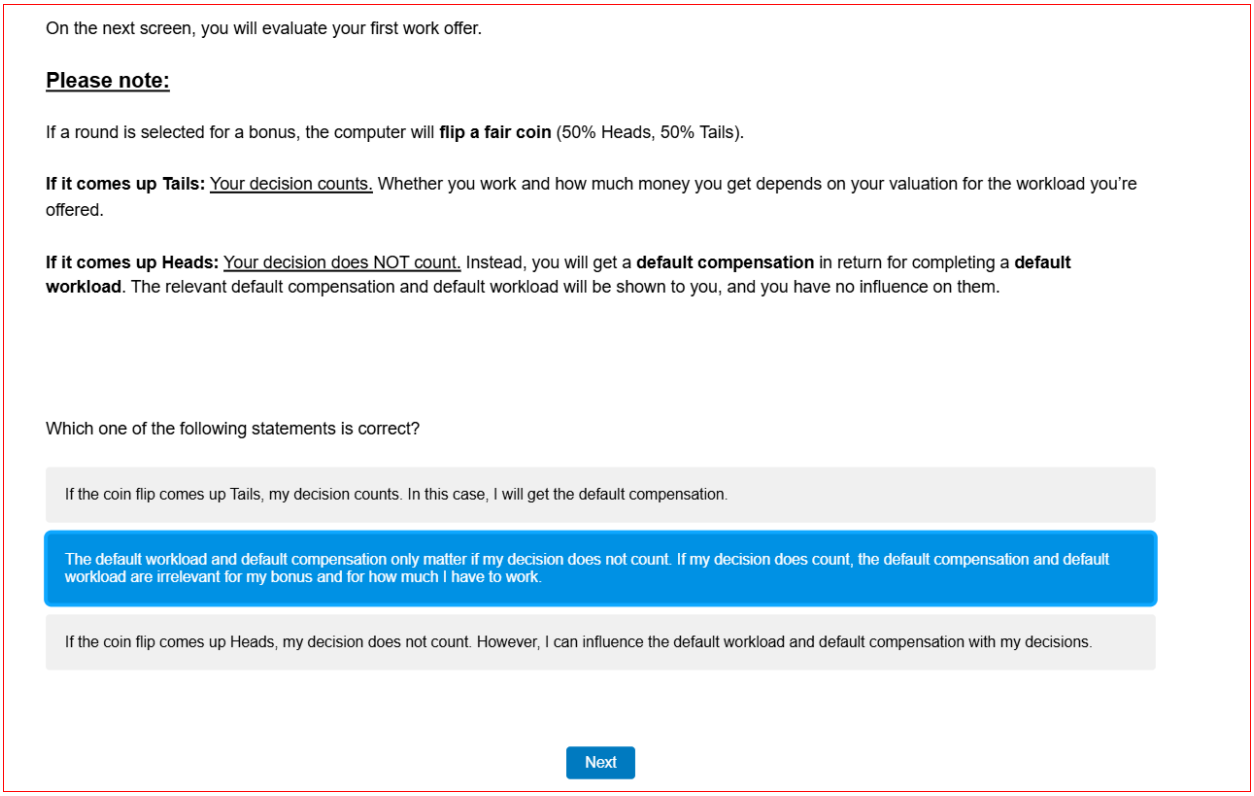

Figure I.22: Treatment announcement — *Expectations CP* (*Effort Valuation*).

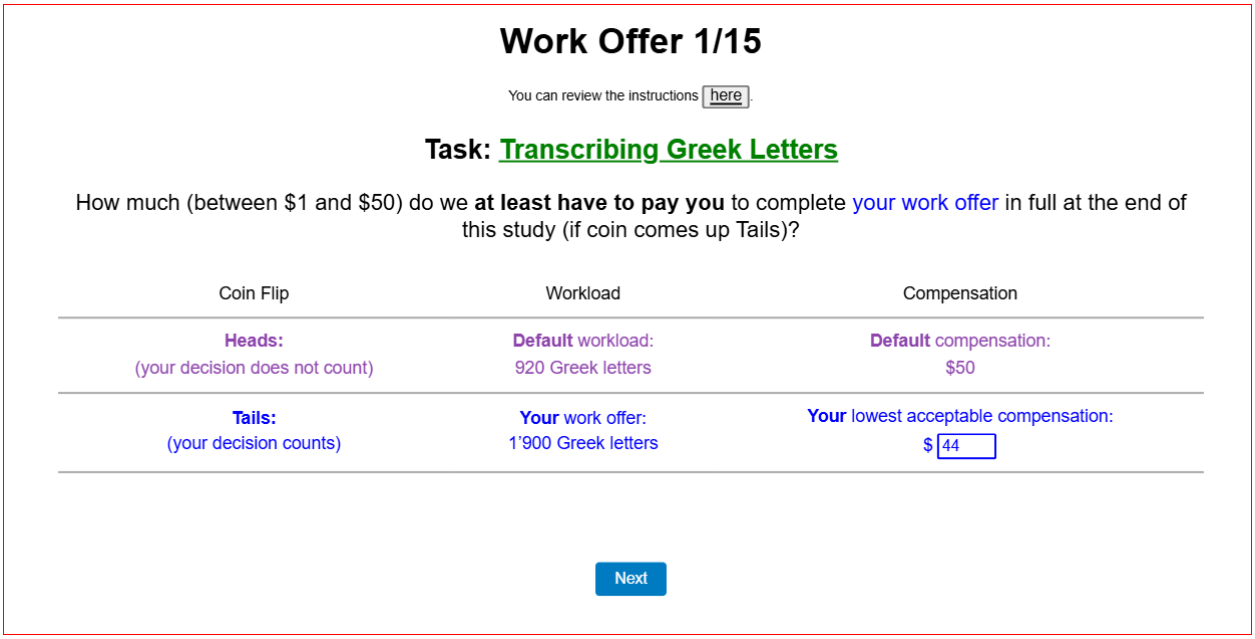

Figure I.23: Decision screen — *Expectations CP* (*Effort Valuation*).

Treatment: Social Both CP

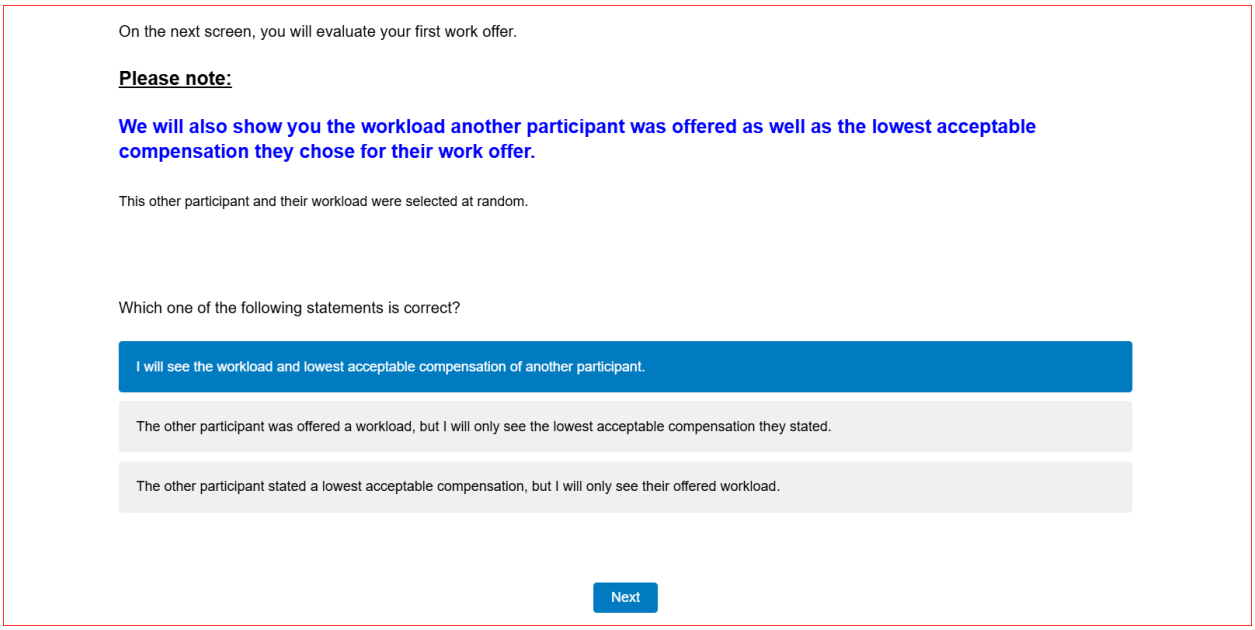

Figure I.24: Treatment announcement — *Social Both CP*.

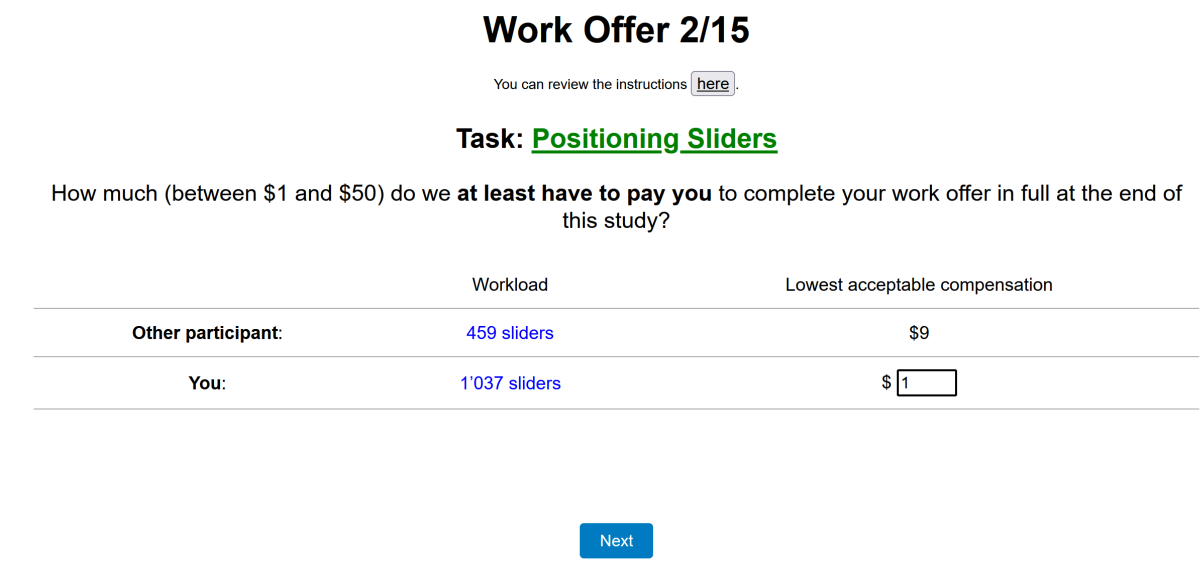

Figure I.25: Decision screen — *Social Both CP* (*Effort Valuation*).

Treatment: Social Input CP

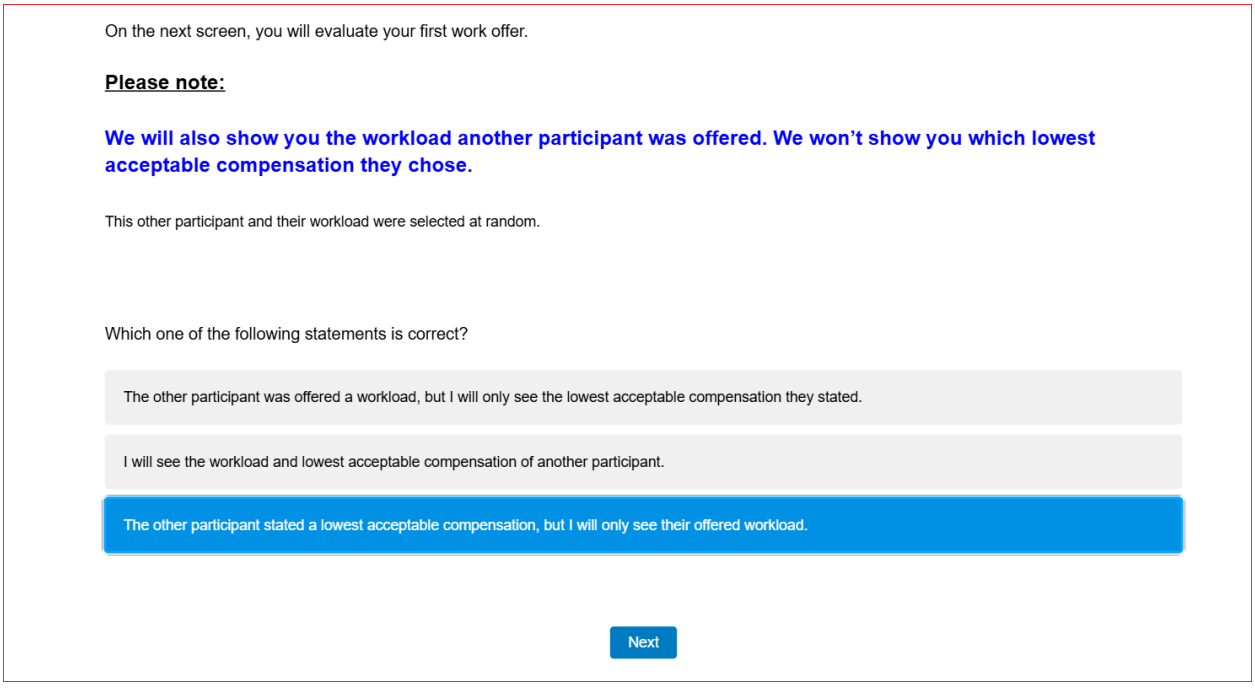

Figure I.26: Treatment announcement — *Social Input CP*.

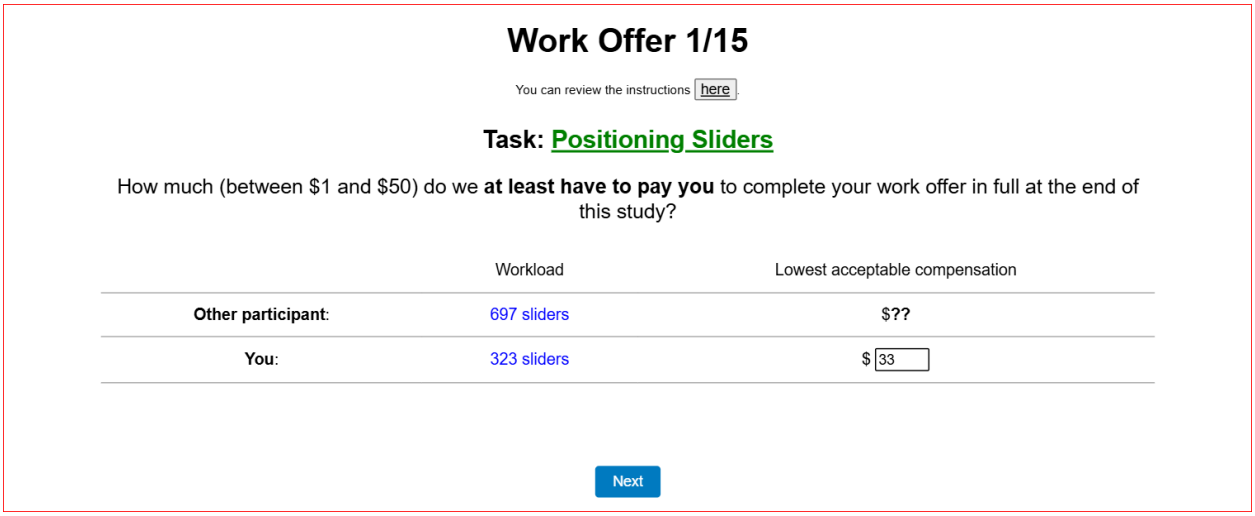

Figure I.27: Decision screen — *Social Input CP* (*Effort Valuation*).

Treatment: Social Output CP

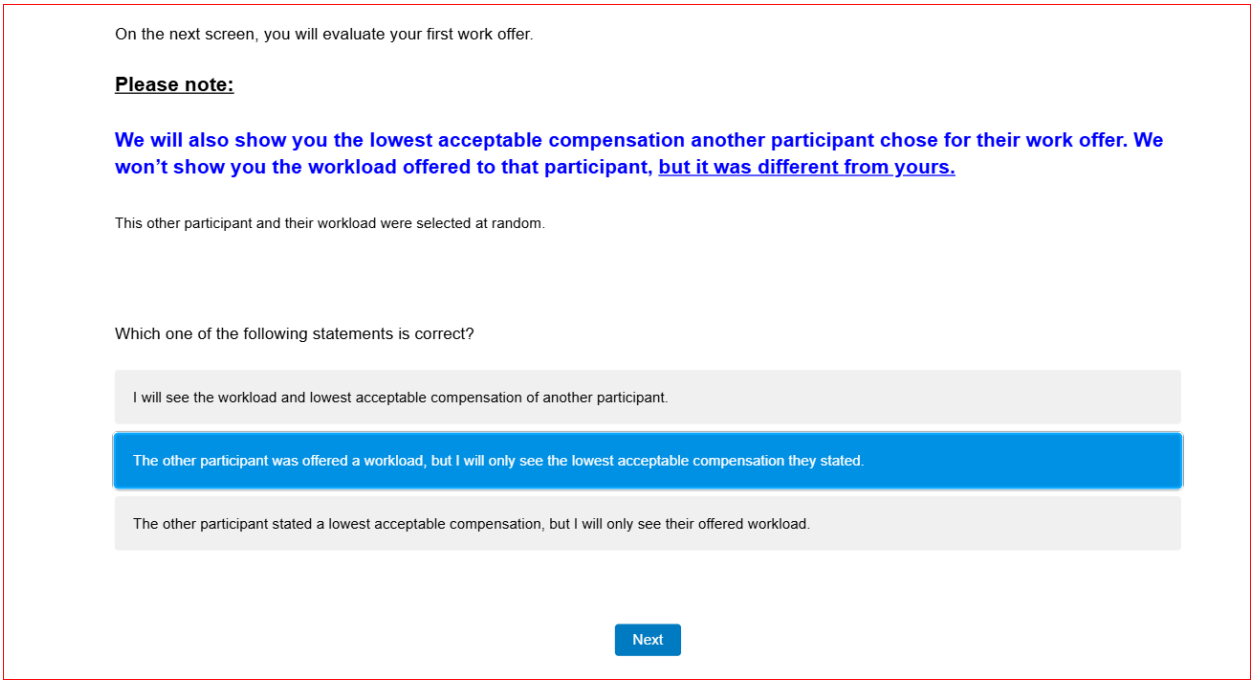

Figure I.28: Treatment announcement — *Social Output CP*.

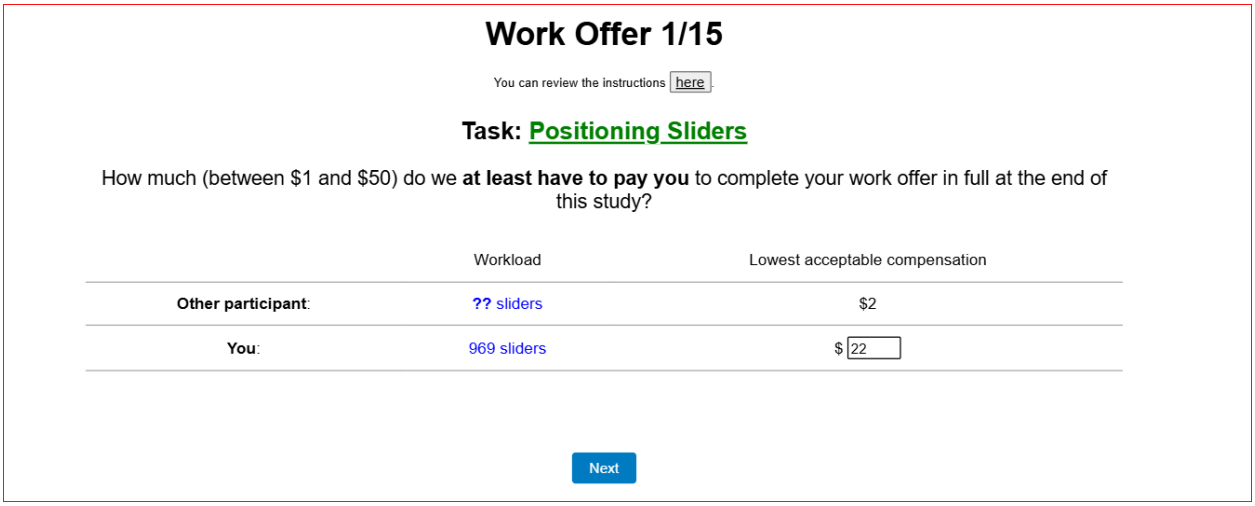

Figure I.29: Decision screen — *Social Output CP* (*Effort Valuation*).

Treatment: Time Treatment

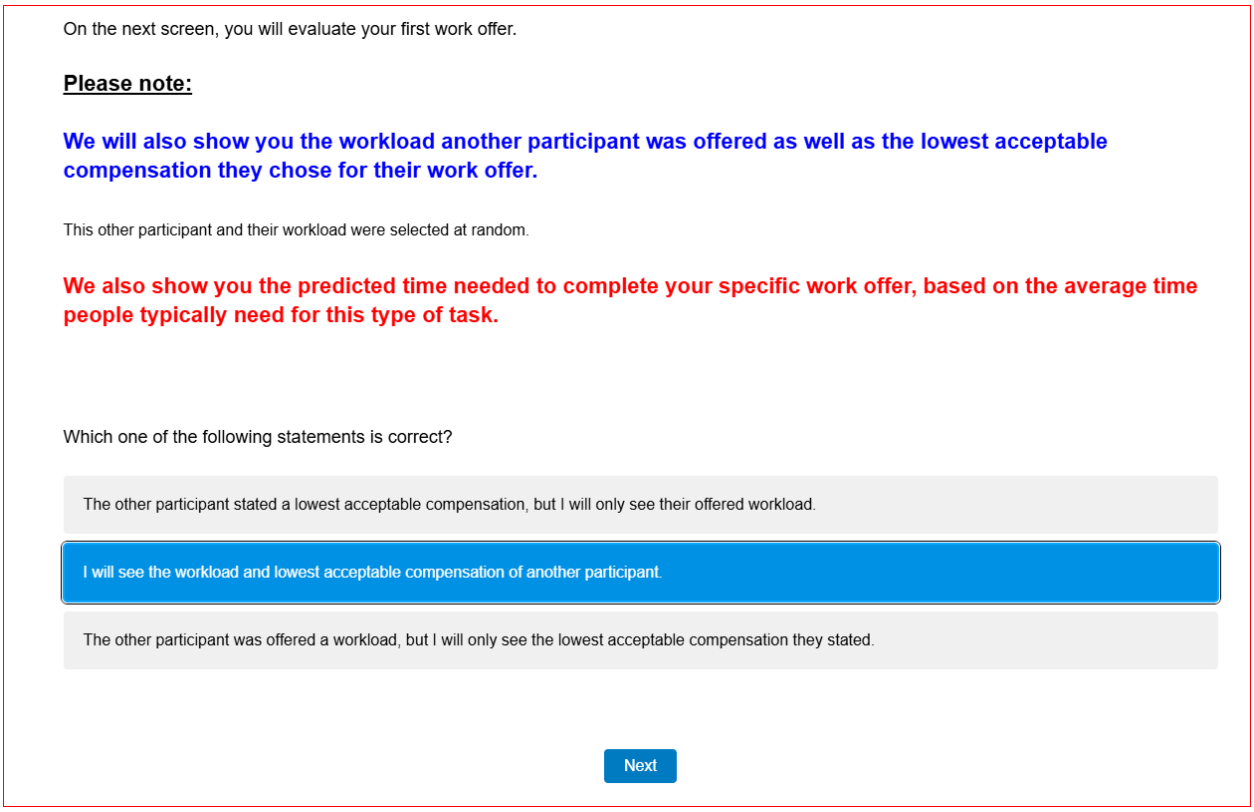

Figure I.30: Treatment announcement — *Time* treatment.

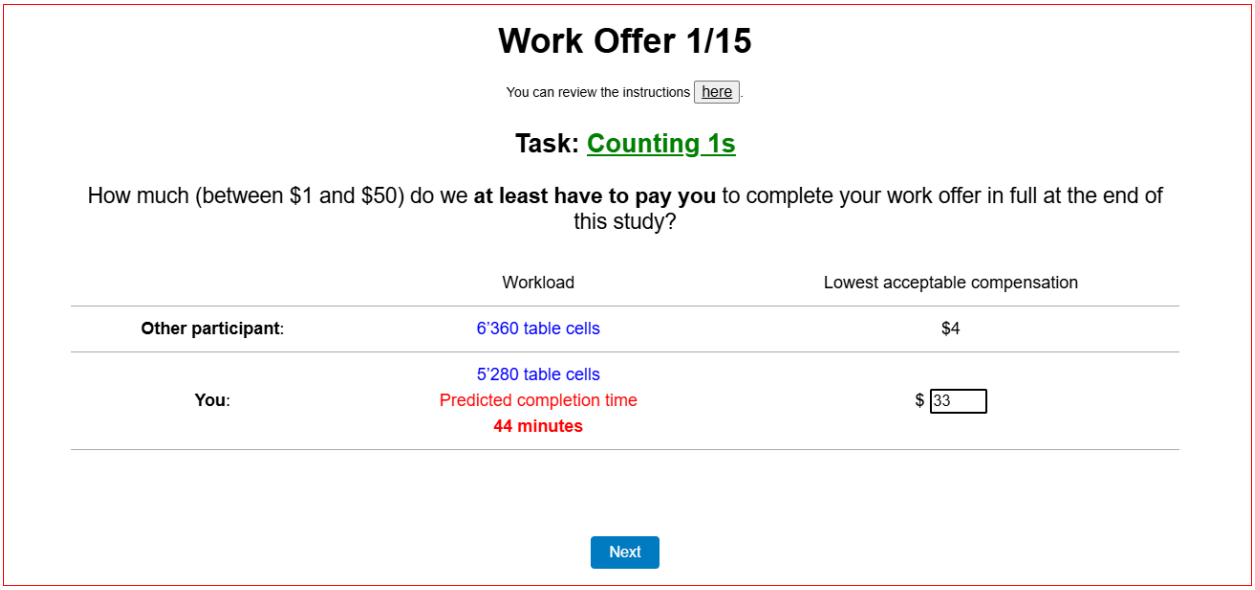

Figure I.31: Decision screen — *Time* treatment (*Effort Valuation*).

Treatment: Random

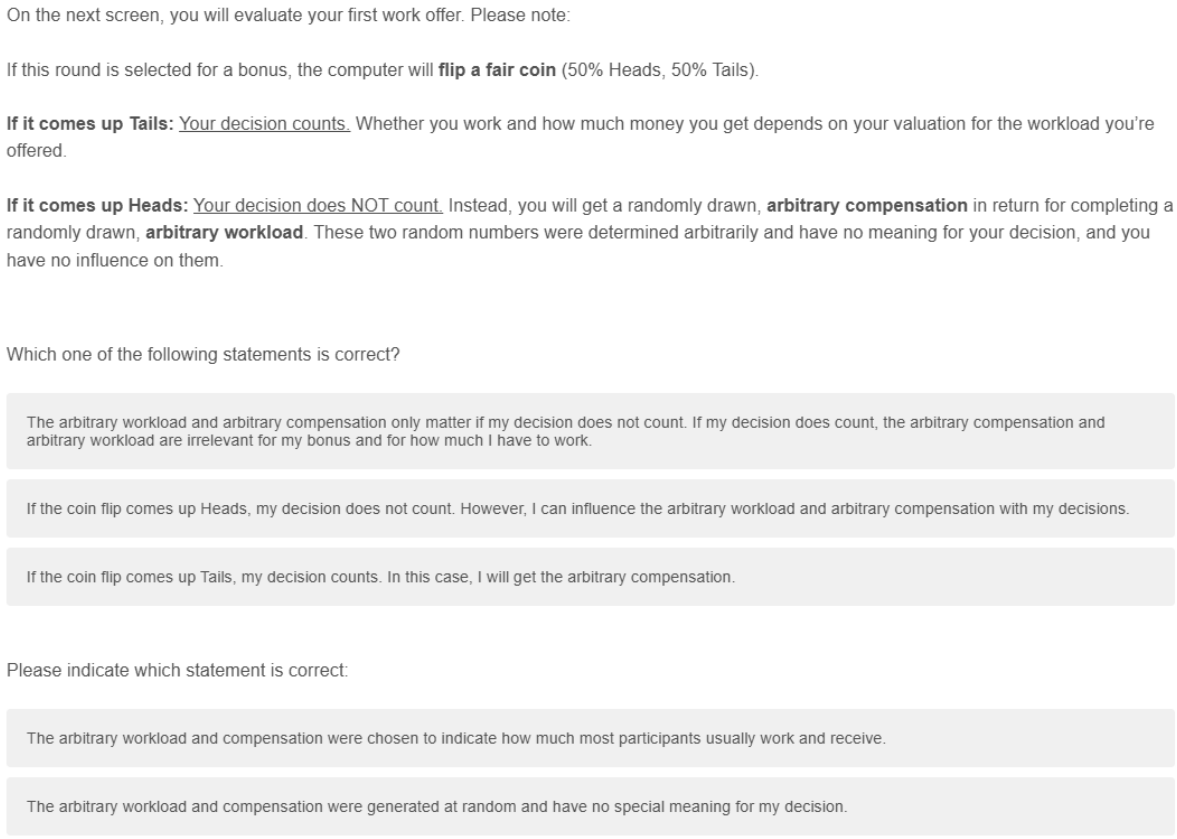

Figure I.32: Treatment announcement — *Random* treatment (*Effort Valuation*).

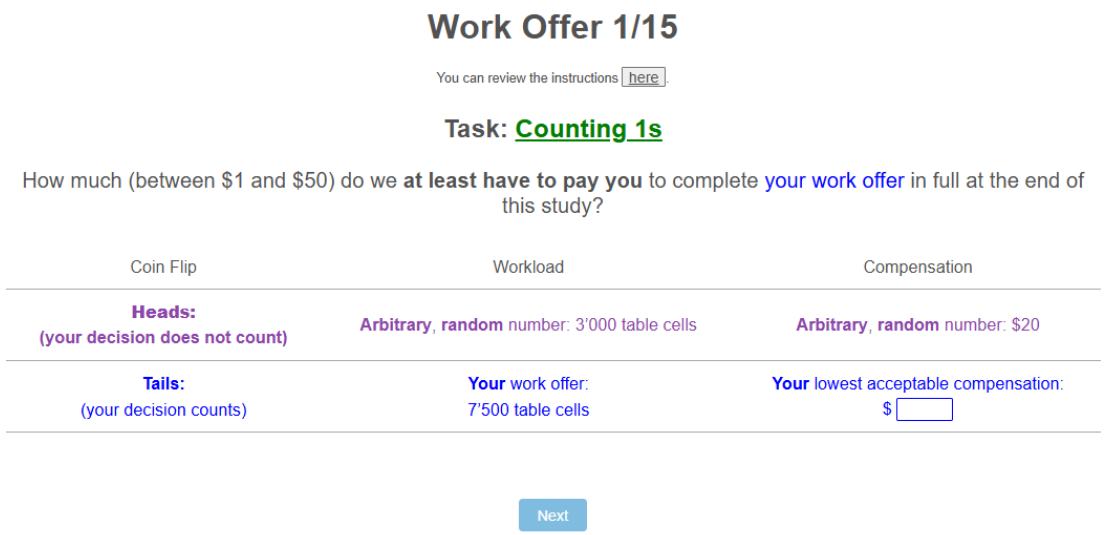

Figure I.33: Decision screen — *Random* treatment (*Effort Valuation*).

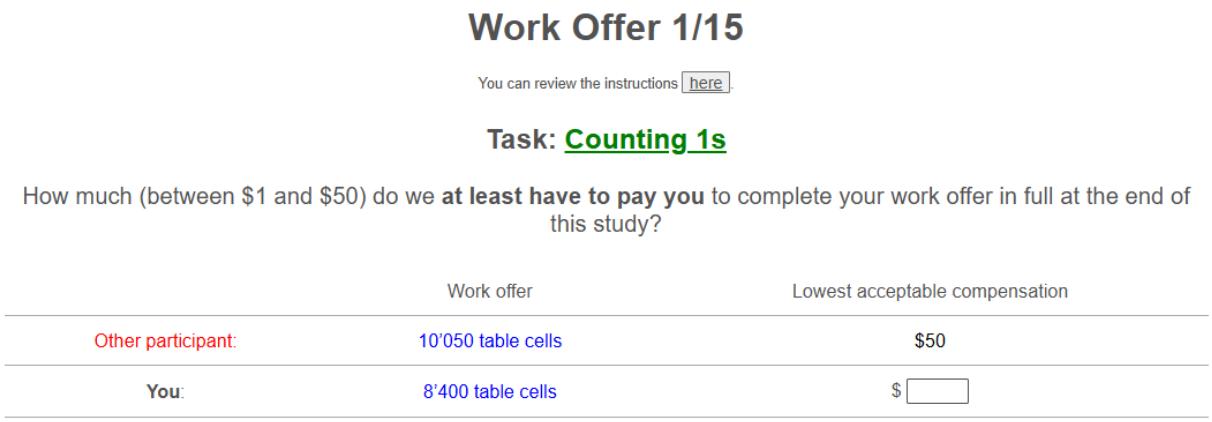

Figure I.34: Decision screen for counting ones — *Deliberation* treatment (*Effort Valuation*).

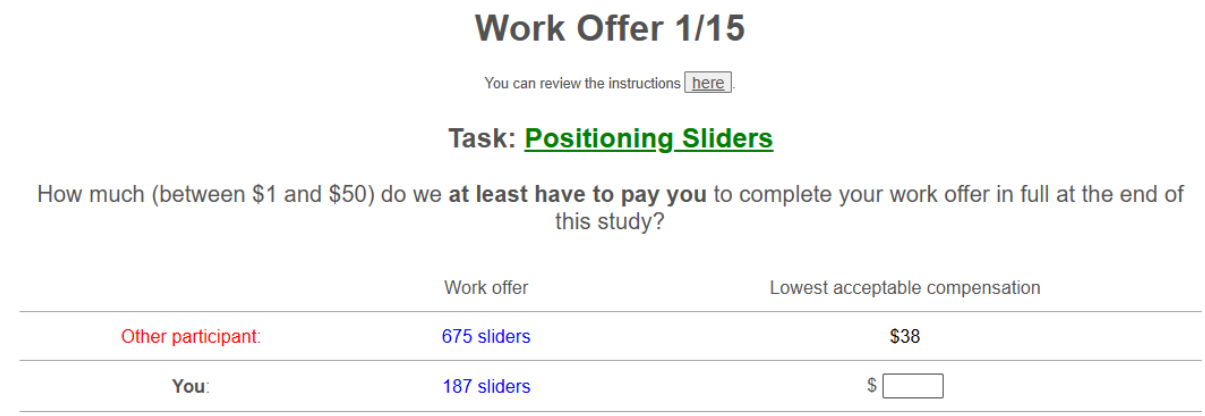

Figure I.35: Decision screen for positioning sliders — *Deliberation* treatment (*Effort Valuation*).

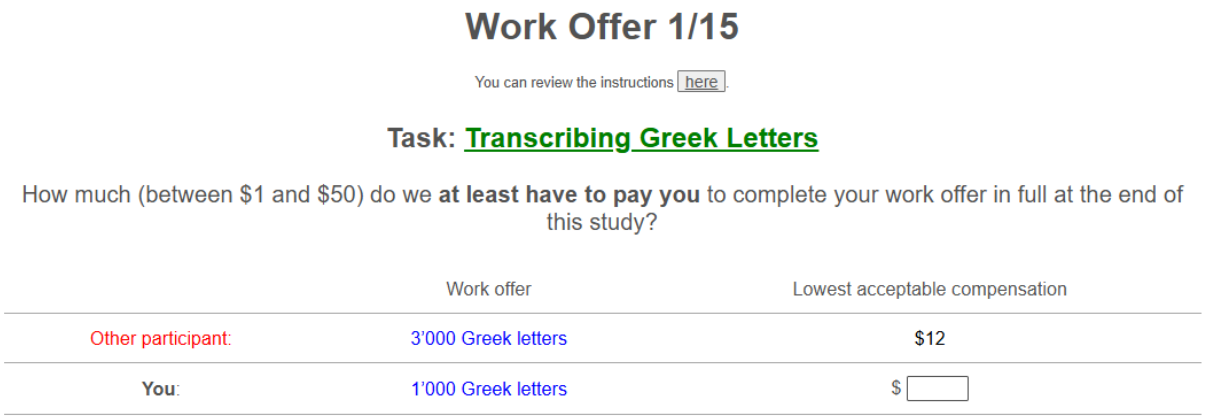

Figure I.36: Decision screen for transcribing Greek letters — *Deliberation* treatment (*Effort Valuation*).

I.2.2 Effort Choice

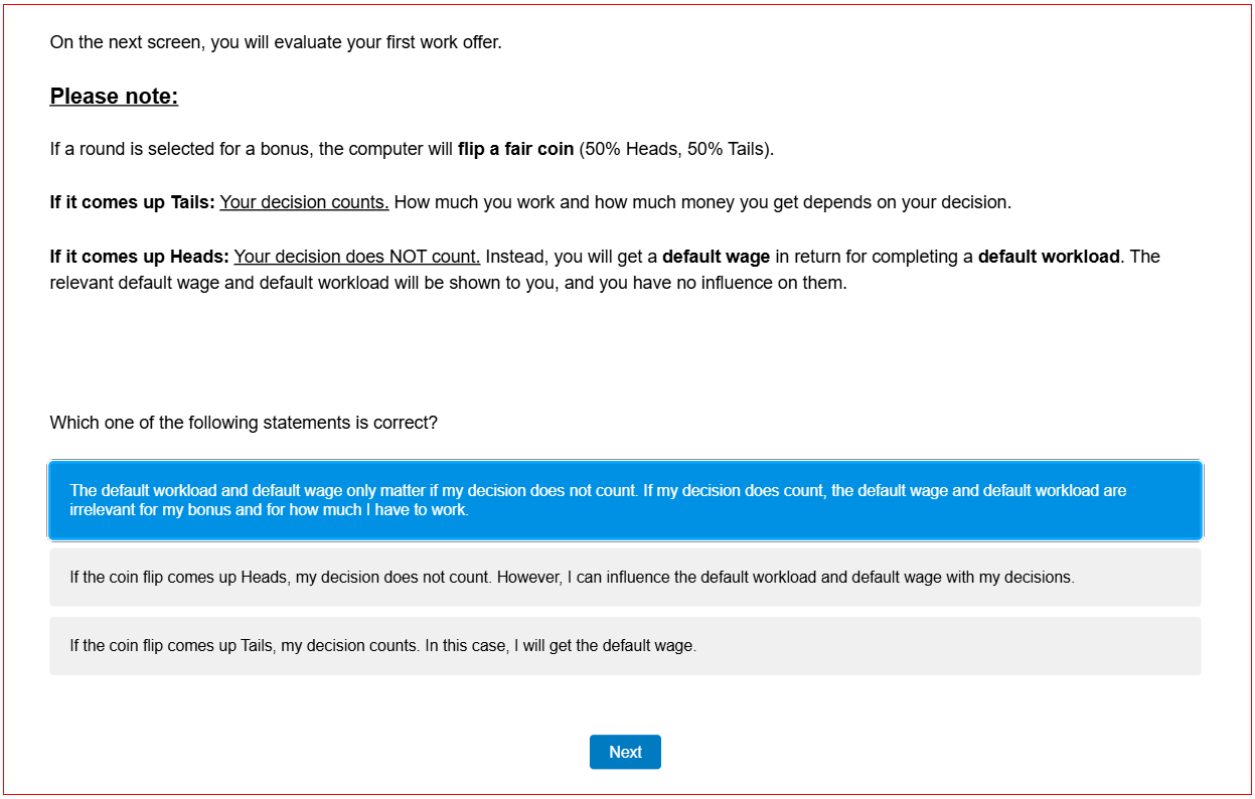

Figure I.37: Treatment announcement screen (*Effort Choice*).

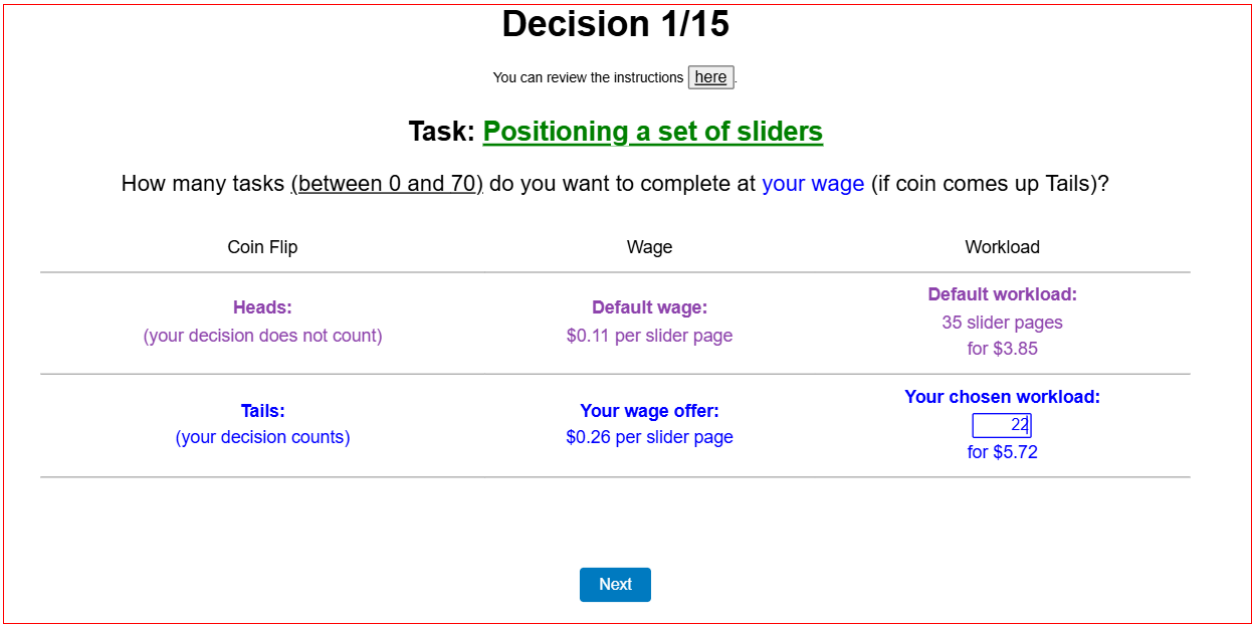

Figure I.38: Decision screen (*Effort Choice*).

I.2.3 Investment

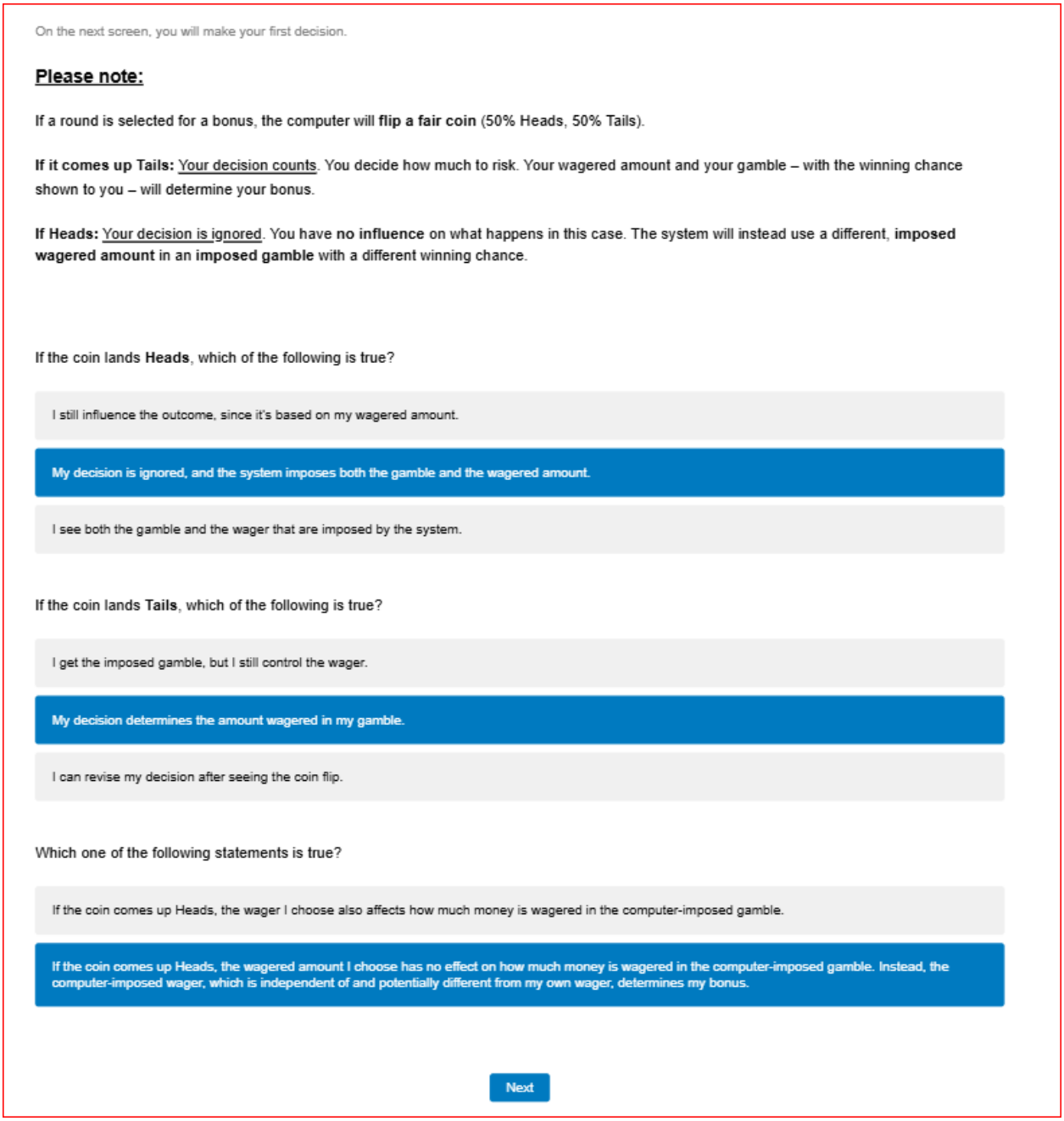

Figure I.39: Treatment announcement screen (*Investment*).

Note: In the comprehension item at the bottom of this screen, the intended correct answer is that the participant's decision is ignored and both the gamble and the wagered amount are imposed. The third response option (that the participant sees both imposed quantities) is also factually accurate—the imposed values were displayed on the decision screen (Figure I.40)—but does not describe the consequence of the coin flip that the item asks about.

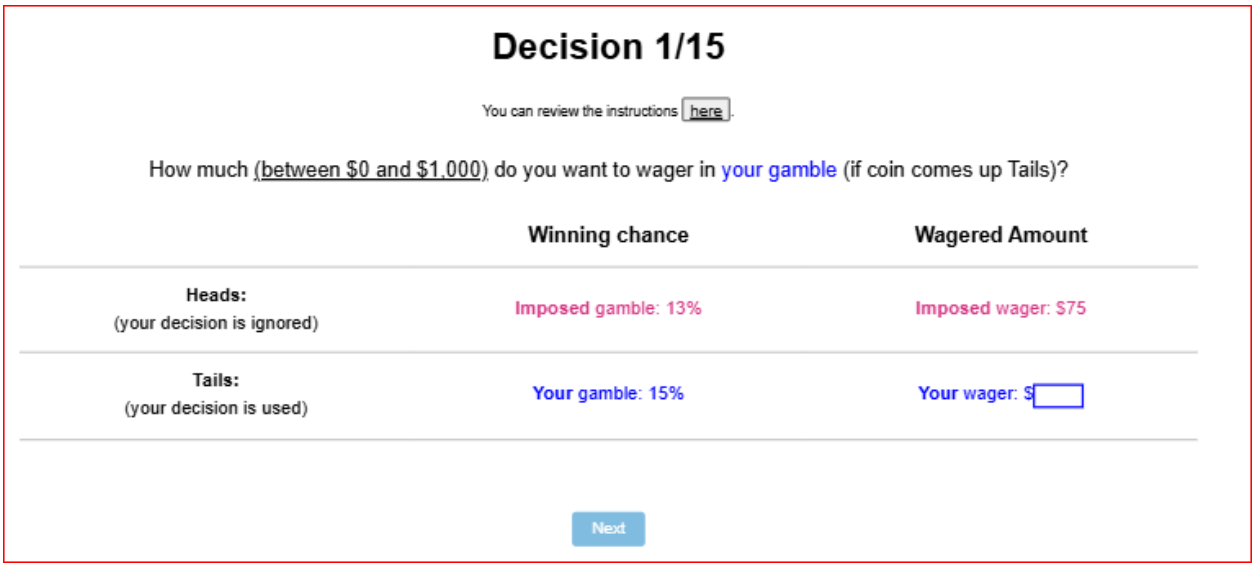

Figure I.40: Decision screen — *Expectations CP* treatment (*Investment*).

I.2.4 Beliefs

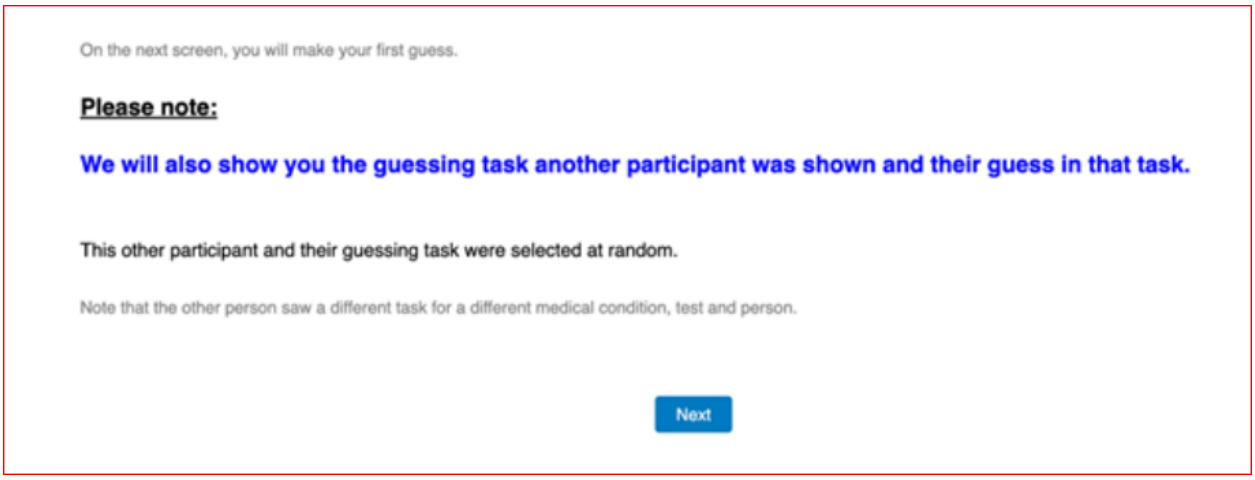

Figure I.41: Treatment announcement screen (*Beliefs*).

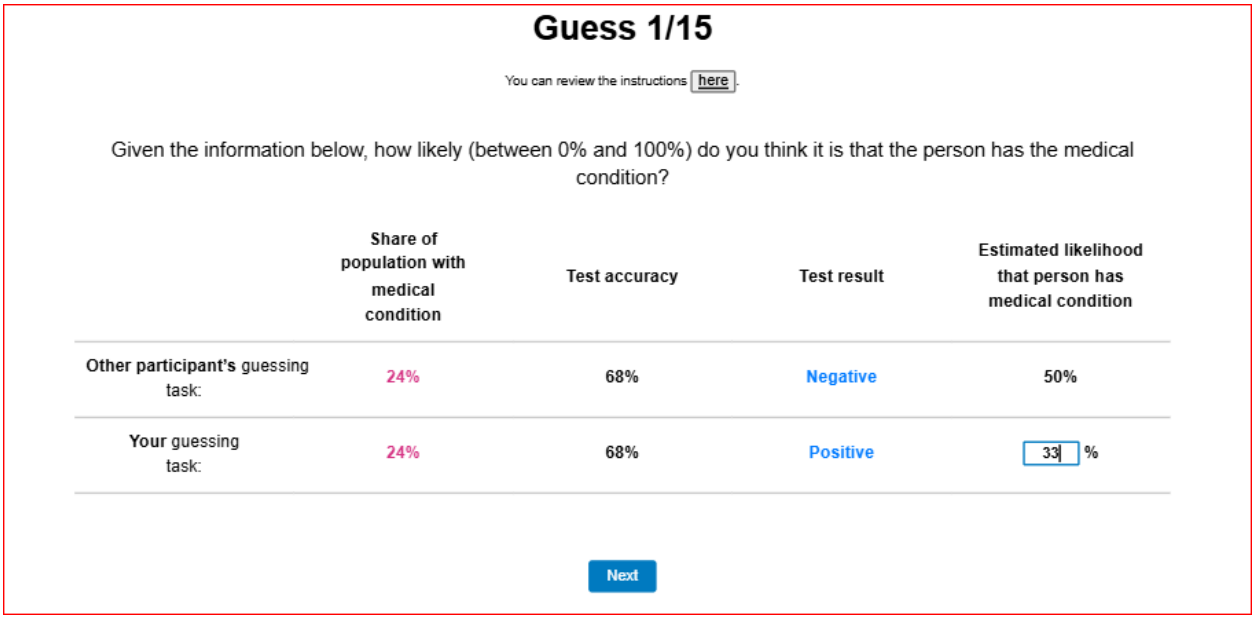

Figure I.42: Decision screen — *Social CP* treatment (*Beliefs*).

I.3 Forecaster Survey (Social Science Prediction Platform)

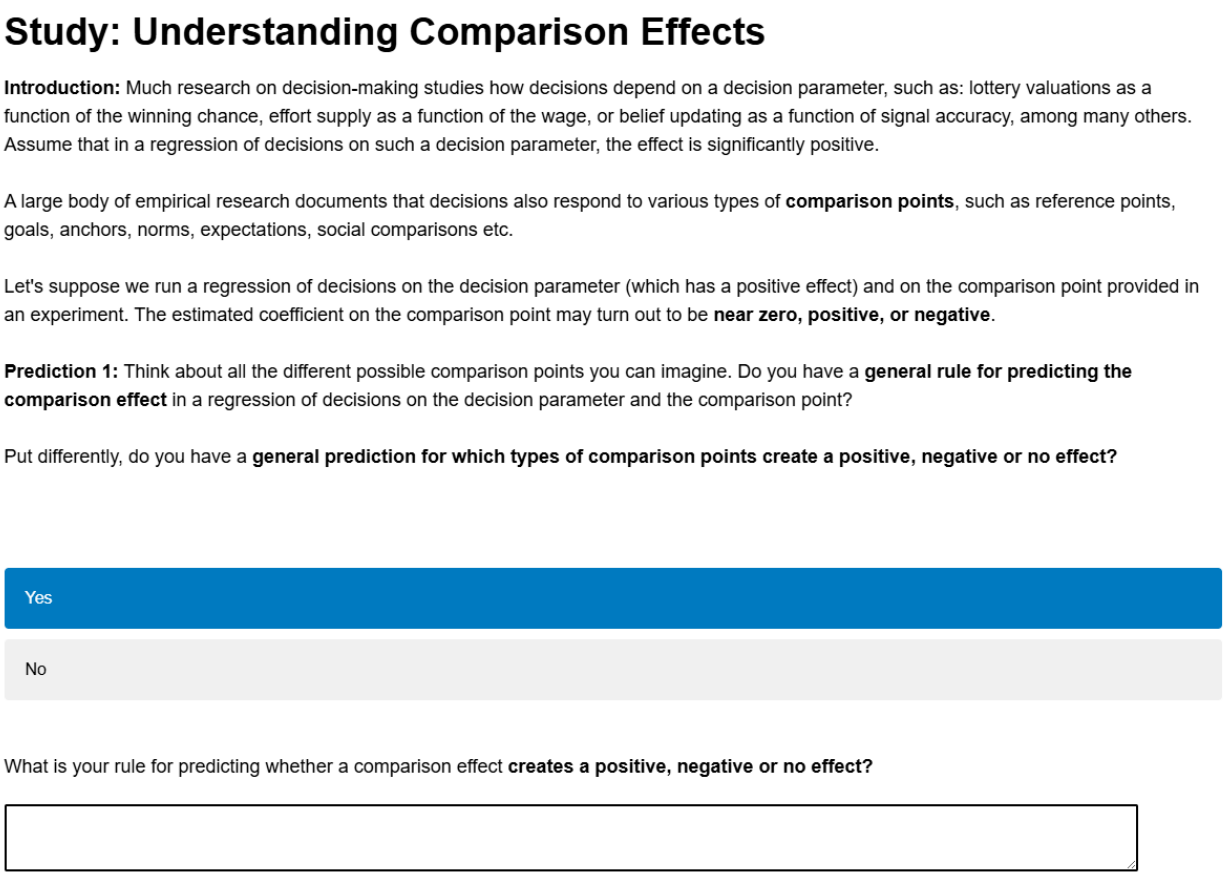

Figure I.43: Forecaster survey introduction and general rule prompt (Prediction 1).

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 I.44: 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 I.45: 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 I.46: 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):

$$\text{Log odds of stated guess} = \alpha + \beta_1 \times \text{Log prior likelihood of disease} + \beta_2 \times \text{Log likelihood ratio of test} + \beta_3 \times \text{Log prior likelihood of disease in other participant's task} + \beta_4 \times \text{Log likelihood ratio of test in other participant's task} + \beta_5 \times \text{Log odds of other participant's stated guess} + \varepsilon$$

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:**
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:**

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

Figure I.47: Beliefs: guessing task screen (Prediction 4).

Note: The regression shown to forecasters labels β_1 and β_3 as the "log prior likelihood of disease." As the surrounding text of the prompt states, the prior was transformed into log *odds*; these labels are shorthand for the log prior odds used in the estimation.

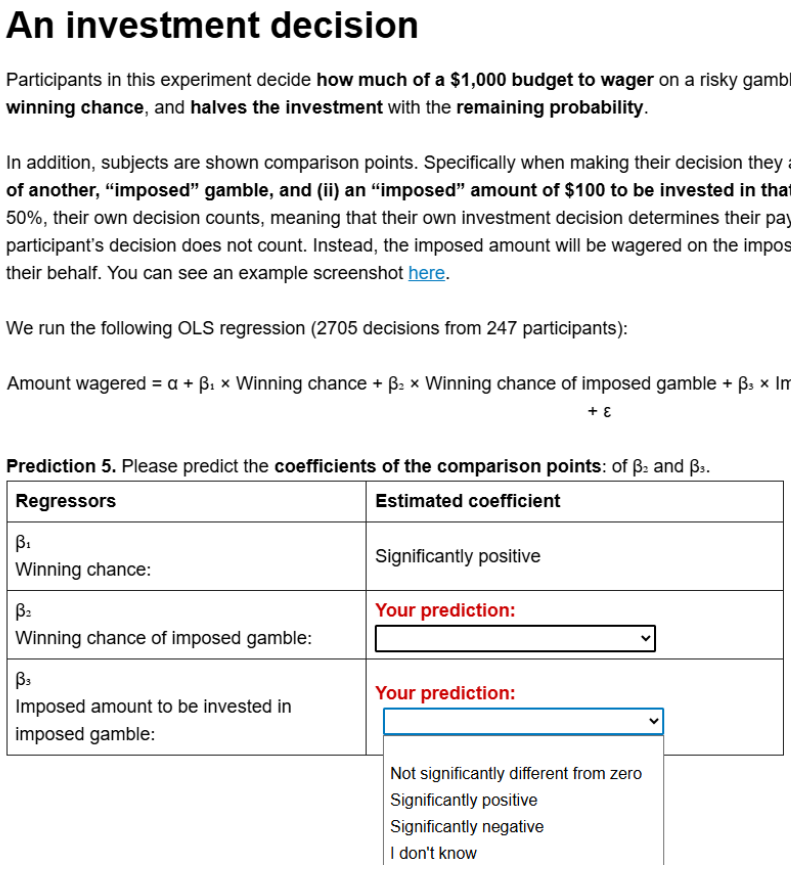

Figure I.48: Investment: decision prompt with imposed gamble and imposed amount (Prediction 5).

Note: Two typos appear in the prompt as fielded. First, the imposed amount is described as "$100," although the imposed amounts varied across rounds in the experiment; because forecasters were asked to predict the coefficient on the imposed amount, this description may have made that particular prediction harder to interpret. Second, the prompt reports "2705 decisions," whereas the estimation sample comprises 3,705 decisions from 247 participants. The regression equation itself was displayed correctly.

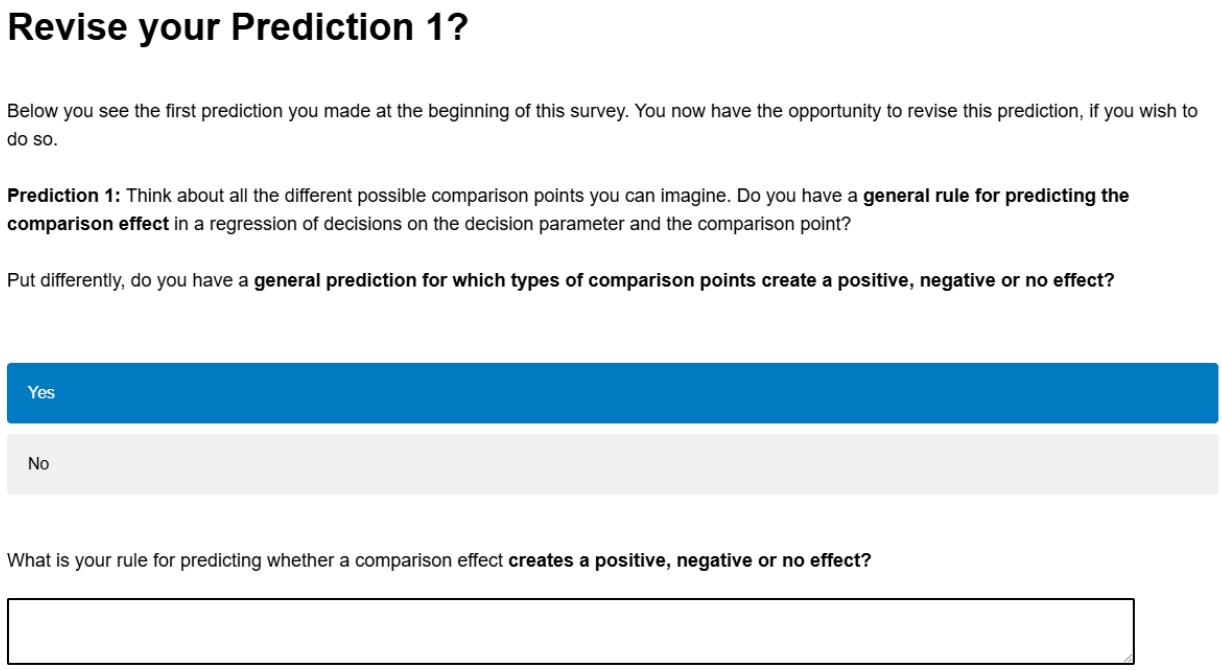

Figure I.49: Opportunity to revise the general rule prediction (Prediction 1).

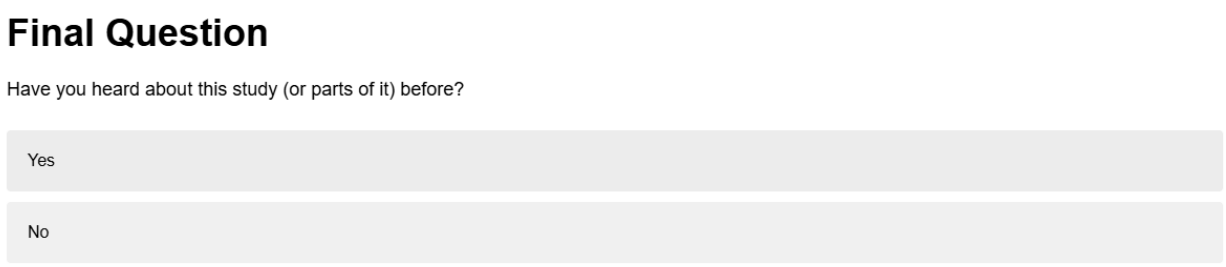

Figure I.50: Final prior exposure question.