Reference Dependence and Attribution Bias: Evidence from Real-Effort Experiments

Benjamin Bushong  Tristan Gagnon-Bartsch*
Michigan State University  Harvard University

February 26, 2020

Abstract

Reference points change with time. When learning about our tastes, we may fail to account for how our prior experiences were shaped by the reference points we previously held. In this paper, we experimentally demonstrate that participants’ impressions of a real-effort task are overly influenced by the surprise or disappointment they felt when they first tried the task. In our baseline experiment, participants learned from experience about one of two unfamiliar tasks, one clearly more onerous than the other. We manipulated participants’ expectations about which task they would face: some participants were assigned their task by chance just prior to their initial experience, while others knew in advance which task they would face. In a second session conducted hours later, we elicited those participants’ willingness to work again on their previously assigned task. Participants assigned the less-onerous task by chance were more willing to work than those who faced it with certainty. Conversely, participants assigned the more-onerous task by chance were less willing to work than those who faced it with certainty. These qualitative results, and the fact that differences in willingness to work were observed hours after first impressions were formed, are consistent with the idea that participants mistakenly attributed sensations of positive or negative surprise (relative to expectations) to the effort cost of their assigned task.

JEL Classification: C91, D03.

Keywords: attribution bias, reference dependence, learning from experience, loss aversion, experiment, real effort.

*E-mails: bbushong@msu.edu and gagnonbartsch@fas.harvard.edu. We thank Ned Augenblick, Katherine Coffman, Benjamin Enke, Christine Exley, David Laibson, Muriel Niederle, Devin Pope, Matthew Rabin, Joshua Schwartzstein, Andrei Shleifer, Charles Sprenger, and seminar audiences at the ASSA annual meeting (2020), Boston College, Boston University School of Management, Caltech, Cornell, Harvard, HBS, Michigan, Michigan State, Norwegian School of Economics (NHH), the North American ESA Conference, Purdue, Stanford, the Stanford Institute for Theoretical Economics, Tennessee, and UCSD Rady for comments. We thank Alexander Millner for sharing the annoying audio stimulus used in these experiments. We gratefully acknowledge financial support from the Eric M. Mindich Research Fund for the Foundations of Human Behavior.
1 Introduction

Evidence from both the lab and field suggests that our hedonic experiences are reference dependent: how we feel about an outcome depends on both its intrinsic value and how that value compares to a reference point (e.g. Kahneman and Tversky 1979; Medvec, Madey, and Gilovich 1995; Card and Dahl 2011; Abeler et al. 2011). But when our reference points change, do we properly account for how our past impressions were shaped by the reference points we used to hold? For instance, after a surprisingly good meal at a dingy restaurant, a diner may not appreciate that he will enjoy it less in the future once his expectations have adjusted upward. Such forecasts are challenging because the person must recall that his prior experience—which offers a glimpse of his intrinsic taste—was confounded by a sensation of surprise.¹ We explore this intuition in two real-effort experiments. Behavior in these experiments suggests a simple mechanism that leads to biased choices: people wrongly attribute sensations of elation or disappointment to the intrinsic (dis)utility of working on the real-effort task. This behavior is consistent with a model of attribution bias over reference-dependent utility developed in Gagnon-Bartsch and Bushong (2019).

To further motivate our experimental design, consider a worker completing a series of short-term jobs. Each day the worker is randomly assigned to one of two tasks—one more desirable than the other. Thus the job she faces each day comes with a positive surprise or a disappointment. A “misattributing” worker fails to account for these sensations as she forms her impressions of the tasks. Consequently, she develops incorrect beliefs about how much she enjoys each. Specifically, when she is fortunate and assigned to the desirable job, she wrongly attributes the positive feelings of surprise to the intrinsic enjoyment of the task. Faced with the opportunity to do this task long-term, her distorted impressions make her too willing to accept the job—when it is fully anticipated and no longer comes as a pleasant surprise, the activity will no longer be as enjoyable. By contrast, when she is unfortunately assigned to the less desirable task, she (mis)attributes the sensation of disappointment to the intrinsic disutility of that task, and becomes too hesitant to work in that role. In both cases, the worker forms biased impressions of the task because she neglects the degree to which her utility experienced in the past depended on her expectations.²

Our experiments explore this form of attribution bias. In each, participants worked on a novel

¹Studies in both psychology and economics demonstrate that people make mistakes when recalling the sources of their feelings. Dutton and Aron (1974) show that opinions of a newly-met person depend on unrelated situational factors—e.g., current state of excitement or fear. Meston and Frohlich (2003) replicate and extend this seminal result to broader settings. Recent evidence in economics (Simonsohn 2007, 2010; Haggag et al. 2019) demonstrates that, when assessing the value of a good or service, people incorrectly attribute state-dependent sensations caused, for instance, by weather or thirst to the underlying quality of the good. We discuss additional evidence for such mistakes in attribution in Section 2.

²While reference dependence absent misattribution captures the notion that potential surprises or disappointments loom large in preferences, we provide a mechanism for why past sensations of surprise continue to loom large in both memory and subsequent forecasts.
real-effort task and formed their initial impressions while facing an exogenously manipulated reference point. Later, they chose how much additional work to do (for additional pay) after their reference point had changed. We examine whether these subsequent choices were influenced by their initial reference point despite it being no longer relevant. In Experiment 1, we endowed participants with differing chances of facing either of two previously-unexperienced tasks, one clearly more enjoyable than the other. In an initial session, participants gained experience working on their assigned task. Then, several hours later, we elicited their willingness to complete additional work. Comparing those who were assigned to the same task, we find that the ex-ante chance of facing each of the two tasks significantly altered participants’ subsequent willingness to work. In our second experiment, we manipulated initial expectations within subjects to examine how willingness to work changed over one week as participants’ expectations adapted. As with Experiment 1, we find that participants’ willingness to work was shaped by the elation or disappointment they experienced while forming their initial impressions. Our evidence therefore suggests a specific, previously unexplored form of attribution bias.3

In Section 2, we present an abridged version of the model from our theoretical companion paper, which guided our experimental design. Following Bell (1985) and Kőszegi and Rabin (2006), we assume the decision maker experiences expectations-based reference-dependent utility composed of two parts: consumption utility, which corresponds to the classical notion of payoffs, and gain-loss utility, which is proportional to the difference between the consumption utility earned and what the person expected. As alluded to above, a “misattributor” correctly recalls how she felt after each experience, but wrongly attributes sensations of surprise or disappointment—i.e., the gain-loss component of her utility—to the underlying outcomes—i.e., the consumption component. She thus forms biased impressions of the outcomes she faced.

In Section 3 we describe Experiment 1, which involved 886 subjects recruited from Amazon’s Mechanical Turk (MTurk). In an initial learning session, each subject listened to audio files of Amazon book reviews and had to determine whether each review was endorsing or criticizing the book. This simple-yet-tedious classification task came in two variants. One variant—which we call noise—included an annoying sound layered on top of the audio review. The second variant—which we call no-noise—had no additional sound added to the audio review.

Before the initial learning session, we endowed participants with different chances of facing either task. One third of participants were assigned to a task from the onset of the experimental

---

3Our experiments also connect to a literature that considers how prior expectations might influence impressions through either assimilation or contrast. This research has highlighted that when outcomes deviate from expectations, a person might either assimilate that experience—interpret it in favor of their current beliefs (as in, e.g., Rabin and Schrag 1999; Fryer, Harms, and Jackson 2018)—or contrast it—interpret the experience against their expectations (e.g., Oliver 1977, 1980; and Boulding et al. 1993). It remains an open question when each force dominates; we find contrast effects are predominant in our environment.
instructions (i.e., they faced no uncertainty). Another third of participants flipped a coin to determine which task they would face moments before their first experience with that task (i.e., they faced a 50% chance of either task). Finally, the remaining participants were assigned to a task with near certainty (i.e., they faced a 99% chance of one task and a 1% chance of the other). Put together, this design generates six groups, which result from crossing the three manipulations in expectations described above with the ultimate task a participant faced: \{control, coin flip, high probability\} × \{noise, no noise\}. After reading the instructions (and, if applicable, resolving any uncertainty), participants completed eight rounds of their assigned task. Knowing that they would later be asked about their willingness to work on this task, these initial trials gave participants an opportunity to learn about their preferences. More than eight hours later, we elicited their willingness to continue working on their assigned task for additional pay.

We first examine how willingness to work differed between participants in the control and coin-flip treatments. Our misattribution model predicts that participants who were assigned the noiseless task via the coin flip would form overly positive beliefs about that task, since their initial impressions came with a sense of positive surprise. Thus, participants in the coin flip + no noise group would be more willing to work than those in the control + no noise group, despite the two groups ultimately facing the same task. By contrast, our model predicts that those assigned the noisy task via the coin flip would form overly negative impressions of the task, as their initial experience was colored by their disappointment. We therefore predict that participants in the coin flip + noise group would be less willing to work than those in the control + noise group.

Indeed we find these effects, as previewed in Figure 1. When the stakes were highest, participants who were assigned the noiseless task via the coin flip were 20% more willing to work than those who faced that task with certainty, while those who were assigned the noisy task via the coin flip were 25% less willing to work than those who faced that task with certainty. We interpret this as evidence that participants misattributed sensations of positive and negative surprise that arose during the initial learning session to their underlying enjoyment of the task. Additionally, we discuss how classical models and reference-dependent models without misattribution struggle to predict these results.\(^4\)\(^5\) Our high-probability treatment helps rule out potential informational ex-
Figure 1: Labor supply curves across treatments. Each point represents the average willingness to work for a fixed payment as elicited using the BDM mechanism. Relative to groups whose assignment did not induce surprise (control groups), those assigned by chance (coin flip groups) demonstrate greater willingness to work when assigned the noiseless task and less willingness to work when assigned the noisy task.

...planations stemming from participants in the control and coin-flip inferring different information from the experimental design itself.6

Experiment 2—presented in Section 4—adopts a within-subject design (n = 87) in a laboratory setting. We elicited each participant’s willingness to work in two different sessions, separated by one week. Our identification of misattribution in this setting stems from changes in a participant’s reference point over the course of the week. In the first session, each participant flipped a coin to determine whether she faced the good or bad task and then immediately completed five trials of that task. Directly after this learning phase, we elicited the participant’s willingness to continue working at that task. One week later, the same participants returned but there was no coin flip:...
each knew ahead of time that she would again face the same task as she did in the first session (with certainty). In that second session, each participant completed five trials of her (previously) assigned task and then stated her willingness to continue working.

We examine the difference in a participant’s willingness to work between Session 1—when her task came as a surprise—and Session 2—when that same task was completely expected. We find that participants who were pleasantly surprised in the first session were less willing to work in the second week than in the first, while those who were negatively surprised in the first session were more willing to work in the second week than the first. Again, this is consistent with misattribution: a participant who was, say, positively surprised (i.e., assigned the good task) may have attributed this sensation to a quality of the underlying task. Upon trying that same task again a week later—when it was no longer pleasantly surprising—the task may have failed to live up to the previous experience, thus lowering willingness to work.

Jointly, our two experimental designs offer a compelling case in favor of attribution bias. Some concerns that might apply to Experiment 1—e.g., effects driven by differences in information or reference points that are slow to adapt—do not apply to Experiment 2, and vice versa. Additionally, while Experiment 1 provides strong evidence of misattribution, Experiment 2 provides a recipe for identifying heterogeneity across subjects (though our current experiment is underpowered for deep conclusions along these lines).

The form of attribution bias we document can capture a number of anomalous (yet well-documented) patterns in belief updating, which we explore in greater detail in Gagnon-Bartsch and Bushong (2019). There, we demonstrate that a misattributor relies too heavily on her personal experience—in particular, recent experiences—when making decisions. Additionally, when comparing her outcomes against past experiences, a misattributor may exhibit sequential contrast effects, whereby she perceives today’s outcomes as better the worse was yesterday’s. Finally, fixing the outcomes she faces, a misattributor forms the most optimistic beliefs after experiencing a sequence of increasing outcomes. Our model provides a mechanism for these effects.

Reference dependence with attribution bias also provides a novel explanation for the findings that willingness to pay for the bundle is significantly higher among those who sampled the better one last. Similarly, Haisley and Loewenstein (2011) show that advertising promotions are most effective when sequenced in increasing order of value—that is, the high-value promotional item is given last. Several authors argue that such assessments follow a mechanism like ours (e.g., Tversky and Griffin 1991; Loewenstein and Prelec 1993; Baumgartner, Sujan, and Padgett 1997). Other forms of sequential contrast effects have been documented in decisions made by teachers (Bhargava 2007), speed daters (Bhargava and Fisman 2014), judges assessing asylum seekers, reviewers of loan applications, baseball umpires (Chen, Moskowitz, and Shue 2016), and investors (Hartzmark and Shue 2018).
of Gneezy and List (2006) and similar experiments. In their field experiment, surprisingly high earnings led workers to increase effort, but this effect diminished over time. The evidence we provide speaks to both why these payments increased effort in the short-term and why surprising wages failed to motivate longer-term changes in behavior after workers’ reference points adapted.

More broadly, our evidence stresses the importance of initial expectations on subsequent judgments. For example, misattribution can have implications for firms’ marketing techniques when consumers are experimenting with new products. Our results suggest that setting a high bar can backfire if consumers judge these new goods against their lofty expectations. Thus, while hyping a product may help early sales, such marketing efforts can hurt if early adopters then underestimate the product’s quality as a result of contrasting it against a high reference point.\textsuperscript{9} This intuition suggests that managers and firms should engage in expectations management—a practice commonly observed in marketing, politics, and finance.\textsuperscript{10}

2 Theoretical Framework and Motivation

In this section, we present a streamlined version of our model of reference dependence with attribution bias (Gagnon-Bartsch and Bushong 2019), which shaped our experimental designs. We also discuss motivating evidence for the central assumptions underlying the model.

Preferences and Misattribution. Following Kőszegi and Rabin (2006; henceforth KR), we assume that the agent’s overall utility has two additively-separable components. The first component, “consumption utility”, corresponds to the material payoff traditionally studied in economics, which we denote by $v \in \mathbb{R}$.\textsuperscript{11} The second component, “gain-loss utility”, derives from comparing $v$ to a reference level of utility or a “reference point”. Following Bell (1985), we take this reference point to be the agent’s expectation of $v$, and we consider a simple piecewise-linear specification of gain-loss utility. Specifically, if the agent believes that consumption utility is distributed according

\textsuperscript{9}Political scientists, for example, have argued that discrepancies between a politician’s performance and citizens’ expectations play a key role in how citizens perceive that politician (see, e.g., Patterson et al. 1969; Kimball and Patterson 1997). Likewise, marketing has emphasized the role of expectations on perceived quality of service (see, e.g., seminal works from Oliver 1977, 1980; and Boulding et al. 1993).

\textsuperscript{10}For example, firms commonly use a variety of mechanisms to “walk down” investors’ expectations prior to earnings announcements, utilizing strategic accounting of working capital and cash flow (Burgstahler and Dichev 1997), sales (Roychowdhury 2006), or distorting analyst forecasts (e.g., Richardson, Teoh, and Wyckoki 2004). Relatedly, Kopalle and Lehmann (2006) study how a firm should optimally restrain quality expectations when consumers have preferences that depend on those expectations.

\textsuperscript{11}We interpret $v$ as if it derives from a classical Bernoulli utility function $u_C : \mathbb{R}_+ \to \mathbb{R}$ over consumption realizations $x \in \mathbb{R}_+$ such that $v = u_C(x)$, but we work directly with consumption utility $v$ to reduce notational clutter.
to CDF $\hat{F}_V$ with a mean value $\hat{E}[V]$, then gain-loss utility from outcome $v$ is

$$
n(v|\hat{E}[V]) = \begin{cases} 
  v - \hat{E}[V] & \text{if } v \geq \hat{E}[V] \\
  \lambda (v - \hat{E}[V]) & \text{if } v < \hat{E}[V],
\end{cases}
$$

where parameter $\lambda \geq 1$ captures loss aversion. Thus, given expectations $\hat{E}[V]$, the person’s total utility is

$$
u(v|\hat{E}[V]) = v + \eta n(v|\hat{E}[V]),
$$

where $\eta > 0$ is the weight given to sensations of gain and loss relative to absolute outcomes.\(^{12}\)

Our notion of misattribution arises in environments where an individual attempts to learn about the consumption utility $v$ she derives from a prospect. We assume the agent uses her total utility to infer $v$. When doing so, a misattributor under-appreciates the extent to which her past experiences were influenced by reference dependence, and she infers $v$ using a misspecified model that weights the gain-loss component of her utility by a diminished factor $\hat{\eta} \in [0, \eta]$.\(^{13}\) That is, she correctly recalls how happy she felt following outcome $v$, but she fails to fully account for how sensations of surprise or disappointment affected her total utility. Specifically, she infers outcome $\hat{v}$ as if her utility function were $\hat{u}(\hat{v}|\hat{E}[V]) = \hat{v} + \hat{\eta} n(\hat{v}|\hat{E}[V])$, and thus $\hat{v}$ solves $\hat{u}(\hat{v}|\hat{E}[V]) = u(v|\hat{E}[V])$.

Equations 1 and 2 imply that this misencoded outcome, $\hat{v}$, takes the following form:

$$
\hat{v} = \begin{cases} 
  v + \left(\frac{\eta - \hat{\eta}}{1 + \hat{\eta}}\right) (v - \hat{E}[V]) & \text{if } v \geq \hat{E}[V] \\
  v + \lambda \left(\frac{\eta - \hat{\eta}}{1 + \hat{\eta}}\right) (v - \hat{E}[V]) & \text{if } v < \hat{E}[V].
\end{cases}
$$

Thus, the encoded outcome is biased upward when the true outcome beats expectations, and biased downward when it falls short of expectations. This bias is proportional to the deviation between the true outcome and expectations and increases in the degree of misattribution—i.e., as $\hat{\eta}$ decreases. Furthermore, when the agent is loss averse—i.e., $\lambda > 1$—a loss is misencoded by a greater extent than an equal-sized gain. Finally, we assume the agent uses this misencoded outcome to update

\(^{12}\)Unlike KR, we do not impose rational expectations; indeed, a key feature of our framework posits that the agent’s (potentially biased) beliefs determine her reference point. Furthermore, our predictions do not substantively depend on whether we assume a deterministic reference point (à la Bell and Equation 1, above), or a stochastic reference point (à la KR, where an outcome is compared to each possible alternative outcome under $\hat{F}_V$ and every such comparison is weighted by the likelihood of the alternative).

\(^{13}\)There are at least two interpretations of how these biased perceptions are formed: (1) the agent improperly encodes each outcome as they happen—which seems most plausible in settings where the determinants of consumption utility are not directly observable (e.g., one’s disutility of working on an unfamiliar task or the quality of a meal); (2) the agent retrieves a distorted memory of an outcome when attempting to recall its value—which seems most plausible in settings where outcomes are easily observed (e.g., one might remember an unexpectedly high price from a previous transaction as higher than it truly was despite knowing the true price when the transaction took place).
her beliefs.

To illustrate the model, recall the example from the introduction wherein a worker’s daily task is assigned at random: some days she faces a relatively enjoyable task and other days she faces a more-onerous one. When the worker is assigned the onerous task, she simultaneously experiences both a bad material outcome and a sensation of disappointment—her task is worse than average. A misattributor fails to fully account for this sensation of disappointment and wrongly attributes it to the underlying disutility of the task. She thus will recall her assigned task as more onerous than it really was. When the worker is assigned the more pleasant task, she simultaneously faces an easier job and a pleasant surprise, and recalls the task as even better than it really was.

For ease of presentation, we have thus far discussed the case where consumption, $v$, is unidimensional. However, our experiments examine a setting with two dimensions—money and effort. To accommodate this, we extend the model outlined above: given expectations $\hat{E}[V^k]$ along each dimension $k \in \{m, e\}$, the agent’s total utility from realization $v = (v^m, v^e)$ is

$$u(v | \hat{E}[V]) = \sum_{k \in \{m, e\}} \left\{ v^k \text{ Consumption utility} + \eta n \left( v^k | \hat{E}[V^k] \right) \right\}.$$  

(4)

The misencoded outcomes, $\hat{v}^k$, are then defined as in Equation 3 along each dimension. That is, a misattributor recalls an outcome $\hat{v}^k$ such that

$$\hat{v}^k + \hat{\eta} n \left( \hat{v}^k | \hat{E}[V^k] \right) = v^k + \eta n \left( v^k | \hat{E}[V^k] \right).$$

Motivating Evidence. Building on studies by Kahneman and Tversky (e.g., 1979), empirical evidence from both psychology and economics has reinforced the underlying idea of reference-dependent utility. Recent papers have demonstrated that reference dependence affects behavior across a wide range of contexts. This evidence spans labor supply among taxi drivers (Camerer et al. 1997; Crawford and Meng 2011; Thakral and Tò 2020), domestic violence resulting from unexpected football losses (Card and Dahl 2011), decisions in game shows and sports (Post et al. 2008; Pope and Schweitzer 2011; Allen et al. 2015; Markle et al. 2015), and even the behavior of capuchin monkeys (Chen et al. 2006).

It is well established that reference points shape behavior, but what exactly determines the reference point remains contested. To discipline their theory, KR assume that the reference point corresponds to forward-looking expectations—that is, the agent’s own future actions shape her current reference point.\(^\text{15}\) Our experimental predictions do not hinge on forward-looking expectations—

\(^{14}\)We assume that a misattributing agent makes decisions that maximize her true expected utility in Equation 2; that is, she makes decisions according to her actual value of $\eta$. Under this interpretation, misattribution only influences behavior via the erroneous beliefs an agent forms when misinterpreting past experiences. While this is our preferred interpretation, it is worth emphasizing that our predictions are robust to the agent making decisions according to her misspecified parameter value, $\hat{\eta}$.

\(^{15}\)The evidence on such forward-looking reference points is mixed. In favor of a forward-looking solution, Abeler
many specifications of the reference point and alternative solution concepts will generate the patterns of behavior we observe. Specifically, so long as participants (at least partially) adopt the reference points we exogenously impose, our theoretical results will hold.

In contrast to reference-dependent preferences, attribution bias has received little attention in the economics literature, though various errors in attribution have been explored in psychology. Often referred to as the “fundamental attribution error” or “correspondence bias” in that literature (e.g., Ross 1977; Gilbert and Malone 1995), these errors may share a common psychology with that of misattribution of reference-dependent utility: transient sensations (e.g., sensations of surprise or disappointment) are incorrectly attributed to an underlying, stable trait. In the economics literature, Simonsohn (2007, 2009) explores the effect of a transient shock (weather) on the subsequent preferences of would-be college students and admissions officers. Simonsohn (2007) demonstrates that college applicants with particularly strong academic qualities were evaluated higher by admissions officers when the weather on that evaluation day was poor. Simonsohn (2009) shows that incoming freshman are more likely to matriculate at an academically rigorous school when the weather on their visit day to that school was cloudy versus sunny. Relatedly, a series of papers show that CEOs (Bertrand and Mullainathan 2003) and politicians (Wolfers 2007; Cole, Healy, and Werker 2012) are rewarded for luck. Recent laboratory experiments have replicated this result (see, e.g., Brownback and Kuhn 2019; Erkal, Gangadharan, and Koh 2019).

Finally, Haggag et al. (2019) provide clean evidence of another novel form of attribution bias: 

et al. (2011) study real-effort provision in the presence of stochastic wages and demonstrate that varying expectations over these wages changes effort. Gill and Prowse (2012) look at a two-person sequential game in which players exert real effort and the probability of winning a prize depends on the total effort exerted. Importantly, the probability of winning in their experiment is linear in effort, meaning that the second player’s behavior should not depend on that of the first. However, the authors find a discouraging effect of low first-player effort that is consistent with a model of expectations-based reference dependence. Finally, Karle et al. (2015) show that food choices depend on the realization of uncertain prices in a way that is consistent with forward-looking expectations. Despite this evidence, the determination of reference points is unclear, and although these studies provide evidence in favor of a forward-looking solution, a growing literature pushes against the evidence above. In an experiment similar to that of Karle et al. (2015), Wenner (2015) finds no evidence for the KR model, but attributes his results to non-equilibrium behavior. And while Ericson and Fuster (2009) demonstrate that the endowment effect is at least partially driven by expectations of future endowments, Heffetz and List (2014), Heffetz (2018), and others provide contradictory evidence.

Although Kahneman and Tversky’s (1979) Prospect Theory supposes that people behave as if they experience reference-dependent sensations or hedonics, those authors do not take a strong stand on whether this behavior truly reflects hedonic sensations. Many studies provide suggestive evidence that sensations of positive and negative surprise are a hedonic phenomenon. Rutledge et al. (2014) shows that a reference-dependent model predicts self-reported happiness during a simple gambling experiment. Additionally, the authors use fMRI to find a neural signal in the midbrain that follows this reference-dependent model. These signals are commonly interpreted as stemming from a non-hedonic reinforcement-learning model that is encoded by midbrain dopamine neurons (Schultz, Dayan, and Montague 1999). These reinforcement-learning models predict a signal very similar to that of the gain-loss function (absent loss aversion). Accordingly, previous neuroscience evidence on reinforcement learning, when reinterpreted through this lens, may be evidence of reference dependence. Finally, recent papers show these reference-dependent signals extend beyond the midbrain to higher levels of cortex in both humans (e.g., Hayden et al. 2011; Hill, Boorman, and Fried 2016) and other primates (e.g., Bayer and Glimcher 2005). Such signals in the ventral medial prefrontal cortex (an area associated with experienced utility) may suggest a neural basis for reference-dependent hedonics.
people wrongly attribute state-dependent fluctuations in utility to some stable underlying quality. In an experiment, the authors show that participants value an unfamiliar beverage more if they first drink it while thirsty rather than sated. Likewise, using field evidence they show that good weather during a person’s visit to a theme park increases the likelihood that person plans to return. The two models (ours and theirs) are complementary and apply attribution bias to different aspects of utility. As such, the models make different predictions. Specifically, errors in attribution in their model lead decision-makers to underestimate the utility difference between two outcomes. Our model predicts—and we observe—the opposite. Moreover, in some settings the biased forecasts that result from Haggag et al.’s formulation may wash out with ample experience. These errors can persist under misattribution of reference dependence. This distinction stems from the fact that Haggag et al. rule out complementaries where past experiences influence today’s consumption utility. Reference dependence naturally introduces this complementarity, as past experiences form the reference point against which today’s consumption is evaluated.

3 Experiment 1

In this section, we present our between-subject experiment, which we conducted on Amazon’s Mechanical Turk (MTurk). We first describe the experimental design. Next, we provide theoretical predictions of both rational-learning models and our model of misattribution. Finally, we analyze our experimental data, noting throughout how the results are consistent with our notion of misattribution yet inconsistent with rational-learning models with or without reference dependence.

3.1 Design

We recruited approximately 900 participants for a two-session experiment. The first session was an “initial-learning phase,” which gave participants experience with our real-effort task. During the second session, we elicited participants’ willingness to complete additional work on the same task they previously faced. Participants took an average of 10 minutes to complete the first session and 15 minutes to complete the second. We paid participants a fixed fee of $4 for successfully completing both sessions, which translates to an hourly wage of approximately $9.60. Participants

17For example, misattribution of state-dependent utility can cause an agent to mislearn the mean outcome in the short run, but the bias will vanish in the long-run if states are independent from the timing of consumption. To illustrate, consider a diner learning about the quality of a restaurant. If she visits the restaurant when she’s both hungry and not hungry, she will correctly learn the average quality. Accordingly, the framework from Haggag et al. (2019) may best apply to situations where choices are based on limited experience. In contrast, misattribution of reference-dependent utility can lead a loss-averse agent to develop persistent misperceptions about the average quality.

18Participants were required to be located within the U.S. and to have completed at least 100 prior jobs on MTurk with a 95% approval rating.
could earn up to $6.50 (total) depending on their willingness to complete additional work and chance.

Each participant worked on one of two tasks. In both tasks, participants listened to audio reviews of books and determined whether each review was positive or negative. After each determination, participants were notified if they were incorrect. Figure 2 depicts this interface. Our two tasks differed in a single way: one version included unaltered audio, while in the other the audio file was overlaid with an annoying noise. This noise was a composite of a fork scraping against a record and a high-frequency tone. The noise played approximately 15 decibels lower than the peak levels of the audio in the review when played at moderate volume. Hence, the noise was annoying but did not hinder participants’ ability to classify the audio reviews.

Our overlaid-audio design has an important feature: participants who faced the annoying noise could not avoid the noise and still successfully complete the task. Additionally, we ensured participants actually listened to the audio reviews using three techniques. First, participants were required to answer at least six out of the eight mandatory classifications correctly during the first session or else they would be removed from the study without pay. Additionally, we hid the response buttons for the first ten seconds of each review, which required participants to listen to a substantial portion of the review before guessing. Finally, many of the reviews featured important details in the late part of the review. To prohibit participants from reloading the web session (and thus generate new random numbers) in an attempt to avoid the noise, we blocked multiple logins and required unique email authentication to access each session of the experiment.

We now provide greater detail for each of the two sessions of the experiment.

**Session 1: Initial-Learning Phase.** Participants completed eight reviews in the initial-learning session of the experiment. We instructed participants that the goal of this session was to learn about how much they enjoyed the task, since they would later have an opportunity to complete additional rounds of that task for extra pay.

Our goal was to investigate how initial expectations altered subsequent evaluations. To this end, we randomly assigned participants into three groups: known assignment ($n = 292$), coin-flip assignment ($n = 294$), and high-probability assignment ($n = 300$). Participants in the known-

---

19 We used digital-voice software to “read” reviews collected from Amazon.com. Reviews were edited to last approximately 20 seconds, to remove any specific references to author names or book titles, and for grammar. Unbeknownst to participants, all reviews were either 1-star reviews or 5-star reviews to make the task straightforward (though tedious). See the Appendix for sample text of the reviews.

20 We ran a small pilot ($n = 12$) with reduced stakes (show-up fee of $1.50) to check the programming and to verify that participants in the noise and no-noise groups could both successfully complete the task. All participants in that pilot successfully completed the task, regardless of whether they faced the noise or not. In our main sample, we find no difference in ability to complete the work depending on task assignment.

21 Given this feature of the reviews, we may have helped participants answer correctly by withholding the response buttons. In our data, patient responders tend to be more accurate. However, there were very few mistakes overall: only two participants were dropped for inaccurate classifications.
assignment group were told from the start which task they would face, while participants in the coin-flip and high-probability groups were initially uncertain. We call these groups control, coin flip, and high probability, respectively. Next we describe Session 1 for each group.

Participants in the control treatment were randomly assigned—unbeknown to them—to one of two subgroups prior to entering the experiment: noise or no noise. Participants in the control + noise group were told that they would hear audio reviews with an annoying noise overlaid. Participants in the control + no noise group completed classifications without the overlaid noise. Participants in each subgroup were not aware of the possibility of facing the alternate task—they were only told about the one they were assigned. Each participant completed eight mandatory trials of their assigned classification task to conclude the first session.22

In contrast, participants in the coin-flip treatment were told that they faced a 1/2 chance of doing the task without noise and a 1/2 chance of doing the task with noise. They were then given a sample task (without noise) and a short sample of the unpleasant noise (8s in duration; repeatable if desired). After the description and samples, each participant “flipped” a digital coin to determine whether she would ultimately face the task with noise or without. Each participant then completed the eight mandatory classifications prescribed by the result of their coin flip.

Lastly, participants in the high-probability treatment were told that they were very likely to face a given task (either noise or no noise). Half of participants were assigned to a “$p = .99$” treatment and the other half were assigned to a “$p = .01$” treatment, where $p$ corresponds to the probability of facing the task with noise. For each participant, we uniformly drew a random integer $z$ from

---

22Prior to completing mandatory work, participants in each subgroup completed one practice trial (which matched their assigned version of the task) to teach them how to use the interface.
Participants in the \( p = .99 \) arm were assigned the task without noise if \( z = 100 \); otherwise, they faced the task with noise. Participants in the \( p = .01 \) treatment were assigned the task with noise if \( z = 100 \); otherwise, they faced the task without noise. As in the groups above, each participant completed eight trials of their assigned task.

Session 2: Willingness to Work. In each group, the first session concluded after the participant completed the eight mandatory trials of her assigned task. We emailed each participant a link to the second session exactly eight hours after they finished the first.\(^{23} \) In the second session, participants were reminded of their prior task assignment (noise or no noise) and given the option to complete additional trials (of that same task) for a bonus payment. We elicited participants’ willingness to continue working in exchange for five different payment values. We utilized the Becker-Degroot-Marshak (BDM) mechanism to incentivize their responses. The mechanism operated as follows: for each possible bonus payment \( m \in \{0.50, 1.00, 1.50, 2.00, 2.50 \} \), we asked participants the maximum number of tasks they would complete in order to receive \( $m \). They responded by using a slider to select any integer \( e \in [0, 100] \), which we call “willingness to work”. We then (uniformly) drew a random integer \( y \in [0, 100] \). If \( y \leq e \), then the participant completed \( e \) additional tasks and received \( $m \). If \( y > e \), then the participant completed no additional tasks and earned no bonus pay. Conditional on a participant’s assigned task, the second session was identical across all treatment groups. Hence, the key difference across treatments was simply the different ex-ante likelihoods of being assigned the noisy task.

3.2 Theoretical Predictions

In this section, we derive theoretical predictions for how a participant’s willingness to work might depend on the ex-ante probability of being assigned the noisy task. Our model of misattribution predicts that the participant’s willingness to work is increasing in this probability. In contrast, rational learning—with or without reference dependence—predicts that willingness to work is independent of this probability. To clearly illustrate our predictions, we make several simplifying assumptions below. Many of these are unnecessary, and we conclude this section with a discussion of the key assumptions and robustness. While this analysis motivates our empirical strategy, the eager reader may skip to the experimental results (Section 3.3).

Setup. We consider a participant who is uncertain about her cost function associated with her assigned task, and who updates her perception of this function based on her work experience. Mirroring our experimental design, there are two periods. In the first period \((t = 1)\), participant \( i \)

\(^{23}\)Fourteen subjects emailed the authors stating that they had not received an invitation to the second session after more than eight hours. All were sent an additional invitation and completed the second session. We suspect that others may have faced the same issue (due to emails getting caught in spam filters or participants providing an old or invalid email address), leading to slightly higher attrition than desirable. Nevertheless, more than 90% of participants returned for the second session.
is randomly assigned to one of two tasks \( a \in \{ h, l \} \), where \( h \) is the noisy task and \( l \) is the noiseless one. Let probability \( p_i \in \{ 0, .01, .5, .99, 1 \} \) denote the participant’s ex ante chance that she will be assigned to task \( a = h \). Participant \( i \) completes 8 trials of her assigned task \( a \) in period 1 and is informed that she will face this same task with certainty in period 2. In the second period \( (t = 2) \), the participant chooses the maximum number of trials of task \( a \) she is willing to complete in exchange for a monetary payment \( m > 0 \) (incentivized via the BDM mechanism).

Along the effort dimension, we assume participant \( i \)’s consumption utility from completing \( e_{i,t} \geq 0 \) rounds of task \( a \) in period \( t \) is

\[
v_{i,t}^e = -[\theta_i(a) + \varepsilon_{i,t}]c(e_{i,t}),
\]

where \( c(\cdot) \) is an increasing function with \( c(0) = 0 \), \( \theta_i(a) \) is a cost parameter that depends on the task \( a \in \{ h, l \} \), and \( \varepsilon_{i,t} \) are i.i.d. mean-zero random cost shocks that are independent of \( \theta_i(a) \). We assume participant \( i \) knows that \( v_{i,t}^e \) has the structure presented in Equation 5 and knows \( c(\cdot) \). However, she is initially uncertain about the cost parameter, \( \theta_i(a) \). Let \( \pi_{i,0}(a) \) denote her prior over \( \theta_i(a) \). We assume that the participant correctly anticipates that \( \theta_i(h) > \theta_i(l) > 0 \)—i.e., the noisy task is more onerous than the noiseless one—and that her prior is independent of her treatment group—i.e., \( \pi_{i,0}(a) \) is independent of \( p_i \).

**Belief Updating.** Since the participant must decide how much to work on task \( a \) in period 2, she seeks to learn about her cost parameter \( \theta_i(a) \) based on her experience working in period 1. We assume the participant cannot separately observe \( \theta_i(a) \) and \( \varepsilon_{i,1} \), so she uses her experienced utility in period 1 as a signal to update her beliefs about \( \theta_i(a) \). Importantly, this experienced utility may depend on the participant’s initial expectations due to reference dependence.

To describe how reference dependence may influence the participant’s experienced utility, we must fully specify her reference point. Because she is assigned task \( a = h \) with probability \( p_i \), the participant’s expected consumption value on the effort dimension entering period 1 (and thus her reference point) is

\[
\hat{E}_{i,0}[V_{i,1}^e] = -[p_i\hat{\theta}_{i,0}(h) + (1 - p_i)\hat{\theta}_{i,0}(l)]c(h),
\]

where \( \hat{\theta}_{i,0}(a) \) denotes the expected value of \( \theta_i(a) \) under her prior. This follows from Equation 5 along with the fact that each participant must complete exactly eight rounds of the task in period 1, so \( e_{i,1} = 8 \). Participant \( i \)’s total experienced utility in period 1 (Equation 4) is thus

\[
u_{i,1} = v_{i,1}^e + \eta n \left( v_{i,1}^e \left| \hat{E}_{i,0}[V_{i,1}^e] \right. \right).
\]

Let \( \hat{v}_{i,1}^e \) denote the participant’s estimated utility of effort inferred from \( u_{i,1} \). As described in Section 2, a misattributor distorts \( \hat{v}_{i,1}^e \) away from the true value, \( v_{i,1}^e \), according to Equation 3. In particular,
if the task is less burdensome than expected (i.e., \( v_{i,1}^e > \hat{\mathbb{E}}_{i,0}[V_{i,1}] \)), then she underestimates the effort cost (i.e., \( \hat{v}_{i,1}^e > v_{i,1}^e \)). If instead the task is worse than expected (i.e., \( v_{i,1}^e < \hat{\mathbb{E}}_{i,0}[V_{i,1}] \)), then she overestimates the effort cost. In contrast, a rational agent who fully appreciates the extent to which her utility depends on expectations (i.e., \( \hat{\eta} = \eta \)) encodes the correct value, \( \hat{v}_{i,1}^e = v_{i,1}^e \).

We now describe how the participant forms her updated expectation of \( \theta_1(a) \)—denoted by \( \hat{\theta}_{i,1}(a) \)—based on her estimated period-1 (dis)utility of effort, \( \hat{v}_{i,1}^e \). To ease exposition and to provide simple closed-form solutions, suppose that the participant’s prior over \( \theta_1(a) \) is \( N(\hat{\theta}_{i,0}(a), \rho^2) \) and that \( \epsilon_{i,t} \sim N(0, \sigma^2) \). Then

\[
\hat{\theta}_{i,1}(a) = -\alpha \left( \frac{\hat{v}_{i,1}^e}{c(8)} \right) + (1 - \alpha) \hat{\theta}_{i,0} \quad \text{where} \quad \alpha \equiv \frac{\rho^2}{\rho^2 + \sigma^2}. \tag{7}
\]

Hence, the updated expectation has a simple negative linear relationship with \( \hat{v}_{i,1}^e \).25

**Effort Choice.** In period 2, the participant announces how many additional tasks she is willing to do for a bonus payment of \( m \) dollars. Our main hypotheses concern whether and how this willingness to work depends on the likelihood that the participant was assigned to the noisy task, \( p_i \). This likelihood is irrelevant in the rational model given that the participant was told well in advance that her period-2 task will exactly match her period-1 task. Under misattribution, however, sensations of elation or disappointment experienced in period 1 are wrongly attributed to the underlying task, and these sensations of surprise depend on the chance of facing the noisy task, \( p_i \). To allow for such an effect, let \( e_i^+(a|p_i) \) denote participant \( i \)'s willingness to work as a function of \( p_i \).

Throughout our analyses in the main text, we assume that the participant’s response, \( e_i^+(a|p_i) \), represents the number of tasks that renders her indifferent between completing \( e_i^+(a|p_i) \) tasks for a payment of \$m\ and not working at all. That is, \( e_i^+(a|p_i) \) is the number of tasks such that her expected total effort cost is equal to \( m \).26 For sake of a complete analysis, we take an alternative

---

25Assuming normal distributions has the advantage of yielding the tractable updating rule above, but it is worth emphasizing that our basic predictions for both experiments hold under much weaker assumptions that are likely met even if the participant does not precisely follow Bayes’ rule. Namely, we require (1) updating is monotonic in the signal: for any two encoded values of consumption utility \( \hat{v}, \hat{v}' \in \mathbb{R}_+ \), \( \hat{\theta}_{i,1}(a|\hat{v}) > \hat{\theta}_{i,1}(a|\hat{v}') \) if and only if \( \hat{v} < \hat{v}' \); and (2) the participant updates in the direction of her signal: if \( \hat{v} > \hat{\mathbb{E}}_{i,0}[V_{i,1}] \), then \( \hat{\theta}_{i,1}(a|\hat{v}) < \hat{\theta}_{i,0}(a) \); if instead \( \hat{v} < \hat{\mathbb{E}}_{i,0}[V_{i,1}] \), then \( \hat{\theta}_{i,1}(a|\hat{v}) > \hat{\theta}_{i,0}(a) \). That is, when effort is less onerous than expected, beliefs about \( \theta_i(a) \) revise downward; otherwise, they revise upward. Both assumptions—monotonicity and updating in the direction of the signal—are implied by Bayesian updating for a range of distributional assumptions, including the case where \( \theta_1(a) \) and \( \epsilon_{i,t}(a) \) are independent and normally distributed. See Chambers and Healy (2012) for general sufficient conditions for Bayesian updating in the direction of the signal.

26This effort level is the optimal response to the BDM mechanism when the participant has a “standard” utility function that does not exhibit expectations-based reference dependence. However, given that the BDM mechanism creates additional uncertainty over how much the participant will eventually work, the mechanism can conceivably alter the optimal response of a participant with reference dependence who incorporates this BDM-specific uncertainty into her reference point. Despite this caveat, we assume the participant’s effort choice \( e_i^+(a|p_i) \) ignores the uncertainty induced by the BDM mechanism. This approach simplifies the exposition, and it is most consistent with the wording of our survey, which asked participants to truthfully report the maximum number of tasks they are willing to do for
approach in Appendix B, where we assume that a participant with reference dependence incorporates the uncertainty induced by the BDM itself into her reference points along the effort and money dimensions and best responds accordingly. Importantly, either approach gives rise to the same key predictions: under misattribution, \( e^*_i(a|p_i) \) is increasing in the participant’s expectations over her initial task assignment, \( p_i \); under rational updating \( e^*_i(a|p_i) \) is independent of \( p_i \). We flesh out these predictions next.

We first consider predictions under rational learning (i.e., no misattribution). To build intuition using the simplest case, suppose the participant’s utility is not reference dependent (i.e., \( \eta = 0 \)). As described above, the participant chooses effort \( e^*_i(a|p_i) \) so that she is indifferent between completing \( e^*_i(a|p_i) \) tasks for \( m \) dollars and not working at all; hence \( e^*_i(a|p_i) \) solves

\[
\tilde{E}_{i,1} [u_{i,2} | e_{i,2}] = \tilde{E}_{i,1} [V^e_{i,2}] + m = 0,
\]

where \( V^e_{i,2} = \{\theta_i(a) + \varepsilon_{i,2}\} c(e_{i,2}) \) and thus \( \tilde{E}_{i,1} [V^e_{i,2}] = -\tilde{\theta}_{i,1}(a) c(e_{i,2}) \). Condition 8 implies that \( e^*_i \) solves

\[
\tilde{\theta}_{i,1}(a) c(e^*_i) = m, \tag{9}
\]

and \( e^*_i(a|p_i) \) is therefore a decreasing function of her expected cost parameter, \( \tilde{\theta}_{i,1}(a) \). Unlike a misattributing agent, a rational agent’s updated expectation, \( \tilde{\theta}_{i,1}(a) \), is independent of \( p_i \) since her experienced utility, and thus her encoded signal, \( V^e_{i,1} \), do not depend on \( p_i \). Clearly \( e^*_i(a|p_i) \) is therefore independent of \( p_i \).

This result holds for a rational agent with reference-dependent preferences as well. In this case, indifference between completing \( e^*_i(a|p_i) \) tasks for \( m \) dollars and not working at all implies that \( e^*_i(a|p_i) \) solves

\[
\tilde{E}_{i,1} [u_{i,2} | e_{i,2}] = \tilde{E}_{i,1} [V^e_{i,2}] + \eta \tilde{E}_{i,1} \left[ n \left( V^e_{i,2} \big| \tilde{E}_{i,1} [V^e_{i,2}] \right) \right] + m = 0. \tag{10}
\]

Building on Equation 10, we show in Appendix A that \( e^*_i(a|p_i) \) solves

\[
h \left( \tilde{\theta}_{i,1}(a) \right) c(e^*_i) = m, \tag{11}
\]

where \( h(\cdot) \) is an increasing function of \( \tilde{\theta}_{i,1}(a) \) that depends on the participant’s preference parameters \( (\eta, \lambda) \) and her subjective distribution of \( V_{i,2} \), but is independent of \( p_i \). While the condition characterizing \( e^*_i(a|p_i) \) is now more complicated due to reference dependence, the punchline from above remains: the agent properly accounts for how \( p_i \) influenced her experienced utility, and thus
it does not distort her beliefs about \( \theta_i(a) \). Therefore, as in the case absent reference dependence, \( p_i \) does not influence \( e_i^*(a|p_i) \). To summarize:

**Observation 1. Rational Learning with or without Reference-Dependent Preferences.** Let \( e^*(a|p) \) denote the average effort choice among participants who face task \( a \) and held prior beliefs that there was chance \( p \) of facing the noisy task. Both the classical and reference-dependent model without misattribution predict that average effort is independent of \( p \): \( e^*(a|p) = e^*(a|p') \) for all \( p, p' \).

Note that Observation 1 does not say that a participant behaves the same with or without reference-dependent preferences. Rather, it says that—regardless of the underlying preferences—behavior should not depend on the prior probability of facing each task. 27

We now describe \( e_i^*(a|p_i) \) under misattribution. As in the case above, \( e_i^*(a|p_i) \) solves Equation 11. However, the misattributor makes this choice based on her biased assessment of \( \theta_i(a) \). Since she wrongly attributes sensations of elation or disappointment to \( \theta_i(a) \), the misattributor errs when inferring her disutility of effort, \( v_{i,1}^e \), from her experienced utility in period 1. In particular, she encodes an overly optimistic signal \( \hat{v}_{i,1}^e \) whenever the true signal \( v_{i,1}^e \) beats expectations, and she encodes an overly pessimistic signal whenever the true signal falls short of expectations. Thus, fixing the outcome, raising initial expectations leads to a more pessimistic view of the underlying task, and lowering expectations leads to a rosier view of the underlying task. We therefore predict that for each option \( a \in \{h, l\} \), participants’ average willingness to work, \( e^*(a|p) \), is increasing in \( p \).

To illustrate more concretely, consider participant \( i \) who faces a chance \( p > 0 \) of being assigned the noisy task. For sake of example, assume she initially holds “unbiased” priors about the cost of effort—i.e., \( \hat{\theta}_{i,0}(a) = \theta_i(a) \) for both \( a \in \{h, l\} \). Hence, before realizing her assigned task, she expects a (dis)utility of effort in period 1 equal to \( \mathbb{E}_{i,0}[V_{i,1}^e] = -[p \theta_i(h) + (1 - p) \theta_i(l)]c(8) \). Suppose the participant is assigned the no-noise task and \( \epsilon_{i,1} = 0 \). Her total utility in period 1 is then \( u_{i,1} = v_{i,1}^e + \eta n \left( v_{i,1}^e \mid \hat{\mathbb{E}}_{i,0}[V_{i,1}^e] \right) \) where \( v_{i,1}^e = -\theta_i(l)c(8) \). Since \( \theta_i(l) < p \theta_i(h) + (1 - p) \theta_i(l) \), it follows that this outcome beats expectations and her gain-loss utility is positive. Thus, \( u_{i,1} > v_{i,1}^e \)—the participant experiences a utility higher than the intrinsic consumption utility associated with the task—and misattribution leads her to think that this consumption utility was higher than it actually was (i.e., \( v_{i,1}^e > v_{i,1}^e \)). As such, she wrongly infers that the noiseless task is less onerous than it really is and forms an inappropriately low estimate of its cost parameter, \( \hat{\theta}_{i,1}(l) < \theta_i(l) \). Furthermore, if \( p \) is larger—the noisy task is more likely—then the noiseless task generates greater elation and

---

27Note that Observation 1 does not require that participants form rational beliefs per se. We use “rational” simply to connote an agent who does not suffer attribution bias. The predictions in Observation 1 continue to hold if participants suffer other biases in belief updating so long as these biases are independent of treatment assignment—and hence independent of \( p_i \).
the misattributor’s estimate of $\theta_i(l)$ is biased downward by more. The converse is true if the misattributor were instead assigned the noisy task: her estimate of $\theta_i(h)$ is biased upward by more when the noisy task comes as a greater disappointment; that is, the lower is $p$. To summarize:

**Observation 2. Learning with Misattribution of Reference Dependence.** Let $e^*(a|p)$ denote the average utility-maximizing effort choice among participants who face task $a$ and held prior beliefs that there was chance $p$ of facing the noisy task. Suppose $\hat{\eta} < \eta$ and each participant’s prior beliefs over $\theta(a)$ are independent of treatment with $\hat{\theta}_i,0(l) < \hat{\theta}_i,0(h)$. Then average effort $e^*(a|p)$ is increasing in $p$ for each $a \in \{h,l\}$.

The two observations together highlight our empirical strategy. For a given task, we compare the willingness to work of participants across the different assignment probabilities. Observation 2 highlights that, conditional on the assigned task, willingness to work under misattribution is increasing in the ex-ante likelihood of facing the onerous task.

**Discussion of Assumptions**

We now discuss some of the assumptions underlying the results above. First, we clarify the extent to which they rely on a participant’s reference point adjusting between the two sessions. Second, we discuss robustness to participants’ prior beliefs and highlight the relationship between these priors and the motivation behind our high-probability treatment.

**Adjustment of the Reference Point Across Periods.** The observations above assume that the participant’s reference point did not adapt between the coin flip and her initial work. In assuming this, we leverage the fact that the participant began working immediately after the coin flip and thus there was almost no time for a reference point to adapt. We also assume that the ex ante chance of facing the noisy task no longer influenced expectations in the second session, which came more than 8 hours after the resolution of the gamble. While this assumption generates crisp distinctions between effort under misattribution and rational learning (with or without reference dependent preferences), reference points that adapt very slowly can muddy these distinctions. In particular, if participants had sluggish reference points (i.e., expectations still depend on the lottery hours later) and held reference-dependent utility over effort but not money, then reference dependence without misattribution may be able to predict effort patterns similar to those predicted by our model of misattribution. We find this particular constellation of assumptions unlikely; moreover, it is inconsistent with existing evidence demonstrating reference-dependent preferences over money.

We believe our design effectively shuts down this critique. We utilized a relatively long gap between sessions to help ensure that reference points adapted by the time Session 2 began. The evidence to date supports the idea that reference points adapt over modest time periods (and indeed informed our experimental design). As mentioned previously, Song (2016) demonstrates that refer-
ence points incorporate new information over the course of approximately ten minutes. Likewise, Smith (2012) and Buffat and Senn (2015) both provide evidence of relatively quick reference-point changes in laboratory settings with small stakes. Using field evidence from taxi drivers, Thakral and Tô (2020) note that “earnings in the first four hours [of a driver’s shift] have little or no effect on the decision of whether to end a shift at 8.5 hours.” Taken together, we share Song’s (2016) interpretation of the broader literature: for small stakes, reference points seem to adjust on the scale of tens of minutes to a few hours.

Robustness to Poorly-Calibrated Priors. The observations above do not require subjects to have held well-calibrated priors about the tasks (i.e., about $\theta_i(a)$). If prior beliefs were biased on average, our observations continue to hold so long as participants’ priors did not systematically vary across treatment groups. In this case, fixing the task a participant faced, rational learning would lead to the same posterior beliefs regardless of the treatment, since the treatment does not influence the interpretation of signals nor priors. In contrast, misattribution creates an interaction between poorly calibrated priors and the treatment. The predictions from Observation 2 still hold, however, so long as priors were “reasonable”—specifically, so long as participants believed the noisy task was more onerous than the noiseless one. Given that participants (outside the control treatment) sampled each task during the instructions, such priors seem likely.

Priors Independent of Treatment-Group Assignment. The claims above (specifically, Observations 1, 2 and robustness to poorly calibrated priors) rely on independence between a participant’s priors about the tasks and her treatment group (i.e., the likelihood she is assigned the noisy task). However, it is plausible that participants in the coin flip treatment—who were exposed to both tasks during the instructions—formed initial beliefs about a given task that systematically differed from participants in the control group—who were exposed to only a single task. For instance, the existence of both an easy and hard version of the task might have led a participant in the coin-flip group to infer that the noisy task is particularly onerous, while a participant in the control group was only aware of the noisy task and might have expected it to resemble a “typical” MTurk task. This would

---

28 Smith (2012) endowed participants with a lottery to receive a water bottle. Some participants faced a low-probability of winning while others faced a higher chance. Once prizes were awarded, winners revealed their willingness-to-accept (WTA) to sell their bottle, and losers were asked their willingness-to-pay (WTP) for the water bottle. The author highlights that WTA and WTP for the bottle should increase in the probability of winning the water bottle—however, he does not find evidence of such an “attachment effect”. Smith interprets this as evidence that reference points adjust quickly. Buffat and Senn (2015) examine preferences after the resolution of sequential lotteries over money. In that study, all participants faced one of three possible gambles and, after the realization of that gamble, participants stated their WTP for a 50/50 chance to gain CHF 10. In this setting, a slowly-adapting reference point would lead participants to react differently to the three initial gambles—however, the authors find no evidence of this for small stakes. For larger stakes, there is some evidence of a house-money effect, wherein risk attitudes depend on the outcome of the initial gamble.

29 Furthermore, we explore this concern in the empirical analysis below. Recall that subjects could choose when to complete Session 2 so long as they waited at least 8 hours after Session 1. We find no difference between participants who completed Session 2 relatively soon after the mandatory 8-hour waiting period and those who waited longer.
violate our assumption that priors are independent of treatment. This (plausibly) rational-inference story could generate differences in willingness to work that match the predictions of misattribution. Our high-probability treatment addresses this concern: exposure to the two tasks in this treatment exactly matches the coin flip treatment, mitigating concerns about differential inference. In this sense, we use the high-probability group (i.e., participants very likely to face task a) as a less-confounded version of the associated control group (i.e., participants certain to face task a). In both groups, participants strongly expected to face task a, but in the high-probability version they were perfectly aware of the alternative task.

3.3 Results

Guided by the theoretical discussion above, we first utilize a non-parametric analysis which demonstrates that willingness to work in Session 2 strongly depends on participants’ initial expectations regarding their task assignment. We then estimate a reduced-form version of our model and demonstrate that behavior is consistent with participants wrongly learning the underlying difficulty of their assigned task as a function of their priors.

Summary of the Data. Our experimental design generates six subgroups: treatment (i.e. whether participants faced certain assignment, coin-flip assignment, or high-probability assignment) crossed by eventual task assignment (i.e. noise or no noise). For each subgroup, Table 1 shows the demographic characteristics of participants who successfully completed the first session (886 participants in total) and the proportion of those who returned for the second session.30 Note that variability in subgroup sizes resulted from random treatment assignment. Also, while there are some differences in attrition rates across groups (e.g., between the coin-flip + noise and high probability + noise), we discuss below how this pattern is unlikely to drive our results.

We implemented some data-cleaning procedures to form our primary dataset. We removed participants who either (i) did not answer all five elicitations of willingness to work31 (three participants), or (ii) stated a willingness to work equal to the maximum amount (100 tasks) for every payment level, which prevented us from estimating their responsiveness to payment (six participants).32 Additionally, we omit participants who did not return for the second session—and whose

30 There is a significant age difference between the first two treatments and the high-probability treatment. The first two treatments were run approximately 1 month prior to the latter and the high-probability treatment was launched at a slightly later time of day. We suspect time-of-day effects account for the age difference between groups. Our regression analyses control for age and time-of-day effects.

31 This first restriction was the result of coding that should have forced all participants to answer all questions, but did not function properly on some obsolete browsers.

32 Of the six participants dropped due to the latter criterion, three were from control + no noise, two were from coin flip + no noise, and one was from coin flip + noise. We believe these statements likely result from confusion, inattention, or wrongly attempting to manipulate the BDM mechanism. Note that a participant who is supposedly willing to complete 100 tasks for $0.50 is revealing that they command an extremely low hourly wage rate.
willingness to work we therefore did not measure—though we present their demographics where applicable. With this set of restrictions, we are left with a sample of 803 participants.

Nonparametric Analysis

Our main hypothesis was that participants’ willingness to work on a given task would depend on their expectations regarding their task assignment prior to the initial-learning session. As a first step to investigate this hypothesis, we compare the average willingness to work in the control and coin-flip treatments, where we average over both individuals and the five payment levels about which we elicited WTW. This is presented in Columns 1 to 4 of Table 2. This comparison provides a simple assessment of whether uncertainty over task assignment in the initial-learning session affected subsequent behavior. Relative to the control group, participants who faced the noiseless task were willing to work significantly more when their initial impressions were formed after the resolution of the coin flip ($p = .039$ for difference; standard errors clustered at individual level). In contrast, participants who faced the noisy task were willing to work significantly less (relative to control) when their initial impressions were formed after the resolution of the coin flip ($p = .025$ for difference; standard errors clustered at individual level).

While Table 2 gives a rough sense of the treatment effect, we further disaggregate willingness
to work by payment level in Figure 3. Continuing our comparison of the control and coin-flip treatments, the top panel of Figure 3 shows the average willingness to work at each of the five payment levels \{\$0.50, \$1.00, \$1.50, \$2.00, \$2.50\} for each of the four groups (crossing treatment with task assignment). At all payment levels, we find that those who formed initial impressions of the noiseless task when it came as a positive surprise were more willing to work than those who faced the same task with certainty. In contrast, those who formed initial impressions of the noisy task when it came as a negative surprise were less willing to work than those who faced the same task with certainty.

The bottom panel of Figure 3 shows the cumulative distributions of willingness to work in each of the four groups, aggregated over all payment levels (and smoothed using the Epanechnikov kernel). As a validation of our basic setup, a Kolmogorov-Smirnov equality-of-distributions test verifies that control participants were more willing to work on the noiseless task than the noisy one ($D = .1225; p < .001$). Speaking to our main hypotheses, the figure highlights that willingness to work in the control + no noise group was significantly lower than the coin flip + no noise group—the latter almost first-order stochastically dominates the former. By contrast, the cumulative distribution of willingness to work in the control + noise group first-order stochastically dominates that of the coin flip + noise group.

These baseline results reveal economically-meaningful magnitudes. For instance, consider a hypothetical firm seeking workers to complete 25 of our classification tasks. Workers who faced no uncertainty when forming their initial impressions required (on average) $\$1.70 and $\$1.50 to complete 25 noisy and noiseless tasks, respectively. This difference is significantly exaggerated

Footnote: While this test fails to account for redundancy in the data stemming from multiple observations from each individual, we calculated a more conservative version of the statistic by running individual K-S tests for each payment level. Three out of five payment levels showed significant differences between the cumulative distributions of willingness to work for control + noise and control + no-noise; the five p values were .024, .189, .041, .019, .090.
Figure 3: (a) Labor supply curves and (b) cumulative bid distribution by group. Cumulative distribution curves are over all five payment levels and are smoothed using the Epanechnikov kernel.
when workers experience sensations of surprise when forming initial impressions: workers whose initial impressions were confounded by sensations of disappointment or elation required $2.30 and $1.20 to complete 25 noisy and noiseless tasks, respectively. Thus, required payments increased by 35% for the noisy task and decreased by 20% for the no-noise one. Furthermore, the payment premium for the noisy task—the additional payment required to incentivize the noisy task over the noiseless one—increased from $0.20 to $1.10.

We now address three plausible alternative explanations for these baseline results: differential information across treatments, reciprocity toward the experimenter, and attrition.

*Independent Priors Across Treatments.* As discussed above (Section 3.2), the observed differences between the control and coin-flip treatments may reflect differences in information rather than misattribution. Recall that participants in the control treatment were only told about the task they ultimately worked on, while participants in the coin-flip treatment were told about both tasks. This differential exposure to the possible tasks may have generated priors about task difficulty that differed across groups. The high-probability treatment—where participants were exposed to both tasks—helps rule out such concerns. We now compare willingness to work between participants in the coin-flip and high-probability treatments. This allows us to assess the effect of expectations about task assignment while eliminating any potential differences in information between the control and coin-flip treatments.

We first note that participants in the high-probability treatment exhibited a lower willingness to work on the noisy task than on the noiseless task (aggregating across all payment levels and clustering standard errors at the individual level; $p = .064$). This validates that participants perceived a difference in the onerousness of the two tasks.

Turning to a comparison across treatments, Columns 3 to 6 of Table 2 summarizes behavior under the high-probability and coin-flip treatments. As before, we find that participants’ willingness to work depended on the expectations they held prior to the initial-learning session. Participants who were assigned the noiseless task based on the coin flip were, on average, willing to work significantly more than those who strongly expected the noiseless task ($p = .034$ for difference). In contrast, participants who were assigned the noisy task based on the coin flip were willing to work significantly less than those who strongly expected the noisy task ($p = .047$ for difference).

In the comparison above, we use our high-probability treatment as a replacement for the control group in order to equalize information across treatments. However, it is not a direct replacement. Given the (albeit small) uncertainty over task assignment present in the high-probability groups, our model predicts that participants in those groups will demonstrate greater differences in willingness to work across the two tasks than those in the control groups: relative to being assigned the noisy task for sure, the noisy task is slightly more disappointing when expecting a high, but not certain, chance of that task. Thus, fixing the assigned task, our model predicts that the average
willingness to work of those in the high-probability group should fall in between that of the control and coin-flip groups. Indeed, we find suggestive evidence to this effect (see Figure 4) although we are underpowered to properly compare these treatments.\footnote{Furthermore, probability weighting—people’s tendency to overweight small probabilities (e.g. Kahneman and Tversky 1979; Prelec 1998; Gonzalez and Wu 1999)—implies that behavior in the high-probability treatments may substantially deviate from the corresponding control group. Probability weighting would suggest that the 1% chances presented in the high-probability treatment loom much larger than the objective probability. If this is the case, participants may treat high probability as closer to coin flip than is merited by the objective probabilities, hindering our ability to detect differences across these treatments. Thus, although we do find significant differences between the high-probability and coin-flip treatments, the statistical tests are perhaps overstating the likelihood of the null hypothesis being true.}

Differential Attrition Across Treatments. The summary statistics presented in Table 1 suggest that differential attrition—that is, failing to return to the second session—cannot explain our treatment effects. As that table demonstrates, there is not a consistent pattern of attrition between treatments and whether participants were assigned the noisy task. In Table A4 in the appendix, we demonstrate that the observables we collect (e.g., task assignment and demographics) do not predict attrition.\footnote{Two caveats are in order. First, we observe overall lower attrition in the high probability treatment. However, this is not explained by task assignment and we suspect this is due to time-of-day effects, as we ran this treatment at a slightly different time. Second, we note that an alternative type of attrition is possible given the MTurk setting: some participants may have exited the survey when assigned to the noise task without ever completing Session 1. We reviewed all partially completed surveys and found that only nine participants closed the survey prematurely after the}

Figure 4: Labor supply curves across all treatments. Each point represents the average willingness to work for a fixed payment as elicited using the BDM mechanism.
nations are unlikely to explain the observed effects.

Reciprocity. Our findings could plausibly result from reciprocity toward the experimenter: a positive surprise encourages participants to work hard to reward the experimenter, yet a negative surprise leads participants to punish the experimenter through low effort. However, such an explanation requires a set of assumptions that is, in fact, similar to our notion of misattribution. Specifically, fixing the outcome received, the ex-ante probability of receiving that outcome must alter the degree to which a person feels reciprocity towards the experimenter. This explanation furthermore requires that the participant continued to feel positively toward the experimenter more than eight hours later. Given the relatively small stakes involved in this experiment, we suspect this is an unlikely explanation but such probability-dependent reciprocity is not directly ruled out by our design.

Transient Moods. Finally, we note that our experiment was designed with the concern that short-term “transient moods” induced by resolving uncertainty (e.g., anger) might explain our effects. Specifically, the time gap between participants forming their impressions and our elicitation of willingness to work helps distinguish our effect from that of short-term moods. We suspect the coin flip doesn’t influence a participant’s mood more than eight hours later. Furthermore, in Supplemental Tables A1 and A2 we reproduce Table 2 but divide the sample in two: those who returned after more than the median amount of time between Sessions 1 and 2, and those who returned after less than the median return time. We find qualitatively similar results, though our statistical power is greatly diminished.

Parametric Analysis

Motivated by our simple nonparametric results, we now consider a more structured, regression-based approach. We follow the model outlined and discussed in Section 3.2. This allows us to properly account for the fact that effort costs in our experiment may be non-linear. Thus, we provide better estimates of the aggregate effort-supply curves illustrated in Figure 3 with the appropriate confidence intervals and, in effect, address the lack of error bars in the previous figures.

Following the learning model in Section 3.2, we estimate participants’ revealed perception of the underlying cost parameters for each task, $\theta(a)$, conditional on their treatment. For participant $i$ who expected to face the noisy task with probability $p \in \{0, .01, .5, .99, 1\}$ and is ultimately assigned task $a$, let $\hat{\theta}_{i,1}(a|p)$ denote her expectation of $\theta_i(a)$ following Session 1. We estimate the task assignment was revealed. Of those partial-completions, six were assigned to the no-noise task and three were assigned to the noisy task.

36Mood effects stemming from weather and sports outcomes, for example, have been shown to influence investor sentiment (Saunders 1993; Hirshleifer and Shumway 2003; Edmans, Garcia, and Norli 2007).

37This result also suggests that slowly-adapting reference points are likely not the explanation for the behavior we observe.

26
average value of this expectation, denoted $\hat{\theta}_1(a|p)$, among participants in each subgroup; that is, for each relevant combination of ex-ante probability of task assignment, $p$, and assigned task, $a$.

In order to estimate these parameters, we assume $c(e) = (e + \omega)^\gamma$, where $\omega$ is a Stone-Geary background parameter. Given this functional form, identification of the common cost function and the relevant parameter $\theta(a)$ is straightforward. Thus participant $i$ chooses $e_i^*$ such that $\hat{\theta}_1(a|p)(e_i^* + \omega)^\gamma = m$. Rearranging, setting $\omega = 0$, and taking logs yields

$$\log(e_i^*) = \frac{\log(m)}{\gamma} - \frac{\log(\hat{\theta}_1(a|p))}{\gamma}. \tag{12}$$

Assuming an additive error structure, Equation 12 suggests the following regression model:

$$\log(e_i) = \beta_0 \log(m) + \sum_{j=1}^{6} \beta_j (D_i(treatment) \times I_i(noise)) + \delta_i, \tag{13}$$

where $D_i(treatment)$ is a dummy variable indicating when person $i$ was in a particular treatment (control, coin flip, or high probability) and $I_i(noise)$ is an indicator variable designating whether that person ultimately faced the task with or without noise. Variation in payouts, $m$, delivers identification of the curvature parameter, $\gamma$, and variation in treatment assignment crossed with the task the participant ultimately completed delivers identification of $\hat{\theta}_1(a|p)$. Thus mapping Equation 12 onto our econometric specification, we find the parameters of interest are $\gamma = \frac{1}{\hat{\beta}_1}$ and $\hat{\theta}_1(a|p) = \exp\left(-\frac{\beta_1}{\hat{\beta}_0}\right)$. For example, in order to estimate aggregate beliefs of participants in the control + noise subgroup—$\hat{\theta}_1(h|p = 1)$—we combine the coefficient on $D_i(control)I_i(noise)$ with the coefficient on $\log(m)$ as prescribed above.

In Table 3, we present the results of two-limit Tobit regressions with random effects at the individual level, where standard errors are computed using the delta method. This estimation technique is appropriate given that (i) observed willingness to work is censored at a minimum value of 0 tasks and a maximum value of 100, and (ii) we have five observations for each person. Column (1) presents the estimates of the baseline specification in Equation 13. First, we estimate the cost-curvature parameter to be $\gamma = 1.207 (0.023)$; we can accordingly reject a linear cost

---

38 This functional form has been utilized in similar real effort experiments (e.g., Augenblick, Niederle, and Sprenger 2015). For the analysis presented below, we take $\omega = 0$. Although our numerical estimates of the underlying parameters are sensitive to this assumption, our qualitative results are robust. Over a wide range of $\omega$, we estimate significant differences in parameters across our treatments. We present this analysis in Table A3.

39 This effort choice follows from Equation 11, which predicts that a participant chooses $e_i^*$ such that $h(\hat{\theta}_1(a))c(e_i^*) = m$. Thus, our estimates of $\hat{\theta}_1(a|p)$ are technically estimates of $h(\hat{\theta}_1(a|p))$. We drop the $h$ notation going forward to simplify exposition. In Appendix A, we present a closed form solution for $h(\cdot)$ under specific distributional assumptions (e.g., normal priors and noise; see Equation 21). This yields a linear structure, which we implicitly utilize to interpret our results: differences in average estimates of $h(\hat{\theta}_1(a|p))$ across treatments are directly proportional to differences in average expectations across treatments.
function despite the linear appearance of the aggregate data in Figure 3.\textsuperscript{40}

Table 3, Column (1) demonstrates our main result: willingness to work—and accordingly our estimate of perceived effort costs—are shaped by participants’ prior expectations over task assignment. For ease of interpretation, the rows of Table 3 (after the first) are ordered to match the ranking of cost perceptions predicted by our model of misattribution. Participants whose task assignment was determined by coin flip acted as if they formed the most extreme views of the underlying difficulty of the task. Specifically, when participants formed their initial impressions immediately after an unfavorable coin flip, they acted as if they formed more pessimistic views of the underlying task (i.e., of $\hat{\theta}(l)$) than those who faced near-certain task assignment ($\hat{\theta}(h|0.99) - \hat{\theta}(h|0.5) = 0.0142$; $\chi^2(1) = 4.13, p = .042$) or faced no uncertainty prior to task assignment ($\hat{\theta}(h|1) = 0.0149; \chi^2(1) = 4.27, p = .039$). Conversely, when participants formed their initial impressions after a favorable coin flip, they acted as if they formed more optimistic views of the underlying task than those who faced near-certain task assignment ($\hat{\theta}(l|0.01) - \hat{\theta}(l|0.5) = -.0087; \chi^2(1) = 4.06, p = .044$) or faced no uncertainty prior to task assignment ($\hat{\theta}(l|0) - \hat{\theta}(l|0) = -.0064; \chi^2(1) = 2.49, p = .115$).

For robustness, Column (2) of Table 3 controls for demographic characteristics (age, gender, and income) and for the time spent completing the first session, which we view as a coarse proxy for subjective task difficulty. Finally, Column (3) drops participants whose responses were not weakly monotonic in payment—that is, their willingness to work did not weakly increase across all five payment levels. This drops a significant portion of the sample, but the point estimates of our effect remain similar.\textsuperscript{41}

Perhaps most notable from Table 3 is that the ordering of parameter estimates closely matches the predictions of our misattribution model. Indeed, the hypothesis that $\hat{\theta}(a)$ does not depend on $p$ is rejected ($\chi^2(4) = 9.88, p = .043$). Given our non-parametric results in combination with these structural estimates, we conclude that manipulating prior expectations had a significant effect on subsequent willingness to work in a pattern that is consistent with attribution bias of reference-dependent utility.

\textsuperscript{40}As a form of robustness check, we estimated a model that mirrored Column (1) but introduced a more flexible cost function that allowed $\gamma$ to depend on whether the person faced the noise or no-noise task. This did not change the qualitative results. Moreover, in that analysis we fail to reject the null hypothesis $H_0 : \gamma(h) = \gamma(l); \chi^2(1) = 0.24; p = 0.624$.

\textsuperscript{41}Although we observe a seemingly high number of non-monotonic responses, we believe that our response mode (slider) was conducive to small mistakes. 111 responses were non-monotonic—that is, the willingness to work for some higher fixed payment was less than that at a lower fixed payment. The average mistake (that is, the magnitude of the deviation from responses that increase in stakes) was small.
Table 3: Parametric Analysis, Experiment 1

<table>
<thead>
<tr>
<th>Dep var: log (ei) Estimated w/ Random-Effects Tobit Regression</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost curvature parameter, $\gamma$</td>
<td>1.199</td>
<td>1.197</td>
<td>1.159</td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.017)</td>
<td>(.016)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 0.5$)</td>
<td>.0673</td>
<td>.0635</td>
</tr>
<tr>
<td></td>
<td>(.006)</td>
<td>(.0120)</td>
<td>(.007)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 0.99$)</td>
<td>.0531</td>
<td>.0510</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.009)</td>
<td>(.005)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 1$)</td>
<td>.0524</td>
<td>.0493</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.008)</td>
<td>(.006)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0$)</td>
<td>.0408</td>
<td>.0385</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.007)</td>
<td>(.004)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0.01$)</td>
<td>.0431</td>
<td>.0416</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.007)</td>
<td>(.004)</td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0.5$)</td>
<td>.0344</td>
<td>.0325</td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.006)</td>
<td>(.004)</td>
</tr>
</tbody>
</table>

$H_0 : \hat{\theta}_1$ (noise | $p = 0.5$) = $\hat{\theta}_1$ (noise | $p = 0.99$) |

$\chi^2(1) = 4.13$ (p = .042)  $\chi^2(1) = 2.90$ (p = .089)  $\chi^2(1) = 3.92$ (p = .048)

$H_0 : \hat{\theta}_1$ (no noise | $p = 0.5$) = $\hat{\theta}_1$ (no noise | $p = 0.01$) |

$\chi^2(1) = 4.06$ (p = .044)  $\chi^2(1) = 4.77$ (p = .029)  $\chi^2(1) = 2.73$ (p = .098)

Joint test of above $\chi^2(2) = 8.18$ (p = .017)  $\chi^2(2) = 7.47$ (p = .024)  $\chi^2(2) = 6.64$ (p = .036)

Observations | 4020 | 4020 | 3470 |
Clusters | 804 | 804 | 694 |
Demographics and Session 1 Length Controls | No | Yes | No |
Restricted to “Monotonic” Sample | No | No | Yes |

Notes: Recall that $p$ in the left column refers to the ex ante probability of completing the task with noise. Standard errors (in parentheses) are clustered at the individual level and recovered via delta method. 18 observations are left-censored and 43 are right-censored in the main sample; 11 are left-censored and 43 are right-censored in the “monotonic” sample.
Finally, we note that the results above demonstrate a large and economically significant effect of expectations over task assignment on our estimates of perceived effort costs. For example, we estimate a roughly 20 percent difference in perceived effort costs between those participants facing the noisy task in the coin-flip treatment and those participants who ultimately faced the same task but began in the high-probability treatment. This finding mirrors our earlier non-parametric results.

4 Experiment 2

In this section, we present our within-subject experiment, which was conducted at the Harvard Decision Science Lab. We first describe the design, highlighting how the approach allows us to rule out any interaction between treatment and priors that may have taken place in Experiment 1. We then discuss our theoretical predictions, which are similar to those from Experiment 1 when applied to a within-subject setting. Finally, we analyze the experimental data. Experiment 2 yields similar conclusions to Experiment 1, but extends our findings to a different experimental population, albeit with a greatly reduced sample size. Importantly, this design allows us to (noisily) estimate within-subject measures of misattribution, an exercise not possible with Experiment 1.

4.1 Design

We recruited participants from the Harvard student body for a two-session experiment, with sessions separated by a week. A total of eighteen sessions (nine groups) were conducted over the course of one month. Our primary sample consists of 87 subjects. Participants were paid $7 for successfully completing each of two sessions in addition to any earnings from their choices. In order to prevent attrition, we paid participants contingent on completion of both sessions.

Before specifying the details of Experiment 2, we first provide a broad overview of how the design differs from Experiment 1. In the first session, each participant was assigned via coin flip to work on one of two tasks. Each participant then returned one week later to work on that same task in a second session. Thus, participants faced uncertainty over their task assignment in the first session, but not in the second. To ensure that participants did not perceive any uncertainty when entering the second session, we instructed them ahead of time that their coin flip in the first session would apply to both sessions, and we sent them an email reminder of their coin-flip outcome approximately two days before their second session.

42Ex-ante power tests suggested that \( n \approx 100 \) would provide 80% power, assuming a modest effect size. We under-recruited because our sampling window coincided with the end of the academic school year. Additionally, two of the groups that we recruited later in the sampling window had higher-than-average attrition, which we suspect was due to final exams. One participant withdrew moments into the first session due to a scheduling conflict; a second withdrew in the middle of the first session because she did not want to take part in the study (and offered no further explanation). These two participants are excluded from all analyses.
We measured participants’ willingness to work in both sessions of the experiment. Assuming participants’ expectations about task assignment change across sessions, then the change in participants’ willingness to work across sessions allows us to identify misattribution. That is, our variable of interest is the difference in a participant’s willingness to work in week one—when her task came as a surprise—and week two—when that same task was expected.

During both sessions, participants worked on a real-effort task similar to that of Augenblick, Niederle, and Sprenger (2015) and Augenblick and Rabin (2019): “transcribing” handwritten Greek and Russian letters. Each trial of the task consisted of a string of 35 handwritten characters; participants “transcribed” each character by clicking the matching letter from an alphabet of the relevant language. See Figure 5 for a screenshot. Participants were randomly divided into one of two language treatments. Half of them transcribed Greek during the first session and Russian during the second, while the other half faced the opposite order. Aside from variation in the language, each session had the same structure: participants first completed an initial-learning phase which consisted of five mandatory trials, and then we elicited their willingness to complete additional trials for a bonus payment.

As in the coin-flip condition of Experiment 1, we presented each participant with two variants of the task—a noisy version and a noiseless one. In both variants, participants wore headphones while completing transcriptions. In the noisy version, the annoying noise played through the headphones (calibrated to roughly 70-75 decibels); the noise was identical to Experiment 1, except it played on loop for the entire transcription time. In the noiseless version, no sound played through the headphones. In order to endow participants with reasonable priors about each task, the initial instructions included an interactive sample of the transcription task, and participants listened to an eight-second sample of the annoying noise (repeatable if desired).

Session 1: Coin Flip and Willingness to Work. Upon entering the experiment, all participants were told that they faced a 1/2 chance of being assigned the noisy task versus the noiseless one. In order to make this probability salient—and to enhance the sensation of surprise or disappointment—each participant flipped a U.S. quarter to determine their assignment. We instructed participants that a flip of heads would result in the noiseless task, while tails would result in the noisy one. After resolving the coin flip, each participant immediately started their initial-learning phase in which they completed five mandatory trials of their assigned task. 44 participants were ultimately assigned the noiseless task, while 43 faced the noisy one.

After completing the initial-learning phase in Session 1, subjects were given the option to complete additional trials for a bonus payment. As in Experiment 1, we asked each participant how

---

43 Although our task mimics that of Augenblick, Niederle, and Sprenger (2015), we used different visual stimuli which ended up being easier to transcribe. Participants in our study needed 40 seconds on average to complete one trial, while participants in the first week of Augenblick, Niederle, and Sprenger’s study needed 54 seconds on average.
Figure 5: Screenshot of the transcription tasks from Experiment 2. Participants clicked the gray button that matched the handwritten letter to “transcribe” the text. Participants were required to achieve 80% accuracy to advance to the next transcription. Each participant randomly faced one language—Greek or Russian (Cyrillic; top and bottom, respectively)—during their first session, and then faced the other language during their second session.

many additional tasks they were willing to complete for each of five payments: \{\$4, \$8, \$12, \$16, \$20\}. Participants responded by using a slider to select any integer \( e \in \{0, \ldots, 100\} \), and we used the BDM mechanism to incentivize these responses.

Session 2: Willingness to Work (2nd Elicitation). Upon returning to the second session of the experiment, each participant first completed five mandatory trials of the same task variant they faced in Session 1 (i.e., noisy or noiseless). After the five mandatory trials, we elicited participants’ willingness to continue working on that task. The experiment concluded after participants completed any additional trials. Subjects were paid only upon completion of both sessions.

As noted above, participants transcribed a different language in the second session. We introduced this minor variation in the task across sessions so that participants could plausibly form different perceptions of the task over time and hence update their willingness to work. This design feature was intended to help reduce anchoring effects: since participants faced a somewhat differ-
ent task in the second session, they may have been less likely to answer exactly the same as they
did during the first session. It also provides subjects with a potential “cover story” for changing
their responses across sessions.

4.2 Theoretical Predictions

We now sketch how our theoretical predictions from Experiment 1 (presented in Section 3.2) ex-
tend to this within-subject design. In contrast to Experiment 1, a participant in this setting receives
two signals about her cost function, and we measure her willingness to work twice—once after
receiving the first signal, and then again after receiving the second. These two signals derive from
the participant’s consumption utility of effort in the initial-learning phase of Sessions 1 and 2. We
focus our analysis on participants who do not complete additional tasks during the first session,
so these two signals are the only information a participant receives about the tasks (aside from the
instructions).

We first describe the predictions of rational learning. Throughout this section, we maintain the
same basic setup and assumptions introduced in the theoretical analysis of Experiment 1, and we
further assume that participants’ priors over the cost parameters are unbiased on average. In this
case—where participants hold reasonable expectations about the difficulty of the tasks—rational
learning without reference-dependent preferences predicts that participants’ willingness to work
will, in aggregate, remain constant across the two periods. In contrast, rational learning with
reference dependence but without misattribution can lead participants to systematically change
their willingness to work across periods. Namely, as we show in detail in Appendix C, reference
dependence absent misattribution creates an incentive for those facing the noisy task to decrease
effort over time, and those facing the noiseless task to increase it.\footnote{These incentives arise if a participant is loss averse and her reference point at the time of her first decision is still based on the expectations she held prior to learning the outcome of the coin flip. If either of these conditions is not met, then reference dependence absent misattribution has no effect on behavior, resulting in effort choices that are, on average, constant across the two periods. If instead both of these conditions hold, then reference dependence can generate systematic changes in effort across periods when the participant forms forward-looking strategies aimed at mitigating losses. By planning to exert similar effort (in terms of cost) in the first period regardless of the outcome of the coin flip, a participant can avoid feeling large sensations of disappointment no matter which task she is assigned. If the participant’s expectations then adapt to her assigned task by the second period, she no longer has incentive to equalize effort across contingencies. Thus, relative to the first period, she will increase her willingness to work if she were assigned the less-costly (noiseless) task and decrease it if she were assigned the more-costly (noisy) task. See Appendix C for further details.}

Misattribution introduces an opposing effect that leads those assigned the noisy task to increase
effort between periods 1 and 2 and those assigned the noiseless task to decrease effort. These
patterns stem from a participant’s reference point evolving between the two periods and, critically,
the fact that first-period outcomes are misencoded. In the first period, her reference point puts a
50% chance on each of the two tasks. We assume that by the second period, the participant fully
anticipates her assigned task and her reference point adapts accordingly. Thus, the participant’s
two experiences with her assigned task—the initial-learning phase at the start of each session—
happen under different reference points. Misattribution will thus cause her to encode these similar
experiences differently.

More formally, suppose that consumption utility takes the same form as the model underlying Experiment 1 (see Equation 5). Thus, consumption utility from each initial-learning phase—in which the participant completes five trials of her assigned task \( a \in \{h, l\} \)—is \( \nu_{i,j}^e = [\theta_i(a) + \varepsilon_{i,t}]c(5) \).\(^{45}\) On average, a participant assigned the noisy task \( a = h \) will encode these values such that \( \hat{\nu}_{i,1}^e < \hat{\nu}_{i,2}^e \). This is because her first signal incorporates a sense of disappointment—in period 1, she anticipates a 50% chance of facing the better task. But her second signal comes with less
disappointment—in period 2, she fully expects the worse task. Put differently, the participant’s first
experience falls short of expectations by a greater amount than the second and is thus remembered
as worse. In contrast, an average participant assigned to the noiseless \( a = l \) task will encode
values such that \( \hat{\nu}_{i,1}^e > \hat{\nu}_{i,2}^e \): the first signal incorporates a sense of elation from the coin flip, but the
second signal comes with less (if any) such elation.

Misattribution in Experiment 2 additionally generates a contrast effect (that was not possible in
the single-decision setting of Experiment 1). To illustrate, consider a participant who is assigned
to the noiseless task. Following the coin flip, this task comes with a pleasant surprise, and a mis-
attributor therefore underestimates its true cost. Since her stated willingness to work in Session
1 is based on this overly-optimistic perception of the underlying effort cost, it is biased upward
relative to the case without misattribution. This follows from the theoretical discussion of Ex-
periment 1. In Experiment 2, however, the participant has a second experience with her assigned
task in the learning phase of Session 2, and this experience tends to come with an unpleasant
surprise: since her prior expectations (stemming from her Session 1 experience) are inflated, her
second experience—now devoid of the positive surprise from the coin flip—will not live up to
those unrealistic expectations. This typically-bad experience pushes her estimated cost upward,
reducing her willingness to work in the second session. If this “contrast effect” between the first
and second rounds is sufficiently strong, then the noiseless participant’s revealed willingness to
work will decrease over the two sessions. Similar logic implies that a misattributor assigned to the
noisy task will increase her effort across sessions: in the first session, the negative surprise from
her unfavorable task assignment leads her to overestimate the disutility of effort. Her experience
with that same task in the second session, however, will typically surpass these overly-pessimistic
expectations. This positive surprise then increases her willingness to work in the second session.

\(^{45}\) Although we model the cost function in a similar way for each experiment, we do not assume it is literally the
same across the two experiments given that the tasks in each are different. As such, we assume value of \( \theta \) and the
parameters governing \( c(\cdot) \) vary across the two experiments.
The dynamic misencoding described above has direct implications for the participant’s perceptions of the task and effort choices across periods. In particular, participants assigned to the noisy task will typically find their task less onerous in period 2 than period 1—that is, on average, \( \hat{\theta}_{i,2}(h) < \hat{\theta}_{i,1}(h) \). In contrast, those assigned the noiseless task will typically find it more onerous in period 2 than period 1—that is, on average, \( \hat{\theta}_{i,2}(l) > \hat{\theta}_{i,1}(l) \).

The behavioral implications of these predictions, however, are met by a countervailing force stemming from rational reference dependence noted above. Accordingly, mapping these predictions to effort choices—the observables in our experiment—is not trivial. Let \( e^*_{i,t}(a) \) denote the predicted effort choice for participant \( i \) facing task \( a \) in period \( t \). If the countervailing force is relatively weak (for instance, if loss aversion is modest) then the distorted beliefs described above will be reflected in effort choices. Thus, with either strong misattribution or sufficiently weak loss aversion, \( e^*_{i,1}(l) > e^*_{i,2}(l) \) and \( e^*_{i,1}(h) < e^*_{i,2}(h) \) on average. These are the main predictions we empirically test.

**Discussion of Predictions.** Our theoretical discussion above relies on unbiased priors on aggregate. If participants’ priors were systematically biased in a specific direction—namely, they significantly overestimate the disutility of the task with noise and underestimate the disutility of the task without noise—then changes in average willingness to work across sessions may result from rational learning. We believe our assumption of correct priors (on average) is justified from the experimental design: participants were exposed to both versions of the task before commencing work, and therefore should have held reasonably well-calibrated priors. Fortunately, as highlighted previously, this limitation does not apply to Experiment 1.

Additionally, our predictions assume that reference points adapted to the assigned task by the beginning of Session 2. This seems warranted given that there was no uncertainty in task assignment in the second session, and participants knew about their task assignment a week in advance. Furthermore, they were reminded by email mid-way through the week. Before beginning Session 2, all participants were required to verbally state which task they had faced in Session 1, and all participants did so successfully. This suggests that the assignment was salient and memorable.

### 4.3 Results

For our primary analysis, we only consider participants who returned to both sessions. Thus, our data comes from 72 participants who completed the experiment across a total of nine different experimental groups. For completeness, we present a (simple) analysis of participant attrition in Table A6.

We first present nonparametric analyses demonstrating that willingness to work systematically changes over time depending on the resolution of the coin flip in Session 1. We then estimate
the parameters of a reduced-form model similar to Experiment 1, but utilizing the within-subject nature of this design. We conclude by demonstrating that our results are robust to informational explanations stemming from those participants who completed additional tasks in the first session.

Nonparametric Analysis

Sessions 1 and 2 of this experiment mirrored the coin-flip and control treatments from Experiment 1. The difference across sessions stemmed from an uncertain reference point in the first session changing to a certain reference point in the second. Following the analysis of Experiment 1, Table 4 presents participants’ average willingness to work—averaged over the five payment levels—in each of these two sessions.

Our design is not intended to detect between-subject differences, and the aggregate results (Columns 1-4 of Table 4) obscure important within-subjects variation. However, Experiment 2 allows us to account for individual differences in overall willingness to work by examining changes in willingness to work across Sessions 1 and 2. We find that willingness to work significantly changes across sessions (see Columns 5-6 of Table 4). Furthermore, consistent with our theoretical predictions, participants who faced the noiseless task tended to decrease their willingness to work across sessions while those assigned the noisy task tended to increase it. When assigned the noiseless task, participants were (on average) willing to complete 7.5 more tasks in Session 1 than in Session 2 ($p = .004$, standard errors clustered at individual level). In contrast, when assigned the noisy task, participants were (on average) willing to complete 4.3 fewer tasks in Session 1 than in Session 2 ($p = .014$, standard errors clustered at the individual level). Figure 6 depicts this result by plotting the density of $e_{i,1} - e_{i,2}$ for each task, averaged over the five payment levels.46

To assess the economic magnitudes of these results, we again consider a hypothetical firm paying workers to complete 25 transcriptions (as done the discussion of Experiment 1, Section 3.3). To incent the average participant to complete 25 noiseless transcriptions, the firm would have to pay $7.75 right after the worker formed her initial impression (i.e., just after the positive outcome of the coin flip); this would increase to $11 once her assessment of the task is no longer confounded with a sense of elation. In contrast, a firm would have to pay $12 to incent the average participant to do 25 noisy transcriptions right after she formed her initial impression (i.e., just after the negative outcome of the coin flip); this would decrease to $10.50 once her assessment of the task is no longer confounded with disappointment. These effect sizes have similar magnitude to those in Experiment 1.47

46We present these densities using kernel smoothing (Epanechnikov kernel) for readability; raw data appears in the Appendix as A1 and unsmoothed histograms appear as Figure A2.

47There are two important caveats to consider before comparing this calibration exercise to the results of Experiment 1. First, because the task in Experiment 2 is more time-consuming than that of Experiment 1, and because the lab subjects are generally paid more, the magnitudes of payments are quite different across experiments. Second, because
Table 4: 
**Baseline Results, Experiment 2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Session 1</th>
<th>Session 2</th>
<th>$(e_{i,1} - e_{i,2})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>noise=0</td>
<td>noise=1</td>
<td>noise=0</td>
</tr>
<tr>
<td>Observations</td>
<td>215</td>
<td>220</td>
<td>180</td>
</tr>
</tbody>
</table>

*Notes: Standard errors (in parentheses) are clustered at the individual level. Differences between Columns (1)-(3) significant at $p = .026$; between Columns (2)-(4): $p = .865$. Columns (5)-(6) both significantly different from zero: $p = .004$ and $p = .014$, respectively.*

**Parametric Analysis**

We now present our quasi-structural estimation. Given that the experiment closely follows the approach from Experiment 1, the decision problem in each session can be modeled in the same way as the previous experiment, and is thus described by the logic in Section 3.2. Mirroring our parametric approach to Experiment 1—and adopting the previous notation—Equation 12 implies that for each period, $\log(e_{i,t}^*) = \log(m) - \log(\hat{\theta}_i(a|p))$. Since we have multiple observations for each individual, we examine the difference $\log(e_{i,1}) - \log(e_{i,2})$. This difference is independent of $m$, thereby eliminating a potential source of (unmodeled) heterogeneity and increasing our statistical power. Our econometric model is thus

$$\left( \log(e_{i,1}) - \log(e_{i,2}) \right) = \beta I_i(noise) + \epsilon_i. \quad (14)$$

From this specification, we can recover aggregate estimates $\frac{\hat{\theta}_1(a|p)}{\hat{\theta}_2(a|p)} = \exp(-\gamma \beta)$. Since the cost-curvature parameter $\gamma$ is not identified in this specification, we separately model the first session (following Equation 13) to generate an in-sample estimate of $\gamma \approx 1.11$; note this estimate falls close to our estimate from Experiment 1.$^{48}$ We use this value (and the equation above) to numerically estimate the ratio of interest.

Following Experiment 1, we estimate Equation 14 using a random-effect Tobit model. The results are shown in Table 5. Note that our structural estimates closely align to those of Experiment 1. In this experiment, we find that $\frac{\hat{\theta}_1(noise)}{\hat{\theta}_2(noise)} = 1.29$ (Column 1 of Table 5). This is very close to the analogous ratio we found in Experiment 1, $\frac{\hat{\theta}(noise|coin \ flip)}{\hat{\theta}(noise|control)} = 1.28$ (Column 1 of Table 3). Likewise the sample size in Experiment 2 is much smaller, the estimated effect size is very imprecise.$^{48}$ As in Experiment 1, we tested whether $\gamma(h) = \gamma(l)$. Using data from the first session, we fail to reject the null $H_0: \gamma(h) = \gamma(l); \chi^2(1) = 0.59, p = 0.442.$

---

$^{48}$
Figure 6: *Kernel density of the difference in willingness-to-work between the first and second sessions, separated by task faced.* Each underlying observation from this figure is the change in a participant’s willingness to work for a fixed payment between Sessions 1 and 2 of the experiment. The black curve represents participants who were assigned to the no-noise task; the red curve represents participants who were assigned to the noisy task.

The ratio \( \hat{\theta}_1(\text{no noise})/\hat{\theta}_2(\text{no noise}) \) = 0.78 falls close to \( \hat{\theta}(\text{no noise}|\text{coin flip})/\hat{\theta}(\text{no noise}|\text{control}) \) = 0.84. In both studies and across all specifications, we find that uncertain assignment via coin flip distorts willingness to work in the range of approximately 17% to 40% relative to certain assignment.
Table 5:
**PARAMETRIC ANALYSIS, Experiment 2**

<table>
<thead>
<tr>
<th>Estimated ratio</th>
<th>Estimated w/ Random-Effects Tobit Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}_1(\text{noise})$</td>
<td>$\hat{\theta}_2(\text{noise})$</td>
</tr>
<tr>
<td>$\hat{\theta}_1(\text{no noise})$</td>
<td>$\hat{\theta}_2(\text{no noise})$</td>
</tr>
</tbody>
</table>

$H_0 : \frac{\hat{\theta}_1(\text{noise})}{\hat{\theta}_2(\text{noise})} \geq 1$

$\chi^2(1) = 5.15$

$p = .023$

$H_0 : \frac{\hat{\theta}_1(\text{no noise})}{\hat{\theta}_2(\text{no noise})} \leq 1$

$\chi^2(1) = 6.37$

$p = .011$

| Observations | 348 |
| Clusters | 70 |

Notes: Standard errors (in parentheses) are clustered at the individual level and derived via the delta method. 12 observations are left-censored and 26 are right-censored. Dropped observations result from taking logs under the assumption that $\omega = 0$. Each estimate $\frac{\hat{\theta}_1(a)}{\hat{\theta}_2(a)}$ is derived assuming that $\gamma = 1.11$.

**Discussion.** A potential concern in this setting is that the participants who completed additional tasks during the first session may have formed different beliefs than those who did not complete additional tasks. One third of participants completed additional tasks in the first session (recall that the BDM mechanism induces randomness in whether a participant will actually complete any additional tasks). Theoretically, comparing participants who completed additional tasks with those who did not complicates the predictions, as the two groups have accumulated different amounts of experience. We explore this empirically in Supplemental Table A5 (Appendix E). There, we demonstrate that the analogous non-parametric results from Table 4 are robust to (a) controlling for extra tasks using OLS; (b) controlling for extra tasks via two-stage least squares (utilizing the BDM randomness for identification); and (c) simply dropping participants who completed extra tasks. While statistical power decreases when dropping participants, our estimates remain similar.

Finally, as with Experiment 1, we suspect attrition is an an unlikely explanation for our results. In
Supplemental Table A6 (Appendix E) we demonstrate that attrition is independent of whether a participant faced noise, their average willingness to work in Session 1, and whether the participant first faced Russian or Greek. All participants who completed extra tasks in Session 1 returned for Session 2.

5 Conclusion

In this paper, we provide evidence consistent with misattribution of reference dependence—that people retrospectively fail to account for their reference-dependent utility when learning about an unfamiliar real-effort task. In a series of experiments, we manipulate participants’ expectations prior to their initial experiences. Consistent with our model (Gagnon-Bartsch and Bushong 2019), we observe systematic and persistent changes in subsequent willingness to work depending on subjects’ initial expectations, despite the fact that this initial reference point is no longer relevant. We now briefly discuss some reasons for caution in interpreting our results as well as directions for future research.

Our model predicts that loss averse participants will form more distorted perceptions of bad outcomes than good ones. In our first experiment, we find weak but suggestive evidence of loss aversion reflected through misattribution: the average willingness to work for those assigned to the noisy task by chance was more distorted than the willingness to work of those assigned to the no-noise task by chance. Although the aggregate results in Experiment 2 do not demonstrate signs of loss aversion, it is possible that we are unable to see loss aversion because of an overall diminished willingness (among all participants) to work in the second session. Additionally, asymmetric distortion of bad outcomes (relative to good outcomes) may be difficult to observe in our paradigm due to compression of the response scales at low values. With low willingness to work, participants may utilize the response scale differently than those with higher willingness to work, which may make detecting loss aversion more difficult. Loosely, choices may be more finely tuned near the bottom of the scale and hence less susceptible to big changes. As loss aversion is central to our theoretical model and drives a number of predictions for long-run beliefs, future work should address the extent to which losses drive asymmetric belief updating.

Future experimental work could explore an additional theoretical prediction of the model: sequential contrast effects. Taking our Experiment 2 design as an example, our model predicts that the disutility of effort in Session 2 is compared against the wrongly-encoded disutility from Session 1. For participants facing the noisy task, this leads to an increase in willingness to work that may “overshoot” the willingness to work of participants who knew all along that they would face the noisy task. In contrast, for those facing the noiseless task, willingness to work may “undershoot” the willingness to work of participants who knew all along they would face the noiseless task. Our
results hint in this direction—e.g., in Table 4 the difference in willingness to work between noise and no noise in the second session is very small.

Our theoretical companion paper describes a number of further avenues for experimental work. For instance, our model predicts that as bad outcomes become less common, a misattributor will perceive those outcomes as worse. In contrast, as good outcomes become less common, a misattributor will perceive those outcomes as better. A straightforward test of this comparative static—extending our existing experimental paradigm—would involve manipulating prior expectations such that participants face a wider range of probabilities over task assignments. While the varied treatments in Experiment 1 provide a first look at the role of probabilistic assignment in subsequent evaluations, future research could explore this in greater detail.

More broadly, our results suggest that organizations (e.g., firms or political parties) can shape short-run impressions by managing expectations. For instance, our results suggest that employees would form more favorable impressions of undesirable tasks if they knew well ahead of time that they would have to complete them. This accords with evidence on firms that give realistic job previews prior to hiring. As Phillips (1998) shows, employees who face a realistic job preview are higher performing and less likely to leave their job than their peers who do not experience a job preview. Misattribution along the lines discussed in this paper may provide an underlying mechanism for this effect.

References


43


Appendix
FOR ONLINE PUBLICATION

A  Experiment 1: Derivation of Optimal Effort

In this appendix we show that, under reasonable assumptions, a rational participant with reference-dependent utility will choose an effort level in Experiment 1 that is decreasing in her expected value of her cost parameter, \( \theta_i(a) \). This analysis formalizes the predictions regarding effort stated in Observations 1 and 2.

Recall that the predicted effort of a participant with reference-dependent utility assigned to task \( a \) solves Equation 10 in the main text: indifference between completing \( e_i^*(a|p_i) \) tasks for \( m \) dollars and not working at all implies that \( e_i^*(a|p_i) \) is the value of \( e_{i,2} \) that solves

\[
\hat{E}_{i,1}[u_{i,2}|e_{i,2}] = m + \hat{E}_{i,1}[V_{i,2}^e] + \eta \hat{E}_{i,1}[n(V_{i,2}^e|\hat{E}_{i,1}[V_{i,2}^e])] = 0
\]

\[
\Rightarrow \hat{E}_{i,1}[u_{i,2}|e_{i,2}] = m - \hat{\theta}_{i,1}(a)c(e_{i,2}) + \eta \hat{E}_{i,1}[n(V_{i,2}^e|\hat{\theta}_{i,1}(a)c(e_{i,2}))] = 0. \tag{15}
\]

Recall that, conditional on \( e_{i,2} \), the participant’s effort cost in period 2 is a random variable \( V_{i,2}^e = -[\theta_i(a) + \epsilon_i]c(e_{i,2}) \). Define the random variable \( X_{i,2}(a) = \theta_i(a) + \epsilon_i \) and let \( \hat{F}_{i,1} \) denote the participant’s subjective CDF over \( X_{i,2} \) conditional on any information obtained in period 1. Let \( x_{i,2} \) denote the realization of \( X_{i,2} \). Furthermore, note that

\[
n(V_{i,2}^e|\hat{\theta}_{i,1}(a)c(e_{i,2})) = -[x_{i,2}(a) - \hat{\theta}_{i,1}(a)]c(e_{i,2}) \text{ if } x_{i,2}(a) \leq \hat{\theta}_{i,1}(a), \text{ and otherwise } n(V_{i,2}^e|\hat{\theta}_{i,1}(a)c(e_{i,2})) = -\lambda [x_{i,2}(a) - \hat{\theta}_{i,1}(a)]c(e_{i,2}). \]

Thus,

\[
\hat{E}_{i,1}[n(V_{i,2}^e|\hat{\theta}_{i,1}(a)c(e_{i,2}))] = -c(e_{i,2}) \left( \hat{F}_{i,1}(\hat{\theta}_{i,1}(a)) \hat{E}_{i,1}[X_{i,2}(a) - \hat{\theta}_{i,1}(a) | X_{i,2}(a) \leq \hat{\theta}_{i,1}(a)] + \lambda [1 - \hat{F}_{i,1}(\hat{\theta}_{i,1}(a))] \hat{E}_{i,1}[X_{i,2}(a) - \hat{\theta}_{i,1}(a) | X_{i,2}(a) > \hat{\theta}_{i,1}(a)] \right), \tag{16}
\]

and thus

\[
\hat{E}_{i,1}[n(V_{i,2}^e|\hat{\theta}_{i,1}(a)c(e_{i,2}))] = -c(e_{i,2})(\lambda - 1) [1 - \hat{F}_{i,1}(\hat{\theta}_{i,1}(a))] \hat{E}_{i,1}[X_{i,2}(a) - \hat{\theta}_{i,1}(a) | X_{i,2}(a) > \hat{\theta}_{i,1}(a)]. \tag{17}
\]
Substituting Equation 17 back into Equation 15 yields:

\[
\widehat{E}_{i,1}[u_{i,2}|e_{i,2}] = m - \hat{\theta}_{i,1}(a)c(e_{i,2}) - \eta(\lambda - 1)[1 - \widehat{F}(\hat{\theta}_{i,1}(a))]\widehat{E}_{i,1}[X_{i,2}(a) - \hat{\theta}_{i,1}(a) | X_{i,2}(a) > \hat{\theta}_{i,1}(a)]c(e_{i,2})
\]

\[
= m - h(\hat{\theta}_{i,1}(a))c(e_{i,2}),
\]

where

\[
h(\hat{\theta}_{i,1}(a)) \equiv \hat{\theta}_{i,1}(a) + \eta(\lambda - 1)[1 - \widehat{F}(\hat{\theta}_{i,1}(a))]\widehat{E}_{i,1}[X_{i,2}(a) - \hat{\theta}_{i,1}(a) | X_{i,2}(a) > \hat{\theta}_{i,1}(a)].
\]

Recall that we assumed the participant’s prior over \(\theta_i(a)\) and the distribution over \(e_{i,2}\) are both normal. Thus, according to the participant, \(X_{i,2}\) is normally distributed with mean \(\hat{\theta}_{i,1}(a)\); let \(\xi^2\) denote the variance of \(X_{i,2}\). We can then write \(X_{i,2} = \hat{\theta}_{i,1}(a) + \delta\) where \(\delta \sim N(0, \xi^2)\). Substituting this into Equation 19 yields

\[
h(\hat{\theta}_{i,1}(a)) = \hat{\theta}_{i,1}(a) + \eta(\lambda - 1)\frac{1}{2}E[\delta | \delta > 0],
\]

where we have additionally used the fact that \(\widehat{F}(\hat{\theta}_{i,1}(a)) = \frac{1}{2}\) given that \(X_{i,2}\) is symmetric about \(\hat{\theta}_{i,1}(a)\). It is well known that \(E[\delta | \delta > 0] = 2\xi \phi(0) = 2\xi / \sqrt{2\pi}\), where \(\phi\) is the standard normal PDF (see, e.g., Greene 2003). Hence,

\[
h(\hat{\theta}_{i,1}(a)) = \hat{\theta}_{i,1}(a) + \eta(\lambda - 1)\frac{\xi}{\sqrt{2\pi}}.
\]

From Equation 21, it is immediate that \(h\) is increasing in \(\hat{\theta}_{i,1}(a)\). Given that \(e_{i}^*\) is chosen such that \(\widehat{E}_{i,1}[u_{i,2}|e_{i}^*] = 0\), Equation 18 implies that the participant will select \(e_{i}^*\) such that \(h(\hat{\theta}_{i,1}(a))c(e_{i}^*) = m\). Therefore, \(e_{i}^*\) is decreasing in \(\hat{\theta}_{i,1}(a)\). It then follows that if \(\hat{\theta}_{i,1}(h)\) tends to be higher than \(\hat{\theta}_{i,1}(l)\), then participants assigned the noisy task will exhibit lower effort levels than those assigned the noiseless task (i.e., fixing \(p\), \(e^*(h|p) < e^*(l|p)\) on average).

**B Reference Points that Incorporate the BDM Mechanism**

In this appendix we consider how our theoretical predictions of Experiment 1 extend when a participant’s reference point incorporates the uncertainty introduced by the BDM mechanism. In particular, we show that the optimal effort of a participant with reference-dependent utility is still decreasing in her estimate of the effort-cost parameter, \(\hat{\theta}_{i,1}(a)\).

Consider participant \(i\) who has been assigned to task \(a\). Recall her willingness to do additional
work is elicited via a BDM mechanism: the participant announces \( e_i \in [0, 100] \) and then a number \( e \) is uniformly drawn from \([0, 100]\) at random. If \( e < e_i \), the participant completes \( e \) tasks in exchange for a bonus of \( m \) dollars. Otherwise, she does no additional work and does not earn a bonus. Thus, conditional on submitting \( e_i \) to the mechanism, the participant will do additional work with probability \( G(e_i) \), where \( G \) denotes the CDF of a uniform random variable on \([0, 100]\) (and \( g \) denotes the associated PDF). Furthermore, upon submitting \( e_i \), the participant’s expected consumption utilities on the money and effort dimensions are, respectively, \( r^m(e_i) \equiv G(e_i)m \) and \( r^e(e_i) \equiv G(e_i)\hat{E}_{i,1}[V_{i,2}^e | e < e_i] \) where \( \hat{E}_{i,1}[V_{i,2}^e | e < e_i] = \hat{\theta}_{i,1}(a) \cdot \int_0^e c(e) \frac{g(e)}{G(e_i)} de \). Thus, the values \( r^m(e_i) \) and \( r^e(e_i) \) serve as the participant’s reference points along each dimension in period 2. As such, she chooses \( e_i^* \) to maximize

\[
\hat{E}_{i,1}[u_{i,2} | e_i] = G(e_i) \left\{ \hat{E}_{i,1}[V_{i,2}^e + \eta n(V_{i,2}^e | r^e(e_i; \hat{\theta}_{i,1}(a))) | e < e_i] + m + \eta (m - r^m(e_i)) \right\} \\
+ [1 - G(e_i)] \left\{ \eta (0 - r^e(e_i; \hat{\theta}_{i,1}(a))) + \eta \lambda (0 - r^m(e_i)) \right\}, \tag{22}
\]

where the expectation \( \hat{E}_{i,1} \) is with respect to the random number \( e \) drawn by the mechanism, \( e_{i,2}(a) \), and the participant’s updated beliefs over \( \theta_i(a) \). The first term in braces in Equation 22 is the participant’s expected utility conditional on the BDM assigning additional work. In this contingency, her disutility of effort will (on average) come as a loss relative to her expected value on this dimension, \( r^e(e_i; \hat{\theta}_{i,1}(a)) \), since this expectation incorporates a chance of no extra work and hence zero effort. Similarly, the monetary bonus comes as a gain relative to her expected monetary gain, \( r^m(e_i) \), which incorporates a chance of no extra work and hence no bonus. The second term in braces is the participant’s expected gain-loss utility conditional on the BDM assigning no additional work. In this contingency, she experiences a gain on the effort dimension but a loss on the monetary dimension.

Similar to the analysis in the main text (which assumes that the BDM does not influence the person’s reference point), the treatment probability \( p \) may influence \( e_i \) through its affect on the participant’s perception of \( \theta_i(a) \) (i.e., via misattribution). Thus, we will examine how the optimal effort choice, \( e_i^* \), depends on this perception, \( \hat{\theta}_{i,1}(a) \). To simplify the analysis below, we assume the participant forms certain beliefs about \( \theta_i(a) \) following period 1, and thus the contingency in which she is assigned additional work necessarily comes as a loss on the effort dimension.

First consider the case without reference dependence (i.e., \( \eta = 0 \)). The objective function in Equation 22 reduces to

\[
\hat{E}_{i,1}[u_{i,2} | e_i] = G(e_i) \left( \hat{E}_{i,1}[V_{i,2}^e | e < e_i] + m \right) = \hat{\theta}_{i,1}(a) \cdot \int_0^{e_i} c(e) g(e) de + G(e_i)m, \tag{23}
\]
and the first-order condition implies an optimal choice of \( e_i^* = c^{-1}(m/\hat{\theta}_{i,1}(a)) \). Clearly \( e_i^* \) is decreasing in \( \hat{\theta}_{i,1}(a) \).

We now consider the case with reference dependence (i.e., \( \eta > 0 \)). It is helpful to rewrite the objective function in Equation 22 as the sum of two components: the expected monetary benefit from statement \( e_i \), which we denote by

\[
B(e_i) \equiv G(e_i) \left\{ m + \eta \left( m - r^m(e_i) \right) \right\} - \eta \lambda [1 - G(e_i)] r^m(e_i),
\]

and the expected cost from \( e_i \), which we denote by

\[
K(e_i; \hat{\theta}_{i,1}(a)) \equiv -G(e_i) \hat{E}_{i,1} \left[ V_i^e + \eta n \left( V_i^e | r^e(e_i; \hat{\theta}_{i,1}(a)) \right) | e < e_i \right] + \eta [1 - G(e_i)] r^e(e_i; \hat{\theta}_{i,1}(a)).
\]

Thus, the objective in Equation 22 reduces so that the person chooses \( e_i \) to maximize

\[
\hat{E}_{i,1}[u_{i,2}|e_i] = B(e_i) - K(e_i; \hat{\theta}_{i,1}(a)).
\]

Given the objective above, we now analyze when the optimal effort choice, \( e_i^* \), is a decreasing function of \( \hat{\theta}_{i,1}(a) \). Let \( L(e_i; \hat{\theta}_{i,1}(a)) \) denote the first derivative of the objective function with respect to \( e_i \):

\[
L(e_i; \hat{\theta}_{i,1}(a)) \equiv \frac{\partial B(e_i)}{\partial e_i} - \frac{\partial K(e_i; \hat{\theta}_{i,1}(a))}{\partial e_i},
\]

so the first-order condition (FOC) requires \( L(e_i^*; \hat{\theta}_{i,1}(a)) = 0 \). Using the Implicit Function Theorem,

\[
\frac{\partial e_i^*}{\partial \hat{\theta}_{i,1}(a)} = -\left( \frac{\partial L(e_i^*; \hat{\theta}_{i,1}(a))}{\partial e_i^*} \right)^{-1} \frac{\partial L(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \hat{\theta}_{i,1}(a)}.
\]

Thus, so long as the second-order condition (SOC) holds and the FOC thus describes the optimum, then \( \frac{\partial L(e_i^*; \hat{\theta}_{i,1}(a))}{\partial e_i^*} < 0 \) and

\[
\text{sgn} \left( \frac{\partial e_i^*}{\partial \hat{\theta}_{i,1}(a)} \right) = \text{sgn} \left( \frac{\partial L(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \hat{\theta}_{i,1}(a)} \right).
\]

Furthermore, since only the cost component of the objective depends on \( \hat{\theta}_{i,1}(a) \), we have

\[
\frac{\partial L(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \hat{\theta}_{i,1}(a)} = -\frac{\partial^2 K(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \hat{\theta}_{i,1}(a) \partial e_i^*}.
\]

From Equation 25 and the definition of \( r^e(e_i; \hat{\theta}_{i,1}(a)) \) (along with our assumption that the partici-
We now consider the case with loss aversion, so \( \Lambda > 0 \). Together, Equations 29 and 30 imply that \( e_i^* \) is decreasing in \( \hat{\theta}_{i,1}(a) \) if \( \frac{\partial^2 K(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \theta_{i,1}(a) \partial e_i^*} > 0 \). From 32, \( \frac{\partial^2 K(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \theta_{i,1}(a) \partial e_i^*} > 0 \) iff

\[
c(e_i^*) \{ 1 + \Lambda[1 - G(e_i^*)] \} - \tilde{g} \Lambda \tilde{c}(e_i^*) > 0
\]

\[
\Leftrightarrow \{ 1 + \Lambda[1 - G(e_i^*)] \} > \frac{\tilde{g} \Lambda \tilde{c}(e_i^*)}{c(e_i^*)}.
\]

Furthermore, using Equations 33 and 32, the SOC implies that

\[
\frac{\partial^2 B(e_i)}{\partial e_i^2} \bigg|_{e_i = e_i^*} - \frac{\partial^2 K(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \theta_{i,1}(a) \partial e_i^*} \bigg|_{e_i = e_i^*} < 0
\]

\[
\Leftrightarrow 2m\tilde{g} \Lambda < \hat{\theta}_{i,1}(a) \left[ c'(e_i^*) \{ 1 + \Lambda[1 - G(e_i^*)] \} - 2\tilde{g} \Lambda \tilde{c}(e_i^*) \right].
\]

(35)
Maintaining our implicit assumption that \( \hat{\theta}_{i,1}(a) > 0 \), Condition 35 then holds only if

\[
0 < c'(e_i^*) \{1 + \Lambda[1 - G(e_i^*)]\} - 2\bar{g}\Lambda c(e_i^*) \Leftrightarrow \{1 + \Lambda[1 - G(e_i^*)]\} > 2\bar{g}\Lambda \frac{c(e_i^*)}{c'(e_i^*)}.
\] (36)

Substituting inequality 36 into 34 establishes that \( \frac{\partial^2 K(e_i^*; \hat{\theta}_{i,1}(a))}{\partial \hat{\theta}_{i,1}(a) \partial e_i^*} > 0 \) if

\[
2 \frac{c(e_i^*)}{c'(e_i^*)} > \frac{\bar{c}(e_i^*)}{c(e_i^*)} \Leftrightarrow 2c(e_i^*)^2 > c'(e_i^*)\bar{c}(e_i^*).
\] (37)

Condition 37 holds, for instance, for any \( c(\cdot) \) that is a power function, as we assume in our parametric estimation. Under our specification of \( c(e) = e^\gamma \) for \( \gamma > 1 \) (see Section 3.3), Condition 37 is equivalent to

\[
2e^{2\gamma} > \frac{\gamma}{\gamma+1}e^{2\gamma}.
\] (38)

We have therefore shown that, with a power-function cost specification (or any other specification that meets Condition 37), the optimal action \( e_i^* \) is a decreasing function of \( \hat{\theta}_{i,1}(a) \) when the participant’s reference point is the expected value of the lottery induced by the BDM mechanism. Given that \( e_i^* \) is a decreasing function of \( \hat{\theta}_{i,1}(a) \), the predictions of Observations 1 and 2 carry over to this setting. Namely: \( p \) does not directly influence a participant’s objective function, but, under misattribution, \( e_i^*(a|p) \) is an increasing function of \( p \) because \( \hat{\theta}_{i,1}(a) \) is a decreasing function of \( p \).

## C Experiment 2: Predictions of Reference Dependence

In this appendix we consider the predictions of the reference-dependent model absent misattribution in Experiment 2. In particular, we show that expectations-based reference dependence with a “forward looking” reference point (a la Kőszegi and Rabin) generates an effect that pushes effort in our experiment in the opposite direction as misattribution. Namely, reference-dependence causes participants assigned the noiseless task to (on average) increase effort across periods, and those assigned the noisy task to (on average) decrease effort across periods. In Section 4.2 we discussed how sufficiently strong misattribution generates the opposite pattern: participants assigned the noiseless task tend to decrease effort across periods, while those assigned the noisy task tend to increase it.

The analysis in this section builds on Appendix B, where we described how a participant with reference-dependent utility optimally chooses effort when her reference point incorporates the uncertainty introduced by the BDM mechanism. We now extend that analysis to the two-period setting of Experiment 2.
The key differentiating feature of Experiment 2 is that the participant’s reference point when making her first effort decision might still reflect the lottery induced by the coin flip. That is because a participant’s decision about effort comes roughly 10 minutes after the resolution of the coin flip, and this may be too little time for the participant’s reference point to fully adapt to her assigned task. If her reference point does not adapt, then her expected utility on each dimension prior to the coin flip determines her reference point on that dimension. Thus, her reference point—and hence her behavior—depends on the utility she expects from each of the tasks, not just the one she is ultimately assigned. This contrasts with Experiment 1: in that design, a participant chooses effort only once, and that choice happens well after she learned her task assignment. This allows ample time for her reference point to adapt to her assigned task. Thus, in Experiment 1, a person’s reference point at the time of choice depends solely on her realized task assignment and not on what “could have been” if the coin landed differently.

Before analyzing the case where the participant’s reference point does not adapt to the assigned task by the time of her first effort decision, it is worth noting predictions for the case where it does adapt before the first decision. With a quickly adapting reference point, effort in each period is described by the single-decision solution derived in Appendix B. Thus, if we assume that, on average, participants have unbiased priors about $\theta(a)$—implying that the average of participants’ expectations over $\theta(a)$ does not move in a systematic direction over time—then the average effort of those facing a given task is constant across periods. Fixing participants average beliefs over $\theta(a)$, this effort level is given by the value $e^*(a)$ that maximizes Objective 26 in Appendix B. As we showed there, $e^*(l) > e^*(h)$—effort by those assigned the noiseless task is predicted to be greater than those assigned the noisy task.

In contrast, even with unbiased priors (on average), reference dependence absent misattribution can generate systematic aggregate changes in effort across periods when participants’ reference points do not adapt prior to the first decision. The basic intuition is as follows. Let $e^*_t(a)$ denote the optimal effort level a participant reports to the BDM mechanism in period $t$ when assigned to task $a$. In period 2—when the participant’s task is fully anticipated—her optimal effort choice $e^*_2(a)$ will follow the derivation in Appendix B: she exhibits a high willingness to work if assigned the noiseless task, and a low willingness to work if assigned the noisy task. In period 1, however, her optimal strategy involves less effort on the noiseless task relative to period 2 (i.e., $e^*_1(l) < e^*_2(l)$), and more effort on the noisy task relative to period 2 (i.e., $e^*_1(h) > e^*_2(h)$). In other words, the difference in effort across tasks is more compressed in period 1 than it is in period 2 (i.e., $e^*_1(l) - e^*_1(h) < e^*_2(l) - e^*_2(h)$).

The strategy described above is optimal because it mitigate losses. In particular, the participant
chooses to work less on the noiseless task in period 1 (relative to what she would do in period 2) so that her expected payment and effort cost from the noiseless task in period 1 are more similar to what she expects to earn from the noisy task. By equalizing expected payments and effort costs across the tasks, neither assignment will generate large sensations of loss. For example, if the participant instead planned to work as much on the noiseless task in the first period as she would in the second period, then being assigned to the noisy task would come with a substantial loss on the money dimension: she would have earned more if she were assigned the noiseless task because she planned to work a substantial amount on that task. By planning to initially work less on the noiseless task (and more on the noisy task) relative to period 2, she can reduce such sensations of disappointment stemming from her assignment. Notice that this loss-mitigation strategy is only relevant when the participant compares her realized outcome to the expected outcomes from each possible task assignment—that is, when the participant’s reference point has not adjusted to her assigned task. This is why such a strategy is irrelevant for our analysis of effort choices in Experiment 1 and in period 2 of Experiment 2.

We now formalize the intuition described above. As in Appendix B, we simplify the analysis by assuming the participant forms degenerate beliefs about \( \theta(h) \) and \( \theta(l) \) following the initial learning session; we denote these beliefs by \( \hat{\theta}(h) \) and \( \hat{\theta}(l) \), respectively.\(^{50}\) Our approach follows the “person equilibrium” concept introduced by Kőszegi and Rabin (2006): prior to her task assignment, the participant forms effort plans for each possible assignment outcome and period, denoted by \( e_t(a) \), and these plans determine her expectations in each period. A personal equilibrium requires that these plans are consistent: given the reference points induced by her plans, it is optimal for the participant to follow through with these plans. Furthermore, we will focus on the participant’s preferred personal equilibrium, which is the consistent plan that provides the highest expected utility out of all consistent plans.

**Reference Points.** Let \( r^m_t \) and \( r^e_t \) denote the participant’s reference point in period \( t \) over money and effort, respectively. We assume that these reference values are equal to the participant’s expected monetary payment and effort cost in each round. In a personal equilibrium, these values are therefore endogenously determined by the participant’s effort plans. To clarify, recall from our analysis of Experiment 1 (Appendix B) that the chance of working in period \( t \) conditional on announcing \( e_t(a) \) to the BDM mechanism is \( G(e_t(a)) = \bar{g}e_t(a) \) (where \( \bar{g} = \frac{1}{100} \) since our experiment uses \( e \sim \text{Uniform}[0, 100] \)). Furthermore, the expected disutility of effort from task \( a \) conditional on announcing \( e_t(a) \) is \( G(e_t(a))\hat{E}_t[V_t^e|e < e_t(a)] = -\bar{g}\hat{\theta}(a)\hat{c}(e_t(a)) \), where \( \hat{c}(e) = \int_0^e c(x)dx \). These formulae allow us to write \( r^m_t \) and \( r^e_t \) in terms of the participant’s effort plans:

\(^{50}\)This allows us to derive predictions by focusing on a single participant. We could instead allow for uncertainty over \( \theta_i(a) \) and derive aggregate predictions. Given our previous assumption that an individual’s priors are unbiased on average (i.e., \( \mathbb{E}[\hat{\theta}_{i,0}(a)|\theta_i(a)| = \theta_i(a)) \), the average population beliefs should remain constant over time under rational updating.
• In period 1, we assume the participant’s reference point along each dimension matches the expectations she forms prior to the coin flip. Thus, her expected disutility of effort is the average disutility she expects to face from the noiseless and noisy task, and her expected payment is the average earnings she expects from each task. These ex-ante expectations depend on the participant’s effort plan contingent on each outcome of the coin-flip. Thus, if she plans to report \( e_1(a) \) to the mechanism conditional on being assigned to task \( a \), then her reference points are given by:

\[
\begin{align*}
    r^m_1 & = \frac{\bar{g}}{2} [e_1(h) + e_1(l)] \text{ and } r^e_1 = -\frac{\bar{g}}{2} [\hat{\theta}(h)\bar{c}(e_1(h)) + \hat{\theta}(l)\bar{c}(e_1(l))]. \\
\end{align*}
\]

(39)

• In period 2, we assume the participant’s reference point along each dimension has adapted to her task: she expects to work on task \( a \), and therefore she forms her expectations over her disutility of effort and payment conditional on working on task \( a \). This matches our assumption for the reference point in the single-period analysis of Experiment 1. Thus, conditional on being assigned to task \( a \), her reference points are given by:

\[
\begin{align*}
    r^m_2(a) & = m\bar{g}e_2(a) \text{ and } r^e_2(a) = -\hat{\theta}(a)\bar{c}(e_2(a)), \\
\end{align*}
\]

(40)

We next analyze the optimal effort plans given these reference points they induce.

**Objective Function.** As in our analysis of Experiment 1, we can break down the objective function in each period into the expected monetary benefit and expected effort cost. Following Equation 24 from Appendix B, the expected monetary benefit from task \( a \) in period \( t \) is

\[
\begin{align*}
    B^a_t(e_t(a)|r^m_t) & \equiv G(e_t(a)) \left\{ m + \eta \left( m - r^m_t \right) \right\} - \eta \lambda \left[ 1 - G(e_t(a)) \right] r^m_t \\
    & = (1 + \eta)\bar{g}e_t(a)m - \eta\bar{g}e_t(a)r^m_t - \eta \lambda \left[ 1 - \bar{g}e_t(a) \right] r^m_t \\
    & = (1 + \eta)\bar{g}e_t(a)m - \eta r^m_t - \Lambda \left[ 1 - \bar{g}e_t(a) \right] r^m_t, \\
\end{align*}
\]

(41)

where \( \Lambda \equiv \eta(\lambda - 1) \). Following Equation 25 from Appendix B, the expected effort cost from task \( a \) in period \( t \) is

\[
\begin{align*}
    K^a_t(e_t(a)|r^e_t) & \equiv -G(e_t(a)) \left\{ \bar{E}_t[V^e_t|e < e_t(a)] + \eta \lambda \left( \bar{E}_t[V^e_t|e < e_t(a)] - r^e_t \right) \right\} + \eta \left[ 1 - G(e_t(a)) \right] r^e_t \\
    & = (1 + \eta \lambda)\hat{\theta}(a)\bar{c}(e_t(a)) + \eta \lambda \bar{g}e_t(a)r^e_t + \eta \left[ 1 - \bar{g}e_t(a) \right] r^e_t \\
    & = (1 + \eta \lambda)\hat{\theta}(a)\bar{c}(e_t(a)) + \eta r^e_t + \Lambda \bar{g}e_t(a)r^e_t. \\
\end{align*}
\]

(42)
Optimal Effort in $t = 1$. In period $t = 1$, the optimal effort choices, $e^*_t(h)$ and $e^*_t(l)$, jointly maximize $\frac{1}{2}[B'(e_1(h)|r^m_1) - K'(e_1(h)|r_2)] + \frac{1}{2}[B'(e_1(l)|r^m_1) - K'(e_1(l)|r_2)]$. The FOC with respect to $e_1(h)$ is thus

$$
(1 + \eta)\bar{g} - \{2\eta + \Lambda[2 - \bar{g}(e_1(h) + e_1(l))]\} \frac{\partial r^m_1}{\partial e_1(h)} + \Lambda \bar{g} r^m_1
- (1 + \eta \lambda) \hat{\theta}(h) \bar{c}(e_1(h)) - \{2\eta + \Lambda \bar{g}(e_1(h) + e_1(l))\} \frac{\partial r^e_1}{\partial e_1(h)} - \Lambda \bar{g} r^e_1 = 0,
$$

(43)

and the FOC with respect to $e_1(l)$ is

$$
(1 + \eta)\bar{g} - \{2\eta + \Lambda[2 - \bar{g}(e_1(h) + e_1(l))]\} \frac{\partial r^m_1}{\partial e_1(l)} + \Lambda \bar{g} r^m_1
- (1 + \eta \lambda) \hat{\theta}(h) \bar{c}(e_1(l)) - \{2\eta + \Lambda \bar{g}(e_1(h) + e_1(l))\} \frac{\partial r^e_1}{\partial e_1(l)} - \Lambda \bar{g} r^e_1 = 0.
$$

(44)

From the definitions of $r^m_1$ and $r^e_1$ in Equation 39, we have

$$
\frac{\partial r^m_1}{\partial e_1(a)} = \frac{m \bar{g}}{2} \quad \text{and} \quad \frac{\partial r^e_1}{\partial e_1(a)} = -\frac{\bar{g}}{2} \hat{\theta}(a) c(e_1(a)).
$$

(45)

Hence, two FOCs above can be written, respectively, as

$$
m \{1 - \Lambda + \Lambda \bar{g}[e_1(h) + e_1(l)]\} - \hat{\theta}(h) \{1 + \Lambda - \frac{\bar{g}}{2}[e_1(h) + e_1(l)]\} c(e_1(h))
+ \Lambda \frac{\bar{g}}{2} [\hat{\theta}(h) \bar{c}(e_1(h)) + \hat{\theta}(l) \bar{c}(e_1(l))] = 0,
$$

(46)

and

$$
m \{1 - \Lambda + \Lambda \bar{g}[e_1(h) + e_1(l)]\} - \hat{\theta}(l) \{1 + \Lambda - \frac{\bar{g}}{2}[e_1(h) + e_1(l)]\} c(e_1(l))
+ \Lambda \frac{\bar{g}}{2} [\hat{\theta}(h) \bar{c}(e_1(h)) + \hat{\theta}(l) \bar{c}(e_1(l))] = 0.
$$

(47)

Let $L^h_1(e_1(h), e_1(l))$ and $L^l_1(e_1(h), e_1(l))$ denote the functions defined by the left-hand side of Equations 46 and 47, respectively. Restricting attention to interior solutions, the optimal effort plans in period 1, $e^*_1(h)$ and $e^*_1(l)$, must solve the system of equations given by $L^h_1(e^*_1(h), e^*_1(l)) = 0$ and $L^l_1(e^*_1(h), e^*_1(l)) = 0$. Notice, however, that if $L^h_1(e^*_1(h), e^*_1(l)) = L^l_1(e^*_1(h), e^*_1(l)) = 0$, then it is immediate from Equations 46 and 47 that $\hat{\theta}(h) c(e^*_1(h)) = \hat{\theta}(l) c(e^*_1(l))$; that is, the optimal effort levels are chosen to equalize the “consumption utility” of effort across the two tasks. To summarize:
Lemma 1. Given the setup formalized above, the optimal effort choices in period 1, \( e_1^*(h) \) and \( e_1^*(l) \), are such that the participant’s effort cost is the same regardless of her task assignment; that is, \( \hat{\theta}(h)c(e_1^*(h)) = \hat{\theta}(l)c(e_1^*(l)) \). Furthermore, given that \( \hat{\theta}(h) > \hat{\theta}(l) \), the participant’s initial effort on the noiseless task exceeds her initial effort on the noisy task; that is, \( e_1^*(l) > e_1^*(h) \).

Optimal Effort in \( t = 2 \). In period \( t = 2 \), the optimal effort choice conditional on being assigned to task \( a \), \( e_2^*(a) \), maximizes \( B_2^a(e_2(a)|r_2^m(a)) - K_2^a(e_2(a)|r_2^e(a)) \), and thus solves the following FOC:

\[
(1 + \eta) \tilde{g}m - \{ \eta + \Lambda[1 - \tilde{g}e_2(a)] \} \frac{\partial r_2^m(a)}{\partial e_2(a)} + \Lambda \tilde{g}r_2^m(a) \\
- (1 + \eta \lambda) \hat{\theta}(a) \tilde{g}c(e_2(a)) - \{ \eta + \Lambda \tilde{g}e_2(a) \} \frac{\partial r_2^e(a)}{\partial e_2(a)} - \Lambda \tilde{g}r_2^e(a) = 0. \tag{48}
\]

From the definition of \( r_2^m(a) \) and \( r_2^e(a) \) in Equation 40, we have

\[
\frac{\partial r_2^m(a)}{\partial e_2(a)} = m \tilde{g} \quad \text{and} \quad \frac{\partial r_2^e(a)}{\partial e_2(a)} = -\hat{\theta}(a) \tilde{g}c(e_2(a)), \tag{49}
\]

and thus the FOC in Equation 48 can be written as

\[
m\{1 - \Lambda + 2\Lambda \tilde{g}e_2(a)\} - \hat{\theta}(a)\{1 + \Lambda - \Lambda \tilde{g}e_2(a)\}c(e_2(a)) + \hat{\theta}(a)\Lambda \tilde{g}c(e_2(a)) = 0. \tag{50}
\]

Let \( L_2^a(e) \) denote the function on the left-hand-side of the FOC above. Thus, \( e_2^*(a) \) is such that \( L_2^a(e_2^*(a)) = 0 \).

Given that this FOC takes the same form as the one in Appendix B, the analysis from Appendix B implies that \( e_2^*(a) \) is decreasing in \( \hat{\theta}(a) \). Thus, given that \( \hat{\theta}(h) > \hat{\theta}(l) \), we have that \( e_2^*(l) > e_2^*(h) \): similar to period 1, the optimal plan in period 2 calls for greater effort when facing the noiseless task than the noisy one.

Changes in Optimal Effort Across Periods. We have so far shown that a participant who believes \( \hat{\theta}(h) > \hat{\theta}(l) \) will, within each period, exert more effort if she’s assigned the noiseless task rather than the noisy one. But, fixing the task she ultimately faces, how will her effort change across periods? The final step of our analysis compares \( e_1^*(a) \) with \( e_2^*(a) \) for each \( a \in \{h,l\} \). For this step, we will simplify matters by assuming—as in previous sections—that effort costs follow a power function; that is, \( c(e) = e^\gamma \) for some \( \gamma > 1 \). We will first consider changes in effort for the noiseless task \( (a = l) \) and then consider the noisy task \( (a = h) \).

1. Willingness to Work on the Noiseless Task Increases Across Periods. We now show that
\( e_1^*(l) < e_2^*(l) \). From Lemma 1, we have \( \hat{\theta}(l)c(e_1^*(l)) = \hat{\theta}(h)c(e_1^*(h)) \) and thus

\[
e_1^*(h) = c^{-1} \left( \frac{\hat{\theta}(l)}{\hat{\theta}(h)} c(e_1^*(l)) \right) = \left( \frac{\hat{\theta}(l)}{\hat{\theta}(h)} \right)^{1/\gamma} e_1^*(l) = \psi_L e_1^*(l),
\]

(51)

where \( \psi_L \equiv \left( \frac{\hat{\theta}(l)}{\hat{\theta}(h)} \right)^{1/\gamma} \). In light of Equation 51, we can write the FOC \( L_1^l(e_1^*(h), e_1^*(l)) \) characterizing effort in period 1 entirely in terms of \( e_1^*(l) \) by substituting the expression for \( e_1^*(h) \) from Equation 51 into Equation 47, which yields

\[
m\{1 - \Lambda + \Lambda \hat{g} [1 + \psi_L] e_1^*(l)\} - \hat{\theta}(l) \{1 + \Lambda - \Lambda \hat{g}^2 [1 + \psi_L] e_1^*(l)\} c(e_1^*(l)) + \Lambda \hat{g} [1 + \psi_L] \hat{\theta}(l) \hat{c}(e_1^*(l)) = 0.
\]

(52)

Letting \( \bar{L}^l(e_1(l); \psi) \) denote the left-hand side of the equation above as a function of \( e_1(l) \) and the parameter \( \psi \), we have that \( e_1^*(l) \) must satisfy \( \bar{L}^l(e_1^*(l); \psi_L) = 0 \). Notice, however, that the FOC characterizing \( e_2^*(l) \) (Equation 50) is identical to the FOC above except the parameter \( \psi \) takes value 1 instead of \( \psi_L < 1 \); that is, \( e_2^*(l) \) solves \( \bar{L}^l(e_2^*(l); 1) = 0 \). Thus, to show that \( e_1^*(l) < e_2^*(l) \), it suffices to show that the solution to \( \bar{L}^l(e^*; \psi) = 0 \) is increasing in \( \psi \). Using the implicit function theorem,

\[
\frac{\partial e^*}{\partial \psi} = - \left( \frac{\partial \bar{L}(e^*; \psi)}{\partial e^*} \right)^{-1} \frac{\partial \bar{L}(e^*; \psi)}{\partial \psi}.
\]

(53)

Focusing on interior solutions, the second-order condition implies that \( \frac{\partial \bar{L}(e^*; \psi)}{\partial e^*} < 0 \), and hence \( \frac{\partial e^*}{\partial \psi} > 0 \) \( \Leftrightarrow \frac{\partial \bar{L}(e^*; \psi)}{\partial \psi} > 0 \). From Equation 52, we have

\[
\frac{\partial \bar{L}(e^*; \psi)}{\partial \psi} = m\Lambda \hat{g} e^* + \Lambda \hat{g} \hat{\theta}(l) [e^* c(e^*) + \hat{c}(e^*)] > 0,
\]

(54)

which therefore establishes that \( e_2^*(l) > e_1^*(l) \).

2. Willingness to Work on the Noisy Task Decreases Across Periods. We now show that \( e_1^*(h) > e_2^*(h) \). The argument is symmetric to the one above. Namely, from Lemma 1, the optimal effort in period 1 must satisfy \( e_1^*(l) = \psi_H e_1^*(h) \), where \( \psi_H = 1/\psi_L > 1 \). Substituting this

\[51\text{Recall that } \hat{c}(e) = \frac{e^{\gamma+1}}{\gamma+1}. \text{ Hence, } e_1(h) = \psi_L e_1(l) \text{ implies that } \hat{\theta}(h) \hat{c}(e_1(h)) = \hat{\theta}(h) \left( \frac{\hat{\theta}(l)}{\hat{\theta}(h)} \right)^{1+1/\gamma} e_1(l)^{\gamma+1}/(\gamma+1) = \hat{\theta}(h) \left( \frac{\hat{\theta}(l)}{\hat{\theta}(h)} \right)^{1+1/\gamma} \hat{c}(e_1(l)) = \psi_L \hat{c}(e_1(l)). \]
expression for \( e_i^*(l) \) into the FOC in Equation 46 implies that \( e_i^*(h) \) must solve

\[
m\{1 - \Lambda + \Lambda \tilde{g}[1 + \psi_H]e_i^*(h)\} - \hat{\theta}(h)\{1 + \Lambda - \frac{\Lambda \tilde{g}}{2}[1 + \psi_H]e_i^*(h)\}c(e_i^*(h))
\]
\[+\Lambda \frac{\tilde{g}}{2}[1 + \psi_H]c(\hat{e}_i^*(h)) = 0.
\]

Again letting \( \tilde{L}^h(e_1(h); \psi) \) denote the left-hand side of the equation above as a function of \( e_1(h) \) and the parameter \( \psi \), we have that \( e_1^*(h) \) must satisfy \( \tilde{L}^h(e_1^*(h); \psi_H) = 0 \). Equation 50 reveals, however, that \( e_2^*(h) \) must solve \( \tilde{L}^h(e_1^*(h); 1) = 0 \). The implicit-function-theorem argument from the previous part implies that the solution to \( \tilde{L}^h(e_1^*(h); \psi) = 0 \) is also increasing in \( \psi \). Thus, since \( \psi_H > 1 \), we have that \( e_2^*(h) < e_1^*(h) \).

D Additional Theoretical Details

In this appendix we provide details on how misattribution interferes with belief updating when a participant misencodes signals but otherwise follows Bayes’ Rule. These details provide a more formal derivation of the predictions generated by misattribution described in Sections 3.2 and 4.2. We consider two rounds of learning; that is, the participant receives two signals in sequence. We examine beliefs and behavior after the first round to address predictions for Experiment 1, and then consider both the first and second round to address predictions for Experiment 2. Regarding Experiment 1, we demonstrate that a misattributing participant \( i \) will form a systematically distorted estimate of her underlying effort-cost parameter, \( \theta_i(a) \). This estimate undershoots the true value when she is assigned the noiseless task and overshoots it when she is assigned the noisy one. We then demonstrate that the second signal is likely to move the participant’s estimate in the opposite direction of her initial error: her estimated cost of the noiseless task following the second round tends to increase while that of the noisy task tends to decrease.

As in previous sections, we assume that each participant \( i \) has prior beliefs over \( \theta_i(a) \) that follow a normal distribution: \( \theta_i(a) \sim N(\hat{\theta}_{i,0}(a), \rho^2) \), where \( \hat{\theta}_{i,0}(h) > \hat{\theta}_{i,0}(l) \). We also assume that participants’ initial estimates of \( \theta_i(a) \) are unbiased in the population so that for all \( i \) and \( a \), \( \mathbb{E}[\hat{\theta}_{i,0}(a) | \theta_i(a)] = \theta_i(a) \); hence, the initial estimates of \( \theta_i(a) \) averaged across individuals is equal to the true value. A participant’s signals about \( \theta_i(a) \) stem from either the single initial learning session in Experiment 1, or from both learning sessions in Experiment 2.\(^{52}\) Assuming the participant is assigned task \( a \), each learning session \( t \in \{1, 2\} \) provides a signal \( X_{i,t}(a) = \theta_i(a) + \epsilon_{i,t} \), where \( \epsilon_{i,t} \sim N(0, \sigma^2) \).\(^{53} \) Let \( x_{i,t}(a) \) denote the realized signal and let \( \hat{x}_{i,t}(a) \) denote the misattribution’s

\(^{52}\)We focus on a participant who is not assigned to do additional work in the first session of Experiment 2. Hence, the initial learning sessions comprise the participant’s only experience with the task.

\(^{53}\)Notice that \( X_{i,t}(a) = -V_{i,t}^a(a)/c(e_{i,t}) \), where \( V_{i,t}^a(a) \) is defined in Equation 5 and \( e_{i,t} \) is the required number of
encoded value of this signal.

Recall that \( \hat{\theta}_{i,t}(a) \) denotes the participant’s estimate of \( \theta(a) \) entering period \( t = 1, 2 \). Given our normality assumptions, a participant who is Bayesian (aside from misencoding signals) updates her beliefs as follows:

\[
\hat{\theta}_{i,t}(a) = \hat{\theta}_{i,t-1}(a) + \alpha_t \left[ \hat{x}_{i,t}(a) - \hat{\theta}_{i,t-1}(a) \right],
\]

where \( \alpha_t = \frac{p^2}{t(p^2 + \sigma^2)} \). The encoded signal, \( \hat{x}_{i,t} \), is defined by Equation 3, and thus depends on how the true signal, \( x_{i,t}(a) \), compares to the participant’s expected effort cost entering round \( t \). This expectation in turn depends on the participant’s treatment group: those who are initially uncertain which task they will work on will hold different expectations about their eventual effort cost than those who are certain. We will therefore examine how updating differs depending on the participant’s treatment group, \( p \), where where \( p \) denotes the participant’s ex ante likelihood of being assigned the noisy task. Recall that \( p = 1 \) and \( p = 0 \) correspond to the control group, and \( p = 1/2 \) corresponds to the coin-flip group.

We now analyze the encoded signal that a misattributing participant forms based on her assigned task and treatment group. For this exercise, we fix the true signal the participant receives in a given period, \( x_{i,t}(a) \), and consider how she would encode this signal if she were in the coin-flip group versus the control group. We let \( \hat{x}_{i,t}(a|p) \) denote this misencoded signal as a function of the treatment, \( p \). Once we establish the direction in which signals are biased across treatment groups, the updating rule in Equation 56 will then immediately reveal how the average estimate of the cost parameter in a given period should differ across treatments under misattribution.

**Biased Updating in Period 1.** We begin by analyzing how the treatment distorts signals in the first period. The predictions we obtain here primarily relate to Experiment 1. Consider participant \( i \) whose treatment group is such that she expects to face the noisy task with probability \( p \). Her expected cost signal entering period 1 is thus \( \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p] \equiv p\hat{\theta}_{i,0}(h) + (1-p)\hat{\theta}_{i,0}(l) \). Upon realizing signal \( x_{i,1}(a) \), Equation 3 implies that her misencoded signal is

\[
\hat{x}_{i,1}(a|p) = \begin{cases} 
  x_{i,1}(a) + \left( \frac{\eta - \hat{\eta}}{1 + \eta} \right) \left( x_{i,1}(a) - \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p] \right) & \text{if } x_{i,1}(a) \leq \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p] \\
  x_{i,1}(a) + \lambda \left( \frac{\eta - \hat{\eta}}{1 + \eta \lambda} \right) \left( x_{i,1}(a) - \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p] \right) & \text{if } x_{i,1}(a) > \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p]. 
\end{cases}
\]

Letting \( \kappa^G \equiv \left( \frac{\eta - \hat{\eta}}{1 + \eta} \right) \) and \( \kappa^E \equiv \lambda \left( \frac{\eta - \hat{\eta}}{1 + \eta \lambda} \right) \), we can define the following random variable that mea-
There are three cases to consider, depending on the values of $k$ (i.e., $p$ pending on whether the participant is in the coin-flip group (i.e., $p$ upward more than low cost signals are distorted downward.

Thus, we can write the misencoded signal in Equation 57 more simply as

$$ K_{i,1}(a|p) \equiv \begin{cases} \kappa^L & \text{if } x_{i,1}(a) > \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p] \\ \kappa^G & \text{if } x_{i,1}(a) \leq \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p]. \end{cases} \quad (58) $$

Thus, we can write the misencoded signal in Equation 57 more simply as

$$ \hat{x}_{i,1}(a|p) = x_{i,t}(a) + k_{i,1}(a|p)[x_{i,1}(a) - \hat{\mathbb{E}}_{i,0}[X_{i,1}(a)|p]] $$

$$ = x_{i,t}(a) + k_{i,1}(a|p)[x_{i,1}(a) - p\hat{\theta}_{t,0}(h) - (1 - p)\hat{\theta}_{t,0}(l)] , \quad (59) $$

where $k_{i,1}(a|p)$ is the realization of $K_{i,1}(a|p)$. Notice that if the participant does not suffer misattribution (i.e., $\hat{\eta} = \eta$), then $k_{i,1}(a|p)$ is always equal to zero. Furthermore, if the participant does suffer misattribution and is loss averse, then $\kappa^L > \kappa^G$, implying that high cost signals are distorted upward more than low cost signals are distorted downward.$^{54}$

First consider how signals about the noiseless task in particular are differentially distorted depending on whether the participant is in the coin-flip group (i.e., $p = 1/2$) or the control group (i.e., $p = 0$). From Equation 59, we have

$$ \hat{x}_{i,1}(l|p = 1/2) - \hat{x}_{i,1}(l|p = 0) $$

$$ = k_{i,1}(l|1/2)[x_{i,1}(l) - \cdot 5\hat{\theta}_{t,0}(h) - \cdot 5\hat{\theta}_{t,0}(l)] - k_{i,1}(l|0)[x_{i,1}(l) - \hat{\theta}_{t,0}(l)] \quad (60) $$

There are three cases to consider, depending on the values of $k_{i,1}(l|1/2)$ and $k_{i,1}(l|0)$:

i. $x_{i,1}(a) > \cdot 5\hat{\theta}_{t,0}(h) + \cdot 5\hat{\theta}_{t,0}(l)$, in which case $k_{i,1}(l|1/2) = k_{i,1}(l|0) = \kappa^L$;

ii. $x_{i,1}(a) \in [\hat{\theta}_{t,0}(l), \cdot 5\hat{\theta}_{t,0}(h) + \cdot 5\hat{\theta}_{t,0}(l)]$, in which case $k_{i,1}(l|1/2) = \kappa^G$ and $k_{i,1}(l|0) = \kappa^L$;

iii. $x_{i,1}(a) < \hat{\theta}_{t,0}(l)$, in which case $k_{i,1}(l|1/2) = k_{i,1}(l|0) = \kappa^G$.

In cases (i) and (iii), $k_{i,1}(l|0)$ and $k_{i,1}(l|1/2)$ are both equal to the same $\kappa^j \in \{\kappa^G, \kappa^L\}$, and hence 60 reduces to

$$ \hat{x}_{i,1}(l|p = 1/2) - \hat{x}_{i,1}(l|p = 0) = -\frac{\kappa^j}{2}[\hat{\theta}_{t,0}(h) - \hat{\theta}_{t,0}(l)] < 0. \quad (61) $$

$^{54}$Recall that $x_{i,t}(a)$ reflects the participant’s cost in period $t$. Hence, the participant experiences a positive surprise when her signal is less than expected; she experiences a negative surprise when it is greater than expected. This is why the signs in Equation 57 are flipped relative to 3—the latter was written in terms of benefits rather than costs.
In case (ii), \( k_{i,1}(l|0) \) and \( k_{i,1}(l|1/2) \) differ, leading to

\[
\hat{x}_{i,1}(l|p = 1/2) - \hat{x}_{i,1}(l|p = 0) = \kappa^G[x_{i,1}(l) - \theta_{i,0}(l)] + \kappa^L[x_{i,1}(l) - \hat{\theta}_{i,0}(l)]
\]

\[
< \kappa^G[x_{i,1}(l) - \theta_{i,0}(h) - .5\hat{\theta}_{i,0}(l)] - \kappa^G[x_{i,1}(l) - \hat{\theta}_{i,0}(l)]
\]

\[
= \frac{-\kappa^G}{2} [\hat{\theta}_{i,0}(h) - \hat{\theta}_{i,0}(l)]
\]

\[
< 0,
\]

where the first inequality follows because \( x_{i,1}(l) - \theta_{i,0}(l) > 0 \) given that \( x_{i,1}(a) \in [\hat{\theta}_{i,0}(l), .5\hat{\theta}_{i,0}(h) + .5\hat{\theta}_{i,0}(l)] \).

All three cases together imply that participant \( i \) facing task \( l \) will always encode a lower signal if she were in the coin-flip group rather than the control group. An entirely symmetric argument (omitted) implies that participant \( i \) facing task \( h \) will always record a higher signal if she were in the coin-flip group rather than the control group. Thus, assuming participants update according to Equation 56, the preceding results imply that \( \hat{\theta}_{i,1}(l|p = 1/2) < \hat{\theta}_{i,1}(l|p = 0) \) and \( \hat{\theta}_{i,1}(h|p = 1/2) > \hat{\theta}_{i,1}(h|p = 1) \), where \( \hat{\theta}_{i,1}(a|p) \) denotes a participant’s predicted expectation of \( \theta_i(a) \) conditional on her treatment group, \( p \).\(^{55}\) The results of our parametric estimation in Section 3.3 mirror these predictions (see Table 3). Furthermore, given this ordering of beliefs across treatment groups, the analysis of Appendices A and B shows that these beliefs generate the predicted differences in aggregate effort described Observation 2.

**Biased Updating In Period 2.** We now demonstrate how misattribution generates a predictable change in beliefs between periods 1 and 2 depending on a participant’s task assignment. The predictions we obtain here relate exclusively to Experiment 2 since Experiment 1 had only one round of experience prior to the sole effort decision. Recall that in Experiment 2, each participant was assigned her task via a coin flip. Thus, in this analysis, a participant’s period 1 beliefs, \( \hat{\theta}_{i,1}(a) \), correspond to \( \hat{\theta}_{i,1}(a|1/2) \) derived above.

We examine how these beliefs change after another round of learning. We maintain our assumption that the participant’s reference point in the second period adjusts to her assigned task. That is, \( \hat{E}_{i,1}[X_{i,2}(a)] = \hat{\theta}_{i,1}(a) \). Thus, the participant’s encoded signal is

\[
\hat{x}_{i,2}(a) = x_{i,2}(a) + k_{i,2}(a) \left[ x_{i,2}(a) - \hat{E}_{i,1}[X_{i,2}(a)] \right]
\]

\[
= x_{i,2}(a) + k_{i,2}(a) \left[ x_{i,2}(a) - \hat{\theta}_{i,1}(a) \right],
\]

\[(63)\]

\(^{55}\)This prediction invokes our assumption that a participant’s priors over the cost parameters are independent of their treatment assignment.
where $k_{i,2}(a)$ is the realization of

$$K_{i,2}(a) \equiv \begin{cases} \kappa^L & \text{if } x_{i,2}(a) > \hat{E}_{i,1}[X_{i,2}(a)] \\ \kappa^G & \text{if } x_{i,2}(a) \leq \hat{E}_{i,1}[X_{i,2}(a)] \end{cases} \quad (64)$$

From the updating rule in Equation (56) and the expression above for the misencoded signal, the participant’s expected change in beliefs between periods 1 and 2 (from an ex-ante perspective) is thus equal to

$$\mathbb{E}[(\hat{\theta}_{i,2}(a) - \hat{\theta}_{i,1}(a))] = \alpha_2 \mathbb{E}[X_{i,2}(a) - \hat{\theta}_{i,1}(a)] = \alpha_2 \mathbb{E}[(1 + K_{i,2}(a))(X_{i,2}(a) - \hat{\theta}_{i,1}(a))], \quad (65)$$

where $\mathbb{E}[\cdot]$ denotes expectations with respect to the true underlying distributions. Notice that Equation 65 is positive iff $\mathbb{E}[X_{i,2}(a)] > \mathbb{E}[\hat{\theta}_{i,1}(a)]$. Given that signals are independent across periods with mean $\theta_i(a)$, the previous condition holds iff $\theta_i(a) > \mathbb{E}[\hat{\theta}_{i,1}(a)]$. As we argued above, assignment to the noiseless task via the coin-flip implies that $\hat{\theta}_{i,1}(l)$ is biased downward. Hence the previous inequality holds and thus, in expectation, $\hat{\theta}_{i,2}(l) > \hat{\theta}_{i,1}(l)$: the perceived effort cost of the noiseless task tends to increase across periods, reducing the participant’s willingness to work on that task. Similarly, assignment to the noisy task via the coin-flip implies that $\hat{\theta}_{i,1}(h)$ is biased upward. Hence the inequality above fails to hold, and thus, in expectation, $\hat{\theta}_{i,2}(h) < \hat{\theta}_{i,1}(h)$: the perceived effort cost of the noisy task tends to decrease across periods, increasing the participant’s willingness to work on that task. This pattern in beliefs generated by misattribution clearly runs against the predictions of reference-dependence absent misattribution (explored in Appendix C), where effort on the noiseless task tends to increase across periods while effort on the noisy task tends to decrease.

### E  Supplemental Tables and Figures

In this appendix we provide additional empirical results that supplement the main text and provide robustness checks for our primary results.

We first show that dividing the Experiment 1 sample in half according to the total amount of time participants spent on the experiment (from the start of Session 1 to completion) does not have a large effect on our nonparametric results. This is demonstrated in Tables A1 and A2 below.

---

56Our assumption that priors are unbiased in the population is relevant here: this assumption implies that, on average, $\hat{\theta}_{i,1}(l) < \theta_i(a)$; that is, that the average expectation among participants regarding $\theta(l)$ after one round of learning is lower than the true parameter value.
However, these comparisons based on duration are limited due to unequal group sizes. Regression analysis (included in Table 3 in the main text) demonstrates that this effect does not alter the results of our parametric analysis.

We next show that the results of our parametric analysis are robust to changing the Stone-Geary background parameter that appears in the effort-cost function. Although our numerical estimates vary with this parameter, we show in Table A3 that our qualitative results hold for two alternative specifications of the background parameter which vary by an order of magnitude. For the reader’s ease, we omit such analysis for Experiment 2.

Next, we utilize a logit model to explore whether any observables predict attrition in Experiment 1 (Table A4). Although we have overall lower attrition in the high-probability treatment, we do not find that other factors influenced attrition. This effect is easily seen in Table 1 in the main text. We suspect this is due to the fact that we ran the high-probability session at a slightly different time of day.

We then turn to the Experiment 2. To address potential concerns about differential experience and learning, Table A5 presents non-parametric results for Experiment 2 (analogous to the final two columns of Table 4) in which we drop any participants who completed extra tasks in the first session. This analysis leaves far fewer participants in our sample, but our qualitative results hold. We then utilize the random numbers from the BDM in our experiment to instrument for whether a person completed extra tasks. This analysis verifies that, while doing extra tasks may have changed willingness to work in Session 2, our primary conclusions remain for those who did not complete extra tasks.

Finally, following the robustness exercise in Experiment 1 concerning attrition, we estimate a similar logit model for Experiment 2 (Table A6). We did not collect demographic information from participants in Experiment 2, and thus we have fewer potential explanatory variables. That said, we find no convincing link between observables and attrition.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Control</th>
<th></th>
<th>Control</th>
<th></th>
<th>High Prob.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>noise=0</td>
<td>noise=1</td>
<td>noise=0</td>
<td>noise=1</td>
<td>noise=0</td>
<td>noise=1</td>
</tr>
<tr>
<td>Willingness to Work</td>
<td>22.38</td>
<td>20.69</td>
<td>27.67</td>
<td>17.43</td>
<td>23.27</td>
<td>21.25</td>
</tr>
<tr>
<td></td>
<td>(1.458)</td>
<td>(1.876)</td>
<td>(2.034)</td>
<td>(1.689)</td>
<td>(2.203)</td>
<td>(2.759)</td>
</tr>
<tr>
<td>Observations</td>
<td>430</td>
<td>385</td>
<td>365</td>
<td>390</td>
<td>245</td>
<td>195</td>
</tr>
</tbody>
</table>

Notes: Willingness to work is averaged over five payment levels. Standard errors (in parentheses) are clustered at the individual level with 402 clusters.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control</th>
<th></th>
<th>Control</th>
<th></th>
<th>High Prob.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>noise=0</td>
<td>noise=1</td>
<td>noise=0</td>
<td>noise=1</td>
<td>noise=0</td>
<td>noise=1</td>
</tr>
<tr>
<td>Willingness to Work</td>
<td>28.53</td>
<td>24.5</td>
<td>29.81</td>
<td>17.941</td>
<td>24.71</td>
<td>21.38</td>
</tr>
<tr>
<td></td>
<td>(2.846)</td>
<td>(2.670)</td>
<td>(2.617)</td>
<td>(2.253)</td>
<td>(1.596)</td>
<td>(1.412)</td>
</tr>
<tr>
<td>Observations</td>
<td>185</td>
<td>280</td>
<td>280</td>
<td>275</td>
<td>445</td>
<td>545</td>
</tr>
</tbody>
</table>

Notes: Willingness to work is averaged over five payment levels. Standard errors (in parentheses) are clustered at the individual level with 402 clusters.
Table A3:
**EXPERIMENT 1. ROBUSTNESS OF PARAMETRIC ANALYSIS**

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Estimated w/ Random-Effects Tobit Regression</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(\omega = 1)$</td>
<td>$(\omega = 10)$</td>
<td></td>
</tr>
<tr>
<td>Cost curvature parameter, $\gamma$</td>
<td>1.327</td>
<td>2.168</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td>(.031)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 0.5$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0420</td>
<td>.0013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.0002)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 0.99$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0329</td>
<td>.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.0001)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (noise</td>
<td>$p = 1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0324</td>
<td>.0099</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.0001)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0255</td>
<td>.0008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.0001)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0.01$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0267</td>
<td>.0008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.0001)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1$ (no noise</td>
<td>$p = 0.5$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0213</td>
<td>.0006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.0001)</td>
<td></td>
</tr>
<tr>
<td>$H_0 : \hat{\theta}_1$ (noise</td>
<td>$p = 0.5$) = $\hat{\theta}_1$ (noise</td>
<td>$p = 0.99$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\chi^2(1) = 4.59$</td>
<td>$\chi^2(1) = 5.00$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(p = .032)$</td>
<td>$(p = .025)$</td>
<td></td>
</tr>
<tr>
<td>$H_0 : \hat{\theta}_1$ (no noise</td>
<td>$p = 0.5$) = $\hat{\theta}_1$ (no noise</td>
<td>$p = 0.01$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\chi^2(1) = 4.25$</td>
<td>$\chi^2(1) = 4.65$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(p = .039)$</td>
<td>$(p = .031)$</td>
<td></td>
</tr>
<tr>
<td>Joint test of above</td>
<td>$\chi^2(2) = 8.83$</td>
<td>$\chi^2(2) = 9.45$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(p = .012)$</td>
<td>$(p = .009)$</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4020</td>
<td>4020</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>804</td>
<td>804</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors (in parentheses) are clustered at the individual level and recovered via delta method. 18 observations are left-censored and 43 are right-censored.
### Table A4: Experiment 1. Determinants of Returning for Second Session

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>AMEs</th>
<th>Raw</th>
<th>AMEs</th>
<th>Raw</th>
<th>AMEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1(\text{Noise}))</td>
<td>-0.096</td>
<td>-0.007</td>
<td>-0.122</td>
<td>-0.008</td>
<td>-0.132</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
<td>(0.018)</td>
<td>(0.248)</td>
<td>(0.017)</td>
<td>(0.255)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>(1(\text{Coin Flip}))</td>
<td>0.003</td>
<td>0.000</td>
<td>0.031</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.261)</td>
<td>(0.018)</td>
<td>(0.279)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1(\text{High Probability}))</td>
<td>1.069**</td>
<td>0.062**</td>
<td>1.101**</td>
<td>0.079**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.289)</td>
<td>(0.018)</td>
<td>(0.371)</td>
<td>(0.026)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Demographics

|                   |           |           |           |           |           |           |
|                   | X         | X         |           |           |           |           |

**Constant**

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>AMEs</th>
<th>Raw</th>
<th>AMEs</th>
<th>Raw</th>
<th>AMEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1\text{Noise}))</td>
<td>2.523***</td>
<td>2.700***</td>
<td>2.096***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.184)</td>
<td>(0.229)</td>
<td>(0.586)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations**: 887

**Notes**: Standard errors in parentheses. The control treatment forms the baseline comparison group; demographics includes dummies for income of respondent and gender, and age; none are significant.

* \(p < 0.05\), ** \(p < 0.01\), *** \(p < 0.001\)
Table A5:  
**Experiment 2. The Effect of Extra Tasks in Session 1**

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( (e_{i,1} - e_{i,2}) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dropping Extra Tasks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \dagger ) Noise</td>
<td>-5.443***</td>
<td>(1.885)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) No Noise</td>
<td>4.896*</td>
<td>(2.685)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \dagger ) Noise</td>
<td>-5.443***</td>
<td>(1.882)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) No Noise</td>
<td>4.896*</td>
<td>(2.681)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) (Extra Tasks) * ( \dagger ) (Noise)</td>
<td>3.351</td>
<td>(3.730)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) (Extra Tasks) * ( \dagger ) (No Noise)</td>
<td>8.431</td>
<td>(5.164)</td>
<td></td>
</tr>
<tr>
<td><strong>IV using BDM as instrument</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \dagger ) Noise</td>
<td>-8.010*</td>
<td>(4.629)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) No Noise</td>
<td>11.181**</td>
<td>(4.428)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) (Extra Tasks) * ( \dagger ) (Noise)</td>
<td>10.459</td>
<td>(11.765)</td>
<td></td>
</tr>
<tr>
<td>( \dagger ) (Extra Tasks) * ( \dagger ) (No Noise)</td>
<td>-12.137</td>
<td>14.176</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>360</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Notes:* Standard errors (in parentheses) clustered at individual level. All regressions include random effects at individual level. Instruments are random number from BDM and dummies for the randomly selected question (five such variables).

* *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01
Table A6:
EXPERIMENT 2. DETERMINANTS OF RETURNING FOR SECOND SESSION

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>AMEs</th>
<th>Raw</th>
<th>AMEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Noise)</td>
<td>-0.134</td>
<td>-0.019</td>
<td>-0.129</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.568)</td>
<td>(0.080)</td>
<td>(0.659)</td>
<td>(0.080)</td>
</tr>
<tr>
<td>Avg WTW, Session 1</td>
<td></td>
<td></td>
<td>-0.006</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>1 (Russian, Session 1)</td>
<td></td>
<td></td>
<td>0.436</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.674)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.638***</td>
<td></td>
<td>0.803</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.413)</td>
<td></td>
<td>(1.134)</td>
<td></td>
</tr>
<tr>
<td>Session Dummies</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>87</td>
</tr>
</tbody>
</table>

Notes: Standard error in parentheses.  * p < 0.05, ** p < 0.01, *** p < 0.001
Figure A1: *Raw willingness-to-work data from Experiment 2.* Each observation in this figure represents a participant’s willingness to work for a fixed payment in sessions one and two of the experiment. Black dots represent participants who faced the no-noise task; red diamonds represent participants who faced the noisy task.
Figure A2: *Histogram of the difference in willingness-to-work between the first and second sessions in Experiment 2*. Each observation in this figure represents the change in a participant’s willingness to work for a fixed payment between sessions one and two of the experiment. Clear bars represent participants who faced the no-noise task; solid red bars represent participants who faced the noisy task.

F Experimental Instructions

In this section, we provide the full text of our experimental instructions. We use brackets to denote alternative instructions corresponding to different treatments. All instructions commenced with an informed consent form. The research in this study was reviewed by the Human Research Protection Program at Harvard University (protocol numbers: IRB15-0365 and IRB16-0944).

---

57 The Nock Lab at Harvard generated the noise used in our experiments. They used the stimuli in work unrelated to our own. In their studies, this sound was played at modest volume (slightly louder than we played the noise). Participants in their (more extensive) studies found the sound unpleasant, but with no lasting effects (e.g., ringing ears).
F.1 Sample Reviews, Experiment 1

For a full text of the reviews used in Experiment 1, please contact the authors.

“To read this book is to go on a journey to places at once unexpected yet familiar; for example, one point is supported by reference to a diagram of nose shapes and sizes. His books teach rather than exposit; they do not lack for a direct thesis–they make arguments and reach conclusions.”

Score: 5; Positive Review

“Sometimes you don’t go out and find a book; the book finds you. Facing an impending loss without a foundation of faith to fall back on, I asked myself: ‘What is the meaning of life if we’re all just going to die?’ The author answers that question in the most meaningful way possible.”

Score: 5; Positive Review

“To be sure, this is a very quick read. The book is already very tiny, and the inside reveals large font and double spacing. It took me about two hours to finish this book. I believe I am an somewhat slow reader compared to other bookworms. On the other hand, I found many other books to be much more compelling and memorable takes on the meaning of life.”

Score: 1; Negative Review

“Sometimes books like this are a real bore. Even worse, sometimes the science is terrible or inconsistent. I was pleased to find that this book is consistent with the established literature while also providing new insight.”

Score: 5; Positive Review

“This book is nothing you expect it to be. I was looking forward to fun, witty tales of some of the author’s romances. But no. He teamed up with a sociologist, and wrote a sociology textbook. It’s bland and it’s boring, with research percentages and the odd pie chart thrown in to liven things up.”

Score: 1; Negative Review

F.2 Complete Experiment Instructions: Experiment 1

Session 1

We will begin with some simple demographic questions. What is your gender? □ Male □ Female

What is your annual income?

□ less than $15,000
□ $15,000 - $29,999
□ $30,000 - $59,999
□ $60,000 - $99,999
□ $100,000 or more

What is your age (in years)?
What is your zip code? [Format: 00000]

We will not deceive you whatsoever in this experiment. All of the instructions provide examples and guidance for the actual tasks you will do. There will be no surprises or tricks. This study will consist of two sessions. You will do the first session now. You will sign in to do the second session later. In each session, you will do a simple job that takes roughly 3 to 5 minutes. You will earn a fixed payment of $4 for completing both sessions. In the second session, you will have the chance to earn extra pay if you elect to do extra work. You must complete both sessions to earn any pay for this study. There will be absolutely no exceptions to this rule. All payments will be credited to your MTurk account within one week of completing the study.

The second session will be unlocked 8 hours after the first session. In order to unlock the second session, a link will be emailed to you. We ask that you complete the second session as soon as you are able to. You must complete the second session within one week of the email in order to receive payment.

Your task in both sessions will be listening a series of audio recordings of book reviews (from Amazon) to determine whether each review is generally positive or negative.

You must wait at least 10 seconds before any buttons will appear. You must then decide if the review is positive or negative. A positive review means that the reviewer generally liked the book and is providing a recommendation. A negative review means that the reviewer generally disliked the book and is cautioning against reading it.

We will now give you a sample task to practice. Once you have listened to the review and correctly determined if it is a positive or negative review, please close the pop-up window and click the arrow below to continue. Please click the link below for a sample of the task. [LINK]

During each of the two participation sessions, you will have to complete eight tasks. Note: the average time of each recording is about 20 seconds.

During the eight required reviews, you cannot get more than two answers wrong. If you get more than two answers wrong, you will be dropped from the study and will not receive payment. However, if you listen to the entire audio recording, the answers should be quite easy.

During the second session, we will ask you about your willingness to do additional reviews for extra pay. Your job in this first session is to learn about the difficulty of the task and think about your willingness to do additional reviews next session.

[Coin flip: Depending on chance, a background noise may be played on top of the audio review. We’ll describe what determines whether you hear the noise in a moment. However, we’d like to make sure you know what the sound will be. Please click the play button below for a sample of the noise. When you are finished listening to the sample noise, click the arrow below to continue.]

[Coin flip: In a moment, you will begin the eight initial reviews. Before that, however, we must determine if you will have to hear the annoying noise over the audio review. In order to do this, you will flip a (digital) coin. If the coin lands Heads, you will not have to hear the noise. If it lands Tails, you will have to hear the noise.]

[Coin flip: Importantly, your flip today determines what you’ll do on the second session of the experiment. If the coin flip lands Tails and you hear the annoying noise today, you will also hear it next session. If the coin flip lands Heads and you do not hear the annoying noise today, you will not hear it next session. So the result of this coin flip really matters!]

Click the button below to flip the coin: [BUTTON]

Sorry [Congratulations]. You will [not] have to hear the noise while you listen to the audio reviews. We will now begin the eight initial tasks. At the end of the task, you will see a code. You
will need that code to continue. Click the words below to begin. [BEGIN TASK]

Remember - this experiment has two parts. The link to the second session will be emailed to you in 8 hours.

Since you heard [did not hear] the annoying noise today, you will also hear it next session. Please click the arrow to submit your work.

Session 2

Welcome to the second session of the experiment.

As with the first session, if you choose not to participate in the study, you are free to exit. You must finish this session in order to receive payment. As a reminder: we will not deceive you whatsoever in this experiment. All of the instructions provide examples and guidance for the actual tasks you will do. There will be no surprises or tricks.

As with last session, you will listen to an audio recording of a review and must determine whether the reviewer is giving a generally positive or negative review. Be careful to listen to the whole review!

You heard [did not hear] the noise on top of the audio last session, and you will [not] hear it again this session. [Noise only: If you need a reminder of the noise, there is a sample below. To play, click the play button twice.]

As before you will have to complete eight reviews. However, this session you will have the option to complete extra reviews for additional payments. These extra tasks will come after the eight initial reviews. You will first decide how many extra reviews you would like to do on top of the eight initial reviews. You will then do the first eight reviews. Finally, you will have a chance to complete extra reviews if you were willing to do so. We will describe how this is determined on the next slides.

The method we use to determine whether you will complete extra reviews may seem complicated. But, we’ll walk through it step-by-step. The punchline will be that it’s in your best interest to just answer truthfully. First, we will ask you how many additional reviews you are willing to do for a fixed amount of money. For instance, we might ask: ”What is the maximum number of extra reviews you are willing to do for $0.40?” This question means that we will give you $0.40 in exchange for you completing some amount of additional work.

On the decision screen, you will be presented a set of sliders that go between 0 and 100 tasks. You will also see an amount of money next to each slider. You will move each slider to indicate the maximal number of reviews you’d be willing to do for each amount of money. That is, if you would be willing to do 15 additional reviews but not 16, then you should move the slider to 15.

You will make five decisions, but only one will count for real. We will choose which decision counts for real using a random number generator. Therefore, it is in your best interest to take each question seriously and choose as if it were the only question.

Once we determine which question counts for real, we will draw a random number between 0 and 100. If your answer is less than that random number, you will not do additional reviews. However, if your answer is greater than or equal to that random number, you will do a number of additional tasks equal to the random number.

Example: Suppose you indicated you were willing to do 15 additional reviews for $0.40 and this question was chosen as the one that counts. If the random number was 16 or higher, you would do no additional tasks. However, if the random number was 12, you would do 12 additional reviews.
The next pages have a short quiz to help clarify how this works.

Suppose you were asked "What is the maximum number of additional reviews you are willing to do for $0.80?" and you responded 60. If the random number is 17, how many reviews will you complete?

- □ 0 and I will be paid $0 in supplementary payments
- □ 60 and I will be paid $0.80 in supplementary payments
- □ 17 and I will be paid $0.80 in supplementary payments
- □ 17 and I will be paid $2.67 in supplementary payments

[On answering correctly] Correct. You will earn the extra payment if the random number is less than the number you indicated, and you will complete a number of additional reviews equal to the random number.

Suppose you were asked "What is the maximum number of additional reviews you are willing to do for $0.80?" and you responded 60. If the random number is 76, how many additional reviews will you complete?

- □ 0 and I will be paid $0 in supplementary payments
- □ 76 and I will be paid $0.80 in supplementary payments
- □ 60 and I will be paid $0.80 in supplementary payments
- □ 76 and I will be paid $0 in supplementary payments

[On answering correctly] Correct. If the random number is greater than your choice, you will complete zero reviews and you will not receive an extra payment. This method of selecting how many additional reviews you will do might seem very complicated, but as we previously highlighted, there’s a great feature to it: your best strategy is to simple answer honestly. If, for example, you’d be willing to do 20 reviews for $0.40 but not 21, then you should answer 20. You may very well do less than 20 reviews (depending on the random number) but you certainly will not do more than 20. Put simply: just answer honestly.

Remember, you will decide whether to do additional reviews, then complete the eight initial reviews. Then we will draw a random number which determines if you will do extra reviews.

We will now ask you the questions about your willingness to do additional reviews for additional payment. Remember, we are using the method just described, so answer honestly. These are the real questions. One of the sliders will count for payment, so pay close attention.

What is the maximal number of additional reviews you’re willing to complete for:

- $2.50? [SLIDER]
- $2.00? [SLIDER]
- $1.50? [SLIDER]
- $1.00? [SLIDER]
- $0.50? [SLIDER]

We will determine whether you will do additional reviews after you complete the eight initial tasks. We will begin those on the next page.

Like last session, you will [not] have to hear the noise during the audio reviews. We will now begin the eight initial reviews. When you have completed these eight reviews, you will see a code. You will need that code to continue. Click the words below to begin. [BEGIN TASK]

We’ll now draw the random number that determines which question counts for payment.

The random number selected the question where you were asked the maximum number of tasks you would do for [AMOUNT]. You answered [RESPONSE]. We’ll now draw a second random number that determines whether you do additional tasks and, if so, how many.
The random number is: [RANDOM NUMBER]. You answered: [RESPONSE].

[Random number too high: Since the random number was higher than the number you were willing to do, you will not complete any extra reviews and you will not receive any extra payments.] Since the random number was lower than the number you were willing to do, you will complete extra reviews. You will do [RANDOM NUMBER] extra reviews and receive [AMOUNT]. In order to verify that you completed all the additional reviews, we will give you a code when you finish.

[BEGIN SUPPLEMENTAL TASKS]

Thank you for participating. Your MTurk code is on the screen that follows. Payments will be processed within one week. Please click the final button below to submit your work.

High-Probability Treatment Modified Lines

The High probability treatment used the same instructions as above, except the paragraphs labeled Coin flip were replaced with the following:

[High Probability: In a moment, you will begin the eight initial reviews. Before that, however, we must determine if you will have to hear the annoying noise over the audio review. In order to do this, we will draw a random number from 1-100. If the random number is 100, you will not have to hear the noise. If it is any other number, you will have to hear the noise.]

[High Probability: Importantly, the random number today determines what you’ll do on the second session of the experiment. If the number is 1-99 and you hear the annoying noise today, you will also hear it next session. If the random number is 100 and you do not hear the annoying noise today, you will not hear it next session. So the result of this random draw really matters!]]

F.3 Complete Experiment Instructions: Experiment 2

Session 1

In front of you is an informed-consent form to protect your rights as a participant. Please read it. If you choose not to participate in the study, you are free to leave at any point. If you have any questions, we can address those now. We will pick up the forms after the main points of the study are discussed.

We will not deceive you whatsoever in this experiment. All of the instructions provide examples and guidance for the actual tasks you will do. There will be no surprises or tricks. If you have any questions at any time, please raise your hand and we will do our best to clarify things for you.

In this experiment, you will have the chance to earn supplemental payments ranging from $2-$25/hour. It is very important for the study that you participate in both days. Unfortunately, if you miss one of your participation dates, you will forgo any completion payments and supplemental payments and will be removed from the study (you will receive the show-up fee). There will be absolutely no exceptions to this rule, regardless of the reason. Completion and supplemental payments will be made as one single payment in cash at the end of the study.

Your task will be transcribing a line of handwritten text in a foreign language. We will explain the task and then allow you to spend a few moments practicing this job on the computer. Note that the example text may not exactly match what you will face in the experiment.

Letters will appear in a Transcription Box on your screen. For each handwritten letter, you will need to enter the corresponding letter into the Completion Box. In order to enter a letter into the
Completion Box, simply click the letter from the provided alphabet. We refer to one row of text is one task. In order to advance to the next task, your accuracy must be above 90%.

We will now give you a sample task to practice. You will see handwritten characters and must enter the corresponding character into the Completion Box by clicking on the appropriate button. When you have transcribed a whole row, press "Submit". You may spend as much time as you like transcribing the text. If you succeed, a new line of text will appear. Once you have transcribed one row successfully, please close the pop-up window and click the arrow below to continue. Please click the link below for a sample of the task. [SAMPLE TASK]

During each of the two participation days, you will have to complete five tasks (five lines of foreign text). Note: the average time to complete a similar task in a different experiment was about 52 seconds (about 70 tasks/hour).

After completing five initial tasks, you will have the option to complete additional supplementary tasks for supplementary payments. The number of supplementary tasks you must complete on each participation day and the supplementary payment will depend on your own willingness to work. The supplementary tasks will come shortly after the five initial tasks.

In order to determine whether you will complete additional tasks, we will ask you how many additional tasks you are willing to do for a fixed amount of money. For instance, we might ask: "What is the maximum number of additional tasks you are willing to do for $5?" This question means that we will give you $5 in exchange for you completing some amount of additional work. The next few screens describe a pretty complicated system that will determine how many additional tasks you actually do. But the point of this system is simple: there is no way to game the system. It is in your best interest to answer honestly.

On the decision screen, you will be presented a set of sliders that go between 0 and 100 tasks. You will also see an amount of money next to each slider. You will move each slider to indicate the maximal number of tasks you’d be willing to do for each amount of money. That is, if you would be willing to do 15 additional tasks but not 16, then you should move the slider to 15. For example (you need not enter anything) What is the maximal number of additional tasks you’re willing to complete for:

$1? [SLIDER]
$2? [SLIDER]
$3? [SLIDER]
$4? [SLIDER]
$5? [SLIDER]

You will make five decisions, but only one will count for real. We will choose which decision counts for real using a random number generator. Therefore, its in your best interest to take each question seriously and choose as if it was the only question.

Once we determine which question counts for real, we will draw a random number between 0 and 100. If your answer is less than that random number, you will do no additional tasks. However, if your answer is greater than or equal to that random number, you will do a number of additional tasks equal to the random number.

Example: Suppose you indicated you were willing to do 15 additional tasks for $5 and this question was chosen as the one that counts. If the random number was 16 or higher, you would do no additional tasks. However, if the random number was 12, you would do 12 additional tasks. The next page has a short quiz to help clarify this system.

Suppose you were asked ”What is the maximum number of additional tasks you are willing to do
for $10?" and you responded 30. If the random number is 8, how many tasks will you complete?

- □ 0 and I will be paid $0 in supplementary payments
- □ 8 and I will be paid $10 in supplementary payments
- □ 8 and I will be paid $2.67 in supplementary payments

Correct. You will be paid the full amount regardless of the random number, and if the random number is less than the number you indicated, you will only need to complete a number of additional tasks equal to the random number.

Suppose you were asked "What is the maximum number of additional tasks you are willing to do for $10?" and you responded 30. If the random number is 46, how many additional tasks will you complete?

- □ 0 and I will be paid $0 in supplementary payments
- □ 46 and I will be paid $10 in supplementary payments
- □ 0 and I will be paid $10 in supplementary payments
- □ 30 and I will be paid $0 in supplementary payments

Correct. If the random number is greater than your choice, you will complete zero tasks and you will not get paid. This method of selecting how many additional tasks you will do might seem very complicated, but as we previously highlighted, there’s a great feature to it: your best strategy is to simple answer honestly. If you’d be willing to do 20 tasks for $5 but not 21, then you should answer 20. You may very well do less than 20 tasks (depending on the random number) but you certainly will not do more than 20. Put simply: just answer honestly.

Depending on chance, a background noise may be played throughout the transcription process. We’ll describe what determines whether you hear the noise in a moment. However, we’d like to make sure you know what the sound will be. Please click the play button below twice for a sample of the noise. When you are finished listening to the sample noise, click the arrow below to continue.

In a moment, you will begin the five initial tasks. Before that, however, we must determine if you will have to hear that annoying noise during the whole transcription process. In order to do this, you will flip a coin. If the coin lands Heads, you will not have to hear the noise. If it lands Tails, you will have to hear the noise.

Importantly, your flip today determines what you’ll do on the second day of the experiment. If the coin flip lands Tails and you hear the annoying noise today, you will also hear it next week. If the coin flip lands Heads and you do not hear the annoying noise today, you will not hear it next week. So the result of this coin flip really matters!

When you reach this screen, please put your hand up. You may remove your headphones for this stage of the instructions. One of the experimenters will come by and help you. We are using a standard U.S. Quarter. This is not a trick coin and we’re going to ask you to flip it. Please flip it and let it land on the table in front of you. If the coin does not flip more than twice, we will ask you to flip again. You’ll be asked to flip a practice flip, and then you’ll flip the one that counts. Reminder: Heads → No Noise. Tails → Annoying Noise

The experimenter will the answer this question.

- □ Tails
- □ Heads

Enter Code to Advance

[Noise: You will have to hear the noise. Please put your headphones back on. We will now
begin the five initial tasks.] You will not have to hear the noise. However, we ask that you please
put your headphones on so that you do not hear others. At the end of the task, you will see a code.
You will need that code to continue. Click the words below to begin. [BEGIN TASK] Please enter
the code below to continue

We will now ask you some questions about your willingness to do additional tasks for additional
payment. Remember, we are using the system described earlier, so answer honestly. One of the
sliders will count for real payment, so pay close attention.
What is the maximal number of additional tasks you’re willing to complete for:
$20? [SLIDER]
$16? [SLIDER]
$12? [SLIDER]
$8? [SLIDER]
$4? [SLIDER]

We’ll now draw a random number to determine which question counts for payment.
The random number selected the question where you were asked the maximum number of tasks
you would do for [AMOUNT]. You answered [RESPONSE]. We’ll now draw a second random
number that determines whether you do additional tasks and, if so, how many.
The random number is: [RANDOM NUMBER]. You answered: [RESPONSE].
[Random number too high: Since the random number was higher than the number you were
willing to do, you will not complete any extra reviews and you will not receive any extra payments.]
Since the random number was lower than the number you were willing to do, you will complete
extra reviews. You will do [RANDOM NUMBER] extra reviews and receive [AMOUNT]. In order
to verify that you completed all the additional reviews, we will give you a code when you finish.
[BEGIN SUPPLEMENTAL TASKS]

Thank you for participating. [Noise: REMINDER: Since you heard the annoying noise today,
you will also hear it in a week.]

REMEMBER: Since you did not hear the annoying noise today, you will not hear it in a week.

Day 1 of the experiment is complete. Please return at the same time one week from now. Please
click the arrow to submit your work. When you have finished, you may exit the lab.

Session 2

Welcome to the second day of the experiment.
Please turn your cell phones off. If you have a question at any point in the experiment, please
raise your hand and a lab assistant will be with you to help. There will be a short quiz once we
have finished the instructions. If you do not understand the instructions after both the instruction
period and the quiz, please raise your hand and ask for help.
As with the first day, if you choose not to participate in the study, you are free to leave at any
point. If you have any questions, we can address those now.
As a reminder: we will not deceive you whatsoever in this experiment. All of the instructions
provide examples and guidance for the actual tasks you will do. There will be no surprises or
tricks.
Like last week, your task is to transcribe a line of handwritten letters from a foreign language.
This week, you will do a different language. You will the task under the same conditions as last
week.
Noise: You heard the noise last week, and you will hear it again this week. If you need a reminder of the noise, there is a sample below. To play, click the play button twice.

You did not hear the noise last week, and you will not hear it again this week.

As with last week, letters will appear in a Transcription Box on your screen. For each handwritten letter, you will need to enter the corresponding letter into the Completion Box. In order to enter a letter into the Completion Box, simply click the letter from the provided alphabet. We refer to one row of text is one task. In order to advance to the next task, your accuracy must be above 90%.

As before you will have to complete five tasks (five lines of foreign text) and then you will have the option to complete additional supplementary tasks for supplementary payments. The supplementary tasks will come shortly after the five initial tasks.

In order to determine whether you will complete additional tasks, we will ask you how many additional tasks you are willing to do for a fixed amount of money. For instance, we might ask: "What is the maximum number of additional tasks you are willing to do for $5?" This question means that we will give you $5 in exchange for you completing some amount of additional work. It is in your best interest to answer these questions honestly.

Recall we used a random number system to determine how many additional tasks you did (if any). We’ll provide a quick reminder of that system now.

On the decision screen, you will be presented a set of sliders that go between 0 and 100 tasks. You will also see an amount of money next to each slider. You will move each slider to indicate the maximal number of tasks you’d be willing to do for each amount of money. That is, if you would be willing to do 15 additional tasks but not 16, then you should move the slider to 15.

You will make five decisions, but only one will count for real. We will choose which decision counts for real using a random number generator. Therefore, it’s in your best interest to take each question seriously and choose as if it was the only question.

Once we determine which question counts for real, we will draw a random number between 0 and 100. If your answer is less than that random number, you will do no additional tasks. However, if your answer is greater than or equal to that random number, you will do a number of additional tasks equal to the random number.

Example: Suppose you indicated you were willing to do 15 additional tasks for $5 and this question was chosen as the one that counts. If the random number was 16 or higher, you would do no additional tasks. However, if the random number was 12, you would do 12 additional tasks. The next page has a short quiz to help clarify this system.

Suppose you were asked "What is the maximum number of additional tasks you are willing to do for $10?" and you responded 60. If the random number is 17, how many tasks will you complete?

☐ 0 and I will be paid $0 in supplementary payments
☐ 60 and I will be paid $10 in supplementary payments
☐ 17 and I will be paid $10 in supplementary payments
☐ 17 and I will be paid $2.67 in supplementary payments

Correct! You will be paid the full amount regardless of the random number, and if the random number is less than the number you indicated, you will complete a number of additional tasks equal to the random number.

Suppose you were asked "What is the maximum number of additional tasks you are willing to do for $10?" and you responded 60. If the random number is 76, how many additional tasks will you complete?

☐ 0 and I will be paid $0 in supplementary payments
☐ 76 and I will be paid $10 in supplementary payments
☐ 60 and I will be paid $10 in supplementary payments
☐ 76 and I will be paid $0 in supplementary payments

Correct. If the random number is greater than your choice, you will complete zero tasks and
you will not get paid. This method of selecting how many additional tasks you will do might seem
very complicated, but as we previously highlighted, there’s a great feature to it: your best strategy
is to simply answer honestly. If you’d be willing to do 20 tasks for $5 but not 21, then you should
answer 20. You may very well do less than 20 tasks (depending on the random number) but you
certainly will not do more than 20. Put simply: just answer honestly.

[Noise: Like last week, you will have to hear the noise. Please put your headphones back on.] Like last week, you will not have to hear the noise. However, we ask that you please put your
headphones on so that you do not hear others. We will now begin the five initial tasks. At the end
of the task, you will see a code. You will need that code to continue. Click the words below to
begin. [BEGIN TASK] Please enter the code below to continue:

We will now ask you some questions about your willingness to do additional tasks for additional
payment. Remember, we are using the system described earlier, so answer honestly. One of the
sliders will count for real payment, so pay close attention.

What is the maximal number of additional tasks you’re willing to complete for:
$20? [SLIDER]
$16? [SLIDER]
$12? [SLIDER]
$8? [SLIDER]
$4? [SLIDER]

We’ll draw a random number to determine which question counts for payment.
The random number selected the question where you were asked the maximum number of tasks
you would do for [AMOUNT]. You answered [RESPONSE]. We’ll now draw a second random
number that determines whether you do additional tasks and, if so, how many.

The random number is: [RANDOM NUMBER]. You answered: [RESPONSE].

[Random number too high: Since the random number was higher than the number you were
willing to do, you will not complete any extra reviews and you will not receive any extra payments.] Since the random number was lower than the number you were willing to do, you will complete
extra reviews. You will do [RANDOM NUMBER] extra reviews and receive [AMOUNT]. In order
to verify that you completed all the additional reviews, we will give you a code when you finish.

[BEGIN SUPPLEMENTAL TASKS]

Thank you for participating. As you know, the experiment consisted of two days. Our main
hypothesis was whether the chance of getting a different task on the first day changed your per-
ceptions of the task difficulty that day. We did not highlight this specific hypothesis during the
experiment in order to maintain the external validity of the study. We’re excited to analyze the data
and thank you again for your participation. Click the arrow to submit your work.