An Experiment on Interpersonal Projection Bias

Benjamin Bushong
Michigan State University

Tristan Gagnon-Bartsch*
Harvard University

November 27, 2020

Abstract

Using a real-effort experiment, we show that people project their current tastes onto others, even when others' tastes are exogenously manipulated and transparently different. In a first stage, "workers" stated their willingness to continue working on a tedious task. We varied how many initial tasks workers completed before eliciting their willingness to work (WTW): some were relatively fresh when stating their WTW, while others were relatively tired. Later, a separate group of "predictors"—who also worked on the task—guessed the WTW of workers in each state. We find: (i) tired workers were less willing to work than fresh workers; (ii) predictors (in aggregate) accurately forecasted the WTW of workers when they cast their predictions in the same state as the workers about whom they were predicting, but, (iii) when predictors were fresh but guessing about tired workers, they substantially overestimated the WTW of tired workers, and (iv) when predictors were tired but guessing about fresh workers, they underestimated the WTW of fresh workers. Using an additional treatment, we find that workers also mispredicted their own future WTW and that this "intrapersonal" projection bias is likely less severe than "interpersonal" projection bias.

JEL Classification: D91, D84.
Keywords: Projection bias, Real-effort experiments, State-dependent preferences.

*E-mails: bbushong@msu.edu and gagnonbartsch@fas.harvard.edu. We thank Jon Eguia, Christine Exley, Kristof Madarász, Matthew Rabin, Antonio Rosato, Josh Schwartzstein, and audiences at the ESA World Conference in Vancouver, Michigan State University, and the University of Michigan for comments.

1 Introduction

> "When you are asked to 'put yourself in someone's place', what is the implied contrasting condition: what is it that you are implicitly being asked *not* to do? . . . [W]hat is implied is that you shouldn't just *project your own* situation and psychology on the other."
>
> — Robert M. Gordon

Forecasting others' preferences is a ubiquitous part of economic behavior. For example, bidding in an auction requires a model of the other bidders' values, negotiating demands an understanding of the counterpart's motivations, and providing financial or career advice calls for an appreciation of the advisee's objectives. In the workplace, a manager attempting to optimally allocate projects amongst her workers must be mindful of their recent workloads—fatigue will diminish the workers' productivity and willingness to take on additional projects. But forecasting others' preferences is challenging. Evidence suggests that people make systematic errors even when predicting how their own preferences will change—they tend to project their current desires into the future, even though their future self may have (predictably) differing tastes (see, e.g., Read and van Leeuwan, 1998; Badger et al., 2007; Conlin, O'Donoghue, and Vogelsang, 2007; Busse et al., 2015). In this paper, we present experimental evidence that a similar error extends to predictions about other people's tastes.

In our experiment, participants worked on a tedious real-effort task. In a first stage, we elicited participants' willingness to continue working on the task for additional pay. In a second stage, different participants cast incentivized predictions about the willingness to work (WTW) of the first group. We call these two groups "workers" and "predictors", respectively. Critically, we varied how many initial tasks workers completed before eliciting their willingness to work: half completed five tasks—they were relatively fresh when stating their WTW—while the other half completed twenty—they were relatively tired. Predictors also worked on the task, and we similarly varied the number of initial tasks that predictors completed: some did five tasks before guessing the WTW of others—they were relatively fresh when making predictions—while others did twenty—they were relatively tired when making predictions. Our central question is whether (and to what extent) predictors abstracted from their own tiredness state when making guesses about others' preferences.

To elucidate the mechanism driving these guesses, we divided the predictors into three subgroups. These groups varied in how many guesses each predictor cast and the state they were in when they made each guess. Some predictors made guesses about the WTW of fresh workers when they themselves were fresh and, later, made guesses about tired workers when they themselves were tired. Other predictors made guesses when they were "out of phase" with the workers: they guessed about tired workers when they themselves were fresh and guessed about fresh work-

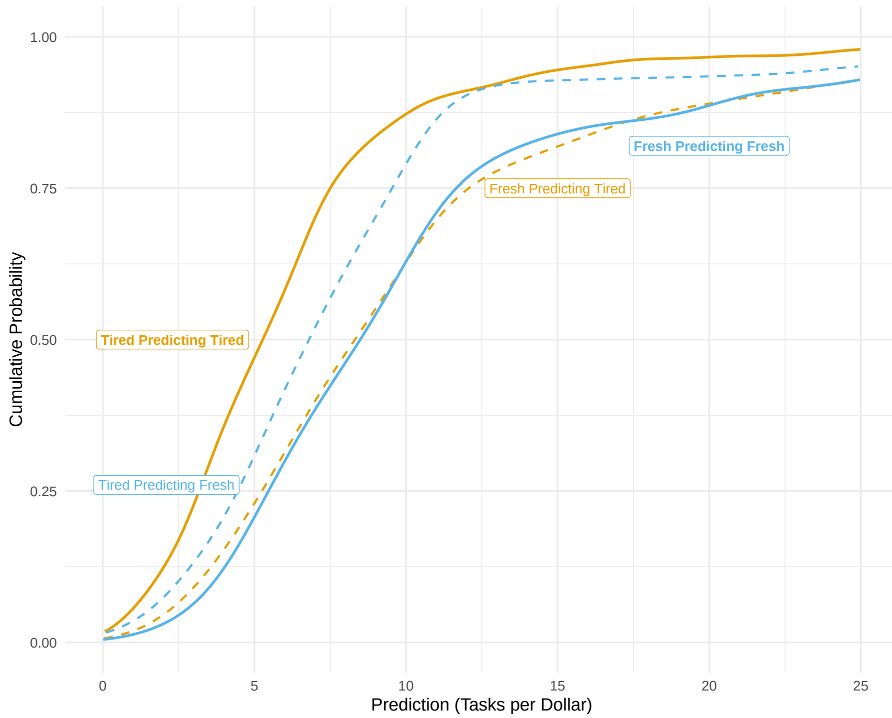

Figure 1: *Empirical (smoothed) CDFs of predictions about fresh (blue) and tired (orange) workers. Predictions are shown in units of "tasks-per-dollar"—the average number of tasks workers are willing to do for each dollar of compensation. (As we discuss below, this normalization is solely to aid the visualization of our results.) The solid lines represent guesses cast by predictors in the same state as the workers about whom they were predicting, while the dotted lines represent the guesses cast by predictors in the other state. Both dotted distributions are significantly different from the relevant solid distributions ($p < .001$ for both; Wilcoxon rank-sum test).*

ers when they themselves were tired. Comparing predictions across these two groups allows us to measure how a predictor's beliefs changed as their own state changed. Finally, a third group made their first prediction—about fresh workers—when they themselves were tired. That is, they made no prediction when they were fresh. This group allows us to further control for anchoring or other order effects by comparing only the initial predictions across groups.

We find five stylized facts about behavior and predictions in our primary experiment: (i) Relatively tired workers were less willing to work than fresh workers. (ii) Predictors (in aggregate) accurately forecasted the WTW of workers when they cast their predictions in the same state as the workers about whom they were predicting. (iii) When predictors were fresh but guessing about tired workers, they substantially overestimated the WTW of tired workers. (iv) When predictors were tired but guessing about fresh workers, they underestimated the WTW of fresh workers. (v) Fresh predictors made a larger error when guessing about tired workers than tired predictors made when guessing about fresh workers.

We highlight two particular results that cleanly support the idea that projection bias underlies the results described above. First, although each predictor made multiple guesses, one of our analyses uses only their initial predictions. This eliminates concerns regarding order effects. We find that results (iii) and (iv) remain: fresh predictors overestimated the WTW of tired workers by approximately 50%, while tired predictors underestimated the WTW of fresh workers by approximately 21%. Figure 1 previews this result by showing the distribution of initial guesses cast by predictors in a different state than workers relative to those cast by predictors in the same state. Moreover, this approach allows us to decompose predictors' erroneous guesses into two components—one due to projection bias, and another due to uncertainty about how onerous the task would become over time. We find that these two components distorted the guesses of fresh predictors by similar magnitudes.

Second, we examine how an individual predictor's guesses changed as they became more tired. When predictors first guessed the WTW of fresh workers when they themselves were fresh, they were, on average, accurate. However, when they performed this same prediction again (i.e., about *fresh* workers) when they themselves became tired, they substantially revised their guesses downward (by approximately 19%; difference significant at $p < .001$). We further show that this was a mistake. By revising their guesses, predictors significantly decreased their accuracy and thus lowered their expected earnings.

Overall, our results paint a consistent picture: participants in our experiment projected their sense of tiredness onto others. We find—through both non-parametric measures and the estimates of a parametric model—that projection induced large and significant errors when predicting the choices of others. Although predictions were accurate when guesses were about others in one's own state, our various analyses suggest that guesses about others in a different state were, on average, wrong by between 21% and 39%.

Finally, we extend the existing literature on projection bias by comparing the magnitude of interpersonal and *intra*personal projection—the tendency for a person's current state to overly influence predictions about their own future behavior. To measure intrapersonal projection bias in the same experimental setting, we ran an additional worker treatment in which workers in the fresh state predicted their own future WTW in the tired state. On average, these fresh workers overestimated their WTW in tired state by approximately 30%. For a comparable measure of interpersonal projection, fresh predictors overestimated the WTW of others in the tired state by approximately 50%. This suggests that interpersonal projection bias may be more severe than the intrapersonal analog.

Despite its potential ubiquity, interpersonal projection bias has received little attention in the economics literature. On the empirical side, Loewenstein and Adler (1995) and Van Boven, Loewenstein, and Dunning (2003) find that sellers in experimental markets project their sense of endowment onto potential buyers; Ambuehl, Bernheim, and Ockenfels (2019) find that subjects in a paternalistic role project their aspirations onto others; and Engelmann and Strobel (2000, 2012) provide experimental tests of the "false-consensus effect"—the tendency to exaggerate the similarity between one's own actions or opinions and those of others. In the following section, we discuss these related studies—and the larger psychology literature on projection—in more detail, and we describe the variety of ways in which our study builds on this literature.[1]

A few of these ways are worth emphasizing here. First, since our predictors completed the same task (in the same quantities) as our workers, we limit information-based explanations that have challenged the interpretation of previous studies. That is, our design reduces the possibility that biased predictions arose simply because predictors were unfamiliar with the experience of being in the fresh or tired state. This highlights a second feature of our design: we transparently induced changes in tastes along a familiar dimension (tiredness).[2] This further limits the scope for informational explanations of our findings. Third, we elicited multiple predictions from each participant, which allows us to highlight how a person's beliefs changed as their own tiredness state changed. Finally, we are—to our knowledge—the first study to measure both inter- and intrapersonal projection bias in the same domain. This allows us to compare the relative magnitudes of these errors.

Our results highlight the potential benefits from greater engagement with the perspectives of others, particularly in domains involving effort provision and fatigue. Returning to our opening example, a manager who suffers the bias we document will systematically underestimate the heterogeneity in her workers' marginal disutility of effort; she may therefore neglect the benefits from strategically tailoring new assignments across her workers based on their recent workloads. Similarly, a biased manager may fail to derive the optimal contract for a worker given her misunderstanding of how the worker's incentives will change over time. This form of projection bias can also have consequences for social judgments, as an observer may too quickly attribute a person's limited effort to intrinsic characteristics (e.g., laziness) rather than their momentary burden.

Moreover, a growing theoretical literature emphasizes the broader implications and importance of the sort of interpersonal projection bias we find. For instance, the projection of political preferences can generate inefficient election outcomes because voters miscalculate the probability of being pivotal (Goeree and Grosser, 2007). The projection of idiosyncratic tastes can lead to overbidding and inefficient allocations in auctions (Gagnon-Bartsch, Pagnozzi, and Rosato, 2020). In

[1] Information projection (e.g. Camerer, Loewenstein and Weber, 1989; Madarász, 2012) is an alternative form of interpersonal projection. This concept captures people's tendency to assume that others share their private information (rather than assuming that others share their preferences). We discuss this concept further—specifically, how information projection does not explain our results—in the conclusion.

[2] In this way, we follow Augenblick and Rabin (2019), who measure *intra*personal projection bias over effort provision in a similar domain.

social-learning contexts, such as the adoption of a new technology, mispredicting others' tastes can prevent people from inferring their optimal action even in settings where rational agents would learn correctly (Gagnon-Bartsch, 2016; Bohren and Hauser, 2020; Frick, Iijima, and Ishii, 2020). Finally, Kaufmann (2019) shows how *intra*personal projection bias over effort can lead people to over-commit, over-work, and experience burnout in contexts similar to our experimental setting.

The paper proceeds as follows. In Section 2, we discuss the existing evidence on interpersonal projection, highlighting how our design builds on previous studies and mitigates important confounds. In Section 3, we provide a detailed description of our experimental design and—using a model that extends Loewenstein, O'Donoghue, and Rabin (2003)—we derive three testable hypotheses. In Section 4, we present evidence supporting these hypotheses. We also present additional analyses involving predictors' experience and confidence that lend further support to projection as the mechanism underlying our results. In Section 5, we present findings on intrapersonal projection bias and compare those results to the existing literature. Section 6 concludes.

2 Related Literature

In this section, we review prior work in both psychology and economics on interpersonal projection bias. We then describe how our experiment mitigates some of the confounds in earlier studies.

Social projection—the tendency for people to believe that others share their tastes or beliefs—has a long history in psychology (see, e.g., Ichheiser, 1949; Cronbach, 1955; Sherif and Hovland, 1961). For instance, Katz and Allport (1931) found that students who cheated on exams typically overestimated the fraction of their peers who also cheated. In a seminal paper examining this "false-consensus effect", Ross, Greene, and House (1977) introduced a methodology that was subsequently adopted by many similar experiments. Loosely, these studies elicited subjects' responses to binary-choice questions (e.g., "Would you vote for a bill to increase space-program funding?") and asked subjects to predict how the general population would answer the same questions. The false-consensus effect is then said to be observed when the average estimate of the fraction that supported a given choice was larger among those who supported that choice than those who did not (e.g., those who voted for space-program funding predict that the bill would receive more support than those who voted against it).[3]

As we will discuss in greater detail below, there are several mechanisms that can generate a false-

[3]Marks and Miller (1987) document the false-consensus effect in 45 different studies published in the decade following Ross, Greene, and House (1977). Mullen et al. (1985) find robust evidence of the effect in a meta-study of 115 tests. A recent large-scale replication study (Klein et al., 2018) shows that two of the primary results from Ross, Greene, and House (1977) replicate with similar effect sizes. The large collection of studies demonstrating a false-consensus effect span a variety of domains, including political preferences (Brown, 1982; Rouhana et al., 1997), opinions on income redistribution (Cruces et al., 2013), and risk attitudes (Faro and Rottenstreich, 2006).

consensus effect. One is that people project their preferences or states onto others. Economists have documented several instances of *intra*personal projection bias, where people project their current preferences onto their future selves and thus exaggerate the similarity between their current and future tastes (Augenblick and Rabin, 2019; Chang, Huang, and Wang, 2018; Busse et al., 2015; Conlin, O'Donoghue, and Vogelsang, 2007; additional evidence discussed in Loewenstein, O'Donoghue, and Rabin, 2003). To the extent that others' tastes are as difficult to predict as one's own, the logic underlying intrapersonal projection bias—that we "mentally trade places" with our future selves and, in doing so, project our current preferences—should similarly apply when empathizing with others.[4]

Indeed, a few studies suggest that the same transient preference states that distort people's prediction of their own future preferences similarly distort their predictions of *others'* preferences. For instance, Loewenstein and Adler (1995) demonstrate that participants who were not endowed with a good failed to predict that endowment would increase the reservation selling prices of others; Van Boven, Loewenstein, and Dunning (2003) further show that endowed sellers overestimated the willingness to pay of potential buyers. Both experiments point toward an egocentric bias in understanding others' tastes.[5,6] In a similar vein, Van Boven and Loewenstein (2003) had participants read a short vignette about three lost hikers stranded overnight in the woods. Participants were then asked to imagine themselves in the place of one of the hikers and answer the following: "Which would be more unpleasant [to you] for the hikers, hunger or thirst?" Some of participants completed at least 20 minutes of vigorous exercise before reading the vignette and answering the question; those who did were significantly more likely to be concerned about thirst when compared to those who did not exercise before answering.

Our experiment drew inspiration from Van Boven and Loewenstein's (2003) design in that we also exgoenously manipulated a predictor's state. Superficially, we focus on projecting fatigue

[4]The problem of understanding others' motivations and mental states (often called "Theory of Mind" or "folk psychology") has long been the object of study in psychology. Gordon (1986) and Goldman (1993) propose similar accounts wherein a person's mental model of others is built from simulation and self introspection. This contrasts with Theory-Theory models (e.g., Morton, 1980) which propose that people try to explain others' behavior by constructing theories of that behavior that are consistent with observed evidence. Although our experiment was not designed to directly test these divergent viewpoints, the evidence we present in Section 4 is suggestive of the simulation theory. See Van Boven et al. (2013) for a deeper discussion of this topic.

[5]Relatedly, Buchanan (2020) explores projection of one's reference point onto others. Participants in her experiment who were endowed with a lower (higher) reference point failed to predict the behavior of those with a higher (lower) reference point. Specifically, the author finds that participants neglected the effect of others' endowments when trying to predict their behavior and instead acted as if others shared their endowment.

[6]Subsequent experiments urge caution when interpreting these previous findings: an apparent egocentric bias in predictions might result from generalized errors in guessing about others. For example, Frederick (2012) shows that people generally overestimate others' willingness to pay for goods. Relatedly, Kurt and Inman (2013) demonstrate that both endowed and unendowed participants are inaccurate in their predictions about others in their *same* state. We will control for generalized errors in predictions (i.e., errors that are not state-dependent) by comparing predictions across states, rather than comparing predictions to the truth.

rather than thirst. But there are several more important design differences that help highlight our contribution to this literature. First, we precisely controlled the tiredness states of the target groups (i.e., fresh and tired workers) and the predictors. Second, predictors attempted to forecast a continuous variable (i.e., willingness to work) chosen by the target groups. Together, these two features allow us to measure projection bias on the intensive margin. Third, since we could actually observe the choices of the target group, we were able to incentivize predictions based on accuracy and analyze how accuracy varied with the predictor's state. Fourth, we elicited multiple predictions from each predictor, which reveals how readily an individual's predictions changed as her own state changed. Fifth, we measured both intra- and interpersonal projection within a shared framework.

Finally—and perhaps most importantly—predictors experienced first-hand the precise situations that the target groups experienced. Predictors worked on the same task as the workers and thus had information about their own preferences in each of the states about which they made forecasts. This aspect of our design addresses a well-known confound in the literature on social projection. As highlighted by Dawes (1989, 1990), the supposed false-consensus effect could stem from rational uncertainty regarding others' tastes and how tastes change across states. With such uncertainty, one's own preference in a given state may provide information about others' preferences in alternative states. Thus, a person may guess that others will behave like herself not because she is projecting, but because she is unfamiliar with other states. As noted above, our predictors completed the same series of tasks as the workers. Furthermore, we told them ahead of time that they would make predictions about the willingness to work of others who had completed 5 and 20 tasks. Thus, predictors (ostensibly) knew that it was in their best interest to pay attention to and recall their sentiment toward work at these two points. While there were likely still idiosyncratic differences in tastes across participants, each predictor credibly knew what it was like "to be in a worker's shoes" in both the fresh and tired state.

A few papers in the economics literature engage with this informational confound. Engelmann and Strobel (2000) examine whether the false-consensus effect survives with incentivized predictions and an informational treatment that provides a signal about others' choices. The authors find that participants tended to underweight their own choice relative to information about others when predicting the average choice of others, thereby eliminating the false-consensus effect. However, the authors' subsequent work (Engelmann and Strobel, 2012) suggests that the artificial nature of that environment—namely, the free acquisition of strong signals about others—led to the null finding in their previous work. Indeed, both Engelmann and Strobel (2012) and Ambuehl, Bernheim, and Ockenfels (2019) find that a significant false-consensus bias remains if subjects face a small cost to acquire information about others' choices.[7]

[7]There are several papers in the economics literature that find indirect evidence of interpersonal projection despite focusing on other questions. One line of papers find that people tend to project their own social preferences, such as

In addition to the rational limited-information explanation noted above, the literature on the false-consensus effect notes a few additional mechanisms that may cause people to seemingly overestimate the similarity between themselves and others (see, e.g., Marks and Miller, 1987). First, a form of availability bias or selection neglect may cause people to excessively extrapolate from the characteristics of their own social circle—which are likely correlated—when estimating the characteristics of a more general population. Second, in domains with a salient social norm, people may derive value from believing that their preferences conform with others'. Hence, due to motivated reasoning, their predictions may reflect this willfully distorted belief. We designed our experiment to mitigate these explanations. In particular, our experiment explores preferences over an unfamiliar yet mundane task. Given its unfamiliarity, it is unlikely that participants had any relevant data from which they could extrapolate or any perception of a social norm. Furthermore, by incentivizing predictions we diminished any benefit from maintaining motivated beliefs.

3 Experimental Design

We recruited a total of 1566 participants using Amazon's Mechanical Turk (MTurk).[8] Our experiment had two distinct stages, which correspond to two different participant roles. Regardless of their decisions or their role, all participants who completed the survey earned at least $3. Participants in the first stage—whom we call "workers"—completed some initial work on a real-effort task and then stated their willingness to perform more work for additional pay. Participants in the second stage—whom we call "predictors"—completed some initial work and then guessed the workers' average willingness to work.

Before providing details on these roles, we first describe the real-effort task. All participants worked on (and, when required, formed predictions about) the same real-effort task. Each round of the task required a participant to count the number of times a particular number or symbol (e.g., 0, 1, ?, !) appeared in a 10×15 matrix of numbers and symbols. See the Figure 2 for a screenshot of the task. On average, it took participants about 75 seconds to complete one round of the task.

We now provide details on the worker and predictor stages of the experiment. Complete experimental instructions are in Appendix B.

guilt aversion and trust (e.g., Ellingsen et al., 2010; Blanco et al., 2014; Butler, Giuliano, and Guiso, 2016).

[8]Participants were selected to meet the following criteria: (i) over 18 years old; (ii) resident of the United States (verified with IP address); and (iii) completion of at least 100 prior HITs on MTurk with a 95% acceptance rate. All data was collected in June 2019, prior to the COVID-19 pandemic.

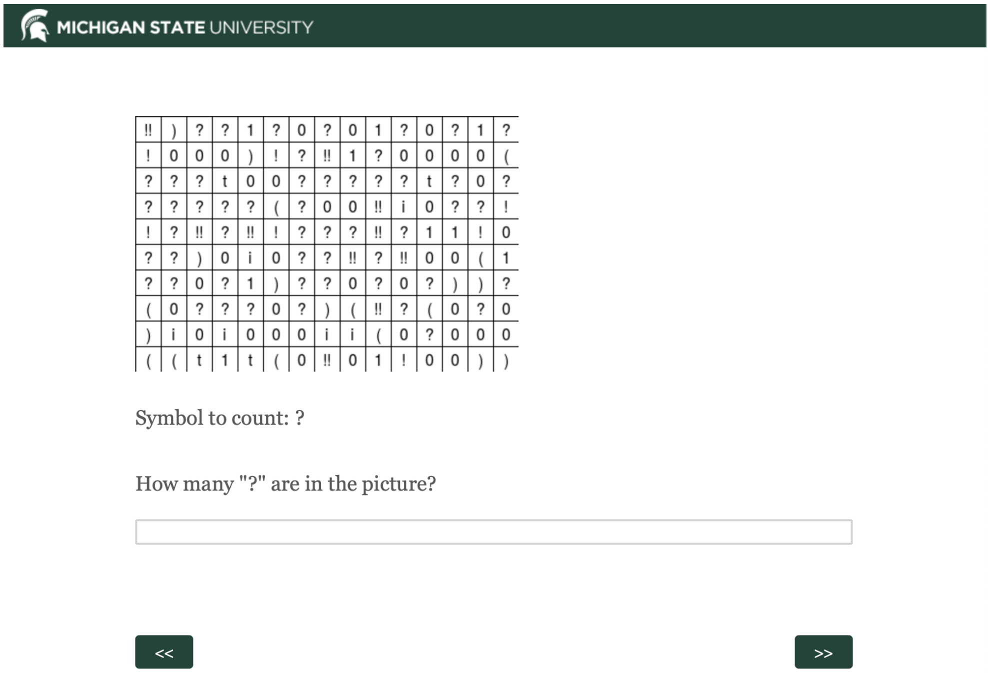

Figure 2: *Screenshot of the counting task.*

3.1 Workers

Participants in the first stage of the experiment first completed a set number of rounds of the task, and then we elicited their willingness to complete additional rounds. Workers were randomized into one of two groups: (i) in the *Fresh* group, workers completed 5 mandatory tasks prior to stating their willingness to complete more; (ii) in the *Tired* group, workers completed 20 mandatory tasks prior to stating their willingness to complete more.

We elicited willingness to work (WTW) using the Becker-Degroot-Marshak (BDM) mechanism. We asked each worker how many additional tasks they were willing to complete for a bonus of $\$m$, where m varied depending on the group (*Fresh* vs. *Tired*).[9] Participants used a slider to select a WTW between 0 and 100. We then drew a random integer z (uniformly) between 0 and 100. If z was below the participant's selected WTW, they had to complete z additional tasks in exchange for a bonus of $\$m$. Otherwise, they did no additional tasks and received no bonus.

To summarize, participants in the worker stage were randomly assigned to one of two groups:

[9] As we discuss below, these measures of WTW were the objects that predictors had to guess. We varied the monetary incentives across groups to ensure that a predictor faced distinct questions when asked about the two groups. Had we asked predictors the same question repeatedly, we may have introduced a consistency bias or a demand effect. In that counterfactual, an inattentive predictor may fail to reassess the question when asked a second time and quickly reply with the same number as their first guess. Alternatively, an attentive (and, perhaps suspicious) predictor may see the same question again and assume there is reason to change their initial guess, leading to a demand effect. Thus, we asked distinct questions about the two groups to promote independent assessments for each prediction.

Fresh Workers ($n = 303$). Participants in this group completed 5 mandatory rounds of the task before we elicited their willingness to complete more rounds. These workers were in a (relatively) fresh state when announcing their WTW. Each Participant i in this group stated how many additional rounds they were willing to complete for $m = \$2$, which we denote by $W_i(\$2,F)$ (where F denotes the fresh state). Let $\overline{W}(\$2,F)$ be the average response among this group.

Tired Workers ($n = 299$). Participants in this group completed 20 mandatory rounds of the task before we elicited their willingness to complete more rounds. These workers were in a (relatively) tired state when announcing their WTW. Each Participant i in this group stated how many additional rounds they were willing to complete for $m = \$3$, which we denote by $W_i(\$3,T)$ (where T denotes the tired state). Let $\overline{W}(\$3,T)$ be the average response among this group.

We also recruited a third group of workers that allow us to measure *intra*personal projection bias. These workers had the same experience as the tired workers described above, except they additionally predicted their own WTW ahead of time. We postpone a detailed description of this group until Section 5, where we compare inter- and intrapersonal projection bias.

3.2 Predictors

Predictors made a series of incentivized guesses about the average WTW of fresh and tired workers; that is, they predicted $\overline{W}(\$2,F)$ and $\overline{W}(\$3,T)$. In order to mitigate confounds from informational asymmetries, predictors also worked on the same task that the workers faced, and they made predictions after completing 5 tasks (i.e., in the fresh state) and after completing 20 tasks (i.e., in the tired state).

Predictors were randomly assigned to one of three groups. In each, participants had to complete 20 rounds of the counting task. The three groups differed based on when participants provided predictions (after completing 5 tasks, 20 tasks, or both) and based on which groups of workers they guessed about (fresh vs. tired).

In particular, our groups differed based on whether or not initial predictions were about workers in the same state as the predictor. Predictors in the "In Group" began by making predictions about others in their own state.

In-Group (Group "I" henceforth; $n = 223$). A predictor in this group made 3 predictions in total. First, after completing 5 tasks himself, he predicted $\overline{W}(\$2,F)$—the WTW of fresh workers. Second, after completing 20 tasks himself, he predicted $\overline{W}(\$3,T)$—the WTW of tired workers. Immediately after the second prediction, he again predicted the behavior of fresh workers, $\overline{W}(\$2,F)$. This final prediction allows us to test whether a predictor changes his view on others simply when he himself changes states.

Note that for the first two predictions above, the predictor guessed the WTW of workers who were in the same state as himself: when the predictor was fresh, he guessed the WTW of fresh workers; when the predictor was tired, he guessed the WTW of tired workers. We call these "in-group" predictions.

Predictors in our other two groups—"Out Groups"—made their first predictions about others in a *different* state than themselves. Our two Out Groups differed in which prediction they made first: one made predictions when both fresh and tired, while the other made predictions only when tired.

Out-Group A (Group "A" henceforth; $n = 221$). A predictor in this group made 3 predictions in total: (1) after completing 5 tasks herself, she predicted $\overline{W}(\$3,T)$—the WTW of tired workers; (2) after completing 20 tasks, she predicted $\overline{W}(\$2,F)$—the WTW of fresh workers; and (3) immediately after the second prediction, she again predicted the behavior of tired workers, $\overline{W}(\$3,T)$.

In contrast to the In-Group, the first two predictions above involved guessing the WTW of workers who were in a *different* state than the predictor: when the predictor was fresh, she guessed the WTW of tired workers; when she was tired, she guessed the WTW of fresh workers. We call these "out-group" predictions.

Out-Group B (Group "B" henceforth; $n = 222$). A predictor in this group made predictions only after completing all 20 tasks. Aside from not eliciting predictions after a predictor completed 5 tasks, this group was otherwise identical to Out-Group A: after a predictor in this group completed 20 tasks herself, she first predicted the WTW of fresh workers, $\overline{W}(\$2,F)$, and then predicted the WTW of tired workers, $\overline{W}(\$3,T)$.

To ensure that the instructions and timing for this group closely mirrored those of Group A, we interrupted each predictor in Group B after she completed 5 tasks. During this interruption, we presented instructions on the BDM mechanism and reminded the participant that she would later make predictions about others who made choices in the state she was currently in (i.e., fresh). Thus, even though a predictor in Group B did not make a numerical guess after she completed 5 tasks, she was still paused and cued to think about others while in the fresh state.

We introduce the following notation for the predictions described above. Let $\widehat{W}_i^g(m,s|s_i)$ denote the guess of Predictor i from Group $g \in \{I,A,B\}$ about workers in state $s \in \{F,T\}$ who face incentives m, where s_i denotes Predictor i's own state at the time of her prediction. For example, $\widehat{W}_i^g(\$2, F|T)$ is Predictor i's guess about $\overline{W}(\$2, F)$ cast while she is in the tired state. Let $\widehat{W}^g(m, s|s')$ denote the average prediction of $\overline{W}(m, s)$ among predictors in group g who were in state $s' \in \{F, T\}$ when making their predictions.

All predictions from each group were incentivized as follows: a participant earned a 50-cent bonus for each prediction that was within 5 tasks of the true value.[10] After each prediction, we also asked participants to rank their confidence in that prediction on a scale from 1 (not at all confident) to 5 (extremely confident). These confidence measures were not incentivized. Finally, after completing all 20 tasks and providing all predictions, we asked predictors about their own willingness to complete more tasks for an additional payment of \$3. This question was phrased identically to the one we asked workers, but it was not incentivized.[11]

Predictors received no feedback during the prediction portion of the experiment. Their payments—which were based on the accuracy of their guesses—were revealed only once the experiment was over.

3.3 Theoretical Predictions

This section derives some immediate implications of interpersonal projection bias, which comprise the hypotheses we test below. We are primarily interested in the extent to which a predictor's own state (fresh vs. tired) influences her predictions about workers' WTW in a different state. Interpersonal projection bias implies that fresh predictors will overestimate the willingness to work of tired workers, and, contrastingly, tired predictors will underestimate the willingness to work of rested workers.

To formalize these hypotheses, first consider the behavior of workers. Suppose Participant i's cost of completing $e \in \mathbb{R}_+$ additional tasks is given by an increasing and convex state-dependent cost function $c(e; s_i, \theta_i)$, where the "state" $s_i \in \mathbb{R}_+$ is the number of tasks i completed beforehand and $\theta_i \in \mathbb{R}$ is i's idiosyncratic attitude toward the task. Additionally, for a fixed effort level e, we assume the cost $c(e; s_i, \theta_i)$ is increasing s_i; that is, effort becomes more costly as the worker grows tired. When asked how many tasks she is willing to complete for a bonus payment of m,

[10]Note that this mechanism is not incentive compatible for eliciting point estimates for beliefs. In particular, any prediction $\widehat{W}_i(m, s|s_i) < 5$ (or $\widehat{W}_i(m, s|s_i) > 95$) is strictly dominated by simply guessing 5 (or 95). However, out of the 1,776 total predictions we collect, only 4% fall outside the interval $[5, 95]$. Dropping these responses does not substantively change any of our results.

[11]Adding incentives to this last question would substantially increase the length of the experiment for predictors, as incentivized work requires that people actually complete that work. Given the budget of the experiment and the (already long) duration, we opted to collect hypothetical WTW instead. It is worth noting that the average WTW from this unincentivized elicitation is similar to the average WTW of our tired (incentivized) workers ($p = 0.443$ for difference).

Participant i chooses e to maximize $\int_0^e [m - c(\tilde{e}; s, \theta_i)] d\tilde{e}$.[12] Participant i's optimal choice is thus implicitly defined by the solution to $c(e; s_i, \theta_i) = m$. As introduced above, this choice is denoted by $W(m, s_i|\theta_i)$. Given our assumptions on c, $W(m, s_i|\theta_i)$ is decreasing in s_i for a fixed monetary bonus—intuitively, a person is less willing to work as she grows tired.

We now consider predictors and present a simple model of interpersonal projection bias. Building from Loewenstein, O'Donoghue, and Rabin's (2003) model of *intra*personal projection, suppose that Participant i wrongly believes that Participant j's cost function is

$$\hat{c}(e; s_j, \theta_j | s_i, \theta_i) = \alpha c(e; s_i, \theta_i) + (1 - \alpha) c(e; s_j, \theta_j), \tag{1}$$

where parameter $\alpha \in [0, 1]$ captures the degree of projection bias. That is, a projecting individual perceives another person's cost as a convex combination of her own cost and that other person's true cost. When $\alpha = 0$, this model collapses to the rational unbiased model.

Under projection bias, Participant i predicts that Participant j will choose an effort level that solves $\hat{c}(e; s_j, \theta_j | s_i, \theta_i) = m$. Hence, i's prediction about the WTW of an average individual in state s_j, denoted by $\widehat{W}(m, s_j | s_i, \theta_i)$, is decreasing in both s_j and s_i whenever $\alpha > 0$.[13] On the other hand, $\widehat{W}(m, s_j | s_i, \theta_i)$ is constant in s_i absent projection bias ($\alpha = 0$).

Although the model allows for a continuum of tiredness states, our experiment focuses on just two: $s = 5$ and $s = 20$. To make the state salient in our notation, we will henceforth denote these two numerical states by F and T, respectively.

We can now state our primary hypotheses regarding the average estimates cast by predictors. Under projection bias ($\alpha > 0$), we have the following:

> **Hypothesis 1:** *Fresh predictors accurately estimate the WTW of Fresh Workers, and tired predictors accurately estimate the WTW of tired workers:* $\widehat{W}^I(\$2, F|F) = \overline{W}(\$2, F)$ and $\widehat{W}^g(\$3, T|T) = \overline{W}(\$3, T)$ for $g \in \{I, A, B\}$.
>
> **Hypothesis 2:** *Relative to tired predictors, fresh predictors overestimate the WTW of tired workers:* $\widehat{W}^A(\$3, T|F) > \widehat{W}^g(\$3, T|T)$ for each $g \in \{I, A, B\}$.
>
> **Hypothesis 3:** *Relative to fresh predictors, tired predictors underestimate the WTW of fresh workers:* $\widehat{W}^g(\$2, F|T) < \widehat{W}^I(\$2, F|F)$ for each $g \in \{I, A, B\}$.

A variety of biases in belief formation *independent* of projection bias could jeopardize Hypothesis 1 above (see Benjamin, 2019 for a comprehensive review). Hypotheses 2 and 3, however, are robust

[12] This follows from the BDM mechanism: if the mechanism draws a random integer $z \in [0, e]$, the worker must complete z tasks in exchange for m dollars; if $z > e$, then the worker does no extra work and receives no bonus.

[13] Note that Participant i's prediction about j, $\widehat{W}(m, s_j | s_i, \theta_i)$, also depends on i's beliefs about θ_j. To structure the model, we assume a predictor is Bayesian aside from the misspecified model of costs presented in (1). Thus, $\widehat{W}(m, s_j | s_i, \theta_i)$ is the expected value of effort that maximizes (1) given Predictor i's beliefs over θ_j.

to such biases because they compare predictions across groups rather than compare the predictions of one group to the truth. If Hypothesis 1 does in fact hold, then we have two immediate corollaries of Hypotheses 2 and 3:

Hypothesis 2A: *Fresh predictors overestimate the WTW of tired workers:* $\widehat{W}^A(\$3,T|F) > \overline{W}(\$3,T)$.

Hypothesis 3A: *Tired predictors underestimate the WTW of fresh workers:* $\widehat{W}^g(\$2,F|T) < \overline{W}(\$2,F)$ for each $g \in \{I,A,B\}$.

3.4 Discussion

Some of our predictions above are potentially confounded by alternative explanations. Here, we discuss how our design addresses these confounds.

First, a fresh predictor may have overestimated the WTW of a tired worker simply because he was uncertain about how onerous the task would become after working longer. Indeed, fresh predictors who guessed about $\overline{W}(\$3,T)$ had never completed 20 tasks themselves, and thus were unfamiliar with the state in which tired workers made decisions. Therefore, we are cautious to interpret an outcome in which $\widehat{W}^A(\$3,T|F) > \overline{W}(\$3,T)$ as stemming purely from projection bias. Such a finding may have stemmed, in part, from uncertainty about the cost function and prior beliefs that the task is less onerous than it really is. Crucially, this limited-information confound is not present—or is at least less relevant—for *tired* predictors who guessed about the behavior of fresh workers. Tired predictors had already experienced the fresh state, and hence they should have understood (to some extent) others' desire to work when fresh. Thus, if limited information caused fresh predictors to overestimate the WTW of tired workers, then we would expect the observed bias in predictions about tired workers while fresh, $|\widehat{W}^A(\$3,T|F) - \overline{W}(\$3,T)|$, to be larger than the bias in predictions about fresh workers while tired, $|\widehat{W}^g(\$2,F|T) - \overline{W}(\$2,F)|$ for each $g \in \{I,A,B\}$. Accordingly, the latter difference provides a cleaner measure of projection bias, and the difference between the two reflects the extent to which limited information distorted predictions.

A variant of the informational confound just discussed could still emerge, however, if tired predictors did not remember what it was like to be in the fresh state. In that case, a tired predictor might have rationally used his current state to approximate the fresh state because he was uncertain (due to limited memory) what that state was like. To address this, our design attempted to make predictors' experience in the fresh state salient and memorable. First, predictors in Groups I and A were interrupted after 5 tasks (i.e., while they were in the fresh state) in order to make predictions. At that point, they likely introspected to some extent. Furthermore, the instructions explicitly told participants that they would later make predictions about the WTW of fresh workers, and hence it behooved them to remember their attitude toward additional work while in that state.

While requiring predictors in Groups *I* and *A* to pause and make predictions when fresh likely provided them with useful information about that state, predictors in Group *B* did not make such predictions. Hence, they may have been less familiar with workers' sentiment in the fresh state (relative to Groups *I* and *A*). To mitigate this concern, we briefly interrupted participants in Group *B* after they completed 5 tasks (i.e., while they were in the fresh state) to deliver some of the instructions. In particular, we reminded them that they would later need to predict the WTW of workers in the fresh state, thereby emphasizing the value of remembering their current attitude toward additional work in that state. Thus, Group *B* was designed to be as similar as possible to Groups *I* and *A* in terms of information. Insofar as participants internalized these instructions, we believe our design accounts for most information-based differences in predictions across states.

Although the design of Group *B* perhaps provided weaker incentives for participants to notice and recall their attitude toward the fresh state, it delivered an important degree of control that Groups *A* and *I* lacked. Namely, participants in groups *A* and *I* made several predictions in various states, and hence their predictions may have exhibited order effects. For instance, participants may have subconsciously anchored later predictions toward their initial guesses, or they may have deliberately chosen later predictions to appear consistent with their initial guesses. Leveraging data from Group *B*—as we do in the next section—allows for a between-subject analysis of projection bias that uses only initial guesses across groups and thus controls for any such order effects. To summarize: we designed Group *A* to control for informational concerns and Group *B* to control for order-effect concerns.

4 Results

In this section, we test our basic predictions of interpersonal projection bias. We first present the baseline WTW of fresh and tired workers and demonstrate that our manipulation of "tiredness" was successful. We then analyze predictors' average guesses about other workers' WTW and discuss evidence supporting our three hypotheses presented above. We then use within-subject variation in predictions to show that participants in Group *I* erroneously decrease their predictions about fresh workers as they become tired. Along these lines, we provide various estimates of the degree of interpersonal projection bias utilizing similar within-subject data. We conclude with a few additional analyses that provide support for projection bias over alternative explanations.

4.1 Willingness to Work Among Workers

We first present the aggregated willingness to work of fresh and tired workers. Table 1 shows the average WTW among these groups. Although raw responses are similar across the two groups

(Row 1), recall that fresh workers stated their WTW for $2 while tired workers stated their WTW for $3. Row 2 accounts for these differential monetary incentives by showing WTW in terms of tasks per dollar. Under this normalization, we see that tiredness had a marked effect: average WTW when fresh was about 10.6 tasks per dollar, when tired it was about 6.8 tasks per dollar (difference significant at $p < .001$; Welch's two-sided t-test used throughout in-body discussion of results).[14] Thus, our tiredness manipulation succeeds at generating a meaningful change in participants' attitude toward work.[15]

Table 1:
AVERAGE WILLINGNESS TO WORK

	Workers' State		
	Fresh (5 tasks)	Tired (20 tasks)	Difference
Number of Tasks	21.29	20.44	0.85
	(1.383)	(1.224)	(1.847)
Tasks Per Dollar	10.64	6.81	3.83***
	(0.692)	(0.408)	(0.803)
Observations	300	299	

Notes: Standard errors are in parentheses. Difference in tasks-per-dollar row significant at $p < .001$ (Welch's two-sided t-test).

4.2 Main Results

We now examine the extent to which predictors anticipated the difference in WTW across the fresh and tired states and the accuracy of their guesses. We evaluate each of our three enumerated hypotheses from Section 3.3 in turn. When presenting results in this subsection, we will continue to normalize WTW (and predictions thereof) in terms of task per dollar. This is purely to aid exposition by preemptively accounting for the different monetary incentives faced by the two worker groups. None of our results rely on this normalization; unnormalized predictions are presented in Appendix A.

[14] This is a conservative test given that our (ex ante) predictions were directional.

[15] Comparing the raw WTW across groups, it appears that the extra fifteen tasks completed by the Tired group almost perfectly counteracted the higher incentives they faced. This was a secondary design consideration, because having roughly equal WTW across groups may have helped avoid forms of anchoring effects among the predictors.

Hypothesis 1: Fresh predictors accurately estimate the WTW of fresh workers, and tired predictors accurately estimate the WTW of tired workers.

Table 2 shows that when predictors were themselves fresh, their average guesses matched the average WTW of fresh workers (difference not significant; $p = 0.855$). Likewise, when predictors were tired, they accurately guessed the WTW of tired workers (difference not significant; $p = 0.704$). Recall that only Group I made a prediction about fresh workers when they themselves were fresh, while all three groups guessed about tired workers when they themselves were tired. Despite these unequal samples in this analysis, our results in Table 2 rule out large or systematic errors in predictions. We therefore find support for Hypothesis 1.

Table 2:
PREDICTIONS OF WILLINGNESS TO WORK, SAME STATE (TASKS PER DOLLAR)

Predictor's State	Prediction	True WTW	Difference
Fresh (after 5 tasks)	10.81	10.64	0.17
	(0.605)	(0.692)	(0.957)
	$n = 223$	$n = 300$	
Tired (after 20 tasks)	6.65	6.81	−0.17
	(0.230)	(0.408)	(0.439)
	$n = 666$	$n = 299$	

Notes: Standard errors are in parentheses. Differences not significant ($p = 0.860$ and $p = 0.704$ from top to bottom; Welch's two-sided t-test).

Hypothesis 2: Relative to tired predictors, fresh predictors overestimate the WTW of tired workers.

Turning to our second hypothesis, we now test whether fresh predictors overestimated the WTW of tired workers. As described in Section 3.3, we control for potential biases in predictions that were independent of a predictor's state by comparing predictions cast by fresh workers to those cast by tired predictors.[16] We present this information in the "Tired" column of Table 3. Note that only Group A cast predictions about tired workers while fresh, yet all three predictor groups cast that prediction while tired. Hence, the top-right cell of Table 3 shows $\widehat{W}^A(\$3, T|F)$, and the bottom-right cell shows $\widehat{W}^g(\$3, T|T)$ averaged over all members of groups $g \in \{I, A, B\}$. We find that fresh predictors significantly overestimated the WTW of tired workers: their guesses were

[16]Our results confirming Hypothesis 1 suggest that any such biases wash out in aggregate. Thus, it is essentially equivalent to test Hypotheses 2–3 or 2A–3A.

more than 50% higher than those cast by tired predictors (difference of 3.57 tasks per dollar or 10.71 total tasks; significant at $p < .001$).

As previously discussed, uncertainty among fresh predictors about how tiredness accumulates could have reasonably contributed to this overestimation. Thus, we take this as an upper-bound on the effect of projection. We elaborate on this point below when we compare initial predictions across groups.

Table 3:
STATE-DEPENDENT PREDICTIONS (TASKS PER DOLLAR)

	Workers' State	
	Fresh	Tired
Predictors' State		
Fresh (after 5 tasks)	10.81 (0.605)	10.22 (0.491)
Tired (after 20 tasks)	9.46 (0.345)	6.65 (0.230)
Difference	1.36** (0.691)	3.57*** (0.489)

Notes: Standard errors are in parentheses. Decreased variance in tired column reflects the fact that all predictors made two guesses when tired. Sample sizes are (clockwise from top-left): 223, 221, 666, 666. Differences significant at $p = .049$; and $p < 0.001$ (left to right; Welch's two-sided t-test).

Hypothesis 3: Relative to fresh predictors, tired predictors underestimate the WTW of fresh workers.

The "Fresh" column of Table 3 confirms this result. There, we show $\widehat{W}^I(\$2, F|F)$ (top-left cell) and $\widehat{W}^g(\$2, F|T)$ averaged over all members of groups $g \in \{I, A, B\}$ (bottom-left cell). We find that tired predictors significantly underestimated WTW (difference of 1.36 tasks per dollar or 2.72 total tasks; significant at $p = .049$).

Notice that tired predictors underestimated the WTW of fresh workers by a smaller degree than fresh predictors overestimated the WTW of tired workers. This reflects our discussion in the previous section: since fresh predictors faced uncertainty about how onerous the task would become, their guesses may have been biased due to limited experience in addition to projection bias. By contrast, for tired predictors, there was minimal scope for limited information to drive the prediction error we find in Table 3. Hence, the error made by tired predictors (about 1.36 tasks per dollar) provides a cleaner measure of projection bias. However, as we discuss next, Table 3 understates the magnitude of this error due to its aggregation of data across Predictor groups.

Analysis of Initial Predictions by Group

We now shift focus to comparing initial predictions to the true WTW of workers. Although our aggregate findings support our ex-ante hypotheses, disaggregating the predictors' data down to the group level yields additional insight (see Table A1 in Appendix A for all disaggregated predictions). In particular, the average second guess among Group A, $\widehat{W}^A(\$2, F|T)$, was relatively high and thus does not exhibit the underestimation prescribed by Hypothesis 3. However, this group does indeed exhibit a substantial difference between their first and second predictions, $\widehat{W}^A(\$3, T|F)$ and $\widehat{W}^A(\$2, F|T)$, respectively, as implied by projection bias. We suspect that the elevated second predictions among Group A stemmed from order effects, e.g., anchoring or consistency bias. Namely, since their first guesses were very high—perhaps due to projection (see top-right cell of Table 3)—their subsequent guesses may have been shifted upward as well. This would partially obfuscate our ability to detect projection amongst Group A's second predictions.

Fortunately, our experimental design allows us to sidestep such order effects by analyzing only the first predictions cast by each group. Predictors in Group I made accurate first guesses, as shown in Table 2. In contrast, Table 4 shows that the first guesses among predictors in Groups A and B were systematically biased.[17] Group A's average prediction of the WTW of tired workers was far too high, while Group B's average prediction of the WTW of fresh workers was too low. Furthermore, when focusing only on first guesses, we find a more pronounced projection error among tired predictors guessing about fresh workers (2.2 tasks per dollar in Table 4 vs 1.36 in Table 3). The previous estimate (Table 3) understates this degree of projection because it includes the upward-biased second guesses of Group A.

Thus, we believe that the difference of 2.2 tasks per dollar—which represents an underestimate of approximately 21% relative to the truth—reflects our cleanest estimate of the effect of projection bias on predictions. This also allows us to loosely approximate the effect that uncertainty had on Group A's guesses. Given that their guesses were approximately 50% too high (3.57 tasks per dollar or 10.71 total tasks; see Table 4), we can decompose their error into two roughly equal-sized parts: an error due to uncertainty, and an error due to projection bias.[18] In the next section, we

[17]Note that the distribution of responses underlying Table 4 is represented in Figure 1.

[18]More precisely, "uncertainty" here and above refers to prior beliefs that were miscalibrated about how onerous the task would become. Specifically, priors that underestimated this onerousness would have caused fresh predictors to overestimate the WTW of tired workers, as in Hypothesis 2.

quantify the degree of interpersonal projection bias using alternative approaches and find similar magnitudes.

Table 4:
FIRST PREDICTIONS VS WORKERS' WTW (TASKS PER DOLLAR)

	Prediction	True WTW	Difference
Fresh Predictors $\rightarrow$ Tired Workers	10.22	6.81	3.40***
	(0.491)	(0.408)	(0.639)
	$n = 221$	$n = 299$	
Tired Predictors $\rightarrow$ Fresh Workers	8.44	10.64	−2.20***
	(0.532)	(0.692)	(0.873)
	$n = 222$	$n = 300$	

Notes: Standard errors are in parentheses. Differences significant at $p < .001$ and $p = .012$ (top to bottom; Welch's two-sided t-test).

4.3 Quantifying Interpersonal Projection Bias

We now provide some simple measures of projection bias. Specifically, we examine both of the following questions: (i) Fixing the group of workers about whom predictors are guessing, how do guesses change once predictors move from fresh to tired? (ii) Fixing the predictors' state, how do guesses about different groups of workers depend on the predictor's own stated WTW? The metrics we calculate in answering these questions also allow us to compare interpersonal projection to existing papers on *intra*personal projection—an exercise we do in Section 5.

First, we consider how predictors in Group I changed their guesses about $\overline{W}(\$2,F)$ after they completed additional tasks and thus became (relatively) tired. In our opinion, this test represents one of our strongest indicators of projection bias. Note that Group I's first guess about $\overline{W}(\$2,F)$ was made while they themselves were in the fresh state. Accordingly, they had the exact same information—and tiredness—as the workers did when stating their WTW. Thus, gaining additional exposure to the task should not lead Group I predictors to change their guesses about this particular value. Nevertheless, we find that Group I predictors significantly lowered their guesses about fresh workers once they themselves became tired. The average revision was $\widehat{W}^I(\$2,F|F) - \widehat{W}^I(\$2,F|T) = 4.2$ total tasks (s.e. = 0.858; difference significant at $p < .001$). Put another way, although predictors' initial guesses about $\overline{W}(\$2,F)$ were well calibrated (see Table

2), they wrongly lowered their guesses about this quantity after they became tired. These adjustments represent a change of approximately 19%, and imply that tired predictors underestimated the WTW of fresh workers by about 21%. As we show below, these adjustments significantly reduced the expected earnings for Group I.

Second, we consider how predictors in Group A changed their guesses about $\overline{W}(\$3,T)$ once they became tired. Recall that Group A cast this guess first while fresh. At that time, they potentially lacked information about how it felt to be tired. We may therefore expect a relatively large change in their predictions, stemming from a combination of this uncertainty with projection. Indeed, the average revision in guesses about tired workers was $\widehat{W}^A(\$3,T|F) - \widehat{W}^A(\$3,T|T) = 8.02$ total tasks (s.e. = 1.018; difference significant at $p < .001$), representing a change of approximately 26% from their first (inflated) guess to their second. Furthermore, this implies that fresh predictors overestimated the WTW of tired workers by about 39%.

The measures above provide a simple non-parametric view regarding the degree of interpersonal projection bias. We now estimate a parametric model motivated by our theoretical framework in Section 3.3. To take that model to the data, we consider a slight variant. Our specification above (and other models of projection bias) assume that the parameter measuring projection bias, α, captures a convex combination of utility functions across states (see Equation 1). However, since we instead observe effort, we estimate a parameter that captures a convex combination of the optimal effort across states.

More specifically, a projector's prediction about a worker's optimal effort is taken to be a convex combination of his own optimal effort in his current state and his unbiased estimate about a worker's effort in the target state. Let $W(m,s|\theta)$ be the utility-maximizing WTW of a participant facing payment m in state s, where θ represents her idiosyncratic taste for the task. Predictor i's guess about the average action of a worker in state s facing pay m is then

$$\widehat{W}(m,s|s_i,\theta_i) = \rho W(m,s_i|\theta_i) + (1-\rho)\mathbb{E}_\theta[W(m,s|\theta)|\theta_i], \tag{2}$$

where $\mathbb{E}_\theta[\cdot|\theta_i]$ denotes Predictor i's subjective expectation over θ conditional on himself having type θ_i, and s_i is Predictor i's state at the time of casting this prediction. That is, Predictor i distorts his prediction toward his own current WTW for \$$m$, and parameter ρ measures the extent of this distortion.

We take this model to our data as follows. First, note that all predictors made two guesses when they were in the tired state. Using Equation (2), we can write these two predictions as

$$\widehat{W}(\$2,F|T,\theta_i) = \rho W(\$2,T|\theta_i) + (1-\rho)\mathbb{E}_\theta[W(\$2,F|\theta)|\theta_i], \tag{3}$$

and

$$\widehat{W}(\$3, T|T, \theta_i) = \rho W(\$3, T|\theta_i) + (1-\rho)\mathbb{E}_\theta[W(\$3, T|\theta)|\theta_i]. \tag{4}$$

Recall that we elicited predictors' own (hypothetical) WTW for \$3 when tired but not for \$2 when tired; thus we measure $W(\$3, T|\theta_i)$ but not $W(\$2, T|\theta_i)$. In order to estimate ρ with this limited data, we leverage our assumption that the effort-cost function is convex in effort. It therefore follows that

$$\frac{2}{3}W(\$3, T|\theta_i) \leq W(\$2, T|\theta_i). \tag{5}$$

This inequality can be used along with Equations (3) and (4) to estimate a lower bound on ρ.

In particular, differencing Equations (3) and (4) isolates the difference $\widehat{W}(\$3, T|T, \theta_i) - \widehat{W}(\$2, F|T, \theta_i)$ as our LHS variable used to estimate ρ.[19] Substituting the inequality from (5) yields the following econometric model:

$$\widehat{W}_i(\$3, T|T) - \widehat{W}_i(\$2, F|T) = \beta_0 + \beta_1\left(\frac{1}{3}W_i(\$3, T)\right) + \varepsilon_i. \tag{6}$$

Thus, our estimate $\widehat{\beta_1}$ provides a lower bound for ρ. Pooling all predictors and estimating via OLS, this analysis yields a lower bound $\rho \geq 0.23$ (s.e. = 0.061).[20]

To summarize, our various measures of interpersonal projection bias stem from two distinct approaches. The non-parametric approach generates estimates based on variation in the predictor's state: as the predictor went from fresh to tired, we observe how their guess about some fixed quantity changed. The second approach fixes the predictor's state, and examines how their prediction is influenced by that state and their own WTW. Although neither approach is beyond reproach, our various estimates suggest that projection distorts predictions by somewhere between 21% and 39%. In Section 5, we discuss how these estimates of interpersonal projection compare to intrapersonal projection from both our own study and previous papers.

4.4 Additional Analyses

In this section, we present a few additional analyses that provide further support for projection as the mechanism underlying our findings. We show that learning from experience with the task was not a primary driver of our effects by examining (i) how the accuracy of a predictor's guesses changed as they accumulated more experience with the task, and (ii) self-reported confidence ratings across guesses. We then explore whether the amount of time it took participants to complete

[19]Note that this approach assumes that a predictor's perceived difference in the expected WTW across states, $\mathbb{E}_\theta[W(\$3, T|\theta)|\theta_i] - \mathbb{E}_\theta[W(\$2, F|\theta)|\theta_i]$, is independent of their own WTW. This holds, for instance, if we assume that the predictor's own WTW influences each expectation term in a similar, additively-separable fashion.

[20]Allowing the intercept to vary for each of the three groups ($I, A,$ and B) does not change this result, and we find $\rho \geq 0.22$ (s.e. = 0.063).

the tasks affected their predictions; we find no such effect.

We first evaluate how participants' guesses improved (or failed to improve) with more task experience. Table 5 shows the mean absolute error in each guess for each group. We see that predictors' guesses tended to become slightly more accurate with time. This improvement, though, is state-dependent: when predictors were guessing about workers who shared their state, they tended to be more accurate than when guessing about workers in the opposite state. Pooling all of the guesses that were cast by tired predictors (i.e. after accumulating experience), we find that same-state guesses were significantly more accurate than different-state guesses (difference 0.730, $p = 0.032$).

Table 5:
PREDICTION ACCURACY (MEAN ABSOLUTE ERROR) BY GROUP

	Group *I*	Group *A*	Group *B*
Mean Abs Error, 1st Prediction	11.29	15.32	-
	(0.942)	(1.256)	
Mean Abs Error, 2nd Prediction	10.33	13.38	11.01
	(0.787)	(1.087)	(0.820)
Mean Abs Error, 3rd Prediction	11.40	12.85	10.42
	(0.763)	(1.081)	(0.895)

Notes: Standard errors are in parentheses.

Furthermore, we find that any improvement in Group *I*'s accuracy (from first to second prediction) vanished by their third prediction. Recall that Group *I*'s first guess was about fresh workers and was cast when they themselves were fresh; Group *I*'s third guess was again about fresh workers, but was cast when they were tired. Once Group *I* became tired, they (mistakenly) lowered their previous estimates about fresh workers, which reduced accuracy. Although this reduction in accuracy does not appear significant in Table 5, an analysis of expected earnings reveals that it came at a considerable cost: Group *I*'s expected earnings significantly decreased between their first and third predictions. Specifically, the number of guesses within ± 5 tasks of the true WTW—and thus guesses that could have increased earnings—fell by approximately 26% (difference significant at $p = .002$).

Overall, we believe that these limited improvements in accuracy—along with our results on confidence, below—suggest that there may have been *some* learning, but that this learning does not fully account for many of the effects that we observe.

We now evaluate predictors' confidence ratings, providing further evidence that learning about the disutility of work does not drive our results. Recall that after each prediction, participants reported their confidence on a five-point scale, where 1 represents "Not at all confident" and 5 represents "Extremely confident". Average responses are reported in Table A2 in Appendix A. We first consider confidence of Groups I and A, as these groups made predictions while both fresh and tired. As shown in Table A2, average confidence did not increase with experience—either in going from the first prediction to the second (and thus accumulating more experience with the task) or in going from the second prediction to the third (and thus accumulating more experience with predicting). In sum, predictors did not grow more confident as they accrued experience.

In an ex-post analysis, we discovered that predictors who were extremely confident tended to be *less* accurate (à la Kruger and Dunning, 1999). While this is consistent with the classic Dunning-Kruger effect, this correlation is also predicted by projection bias: as the extent of projection increases, a predictor believes that she has a more precise assessment of others because she is more confident that others will act like herself. At the same time, an increase in projection leads to a greater bias in predictions; hence, it induces a negative correlation between confidence and accuracy. To test whether this correlation indeed stems from projection, we considered predictors from the In-Group and Out-Group A, and we split them into two groups: (i) "high confidence" predictors who responded with "Extremely confident" to at least one of the confidence questions, and (ii) predictors who never responded with "Extremely confident". We then calculated a crude measure of projection for each predictor: how much, in percentage terms, they revised their first guess after they became tired.[21] Those with extremely high confidence changed their guesses by 26.7% on average, while those with non-extreme confidence changed their guesses by 14.6% on average (difference significant at $p = 0.032$). We believe this provides additional suggestive evidence for projection, insofar as strongly-biased projectors exhibited extreme confidence because they believed—either directly or inattentively—that their own attitude toward work was very informative about others'.

Finally, we briefly consider task completion time and its (null) effect on the degree of projection. Ex ante, we believed that those who took longer to complete the tasks might be more "tired." Thus, we believed that relatively slow predictors might exhibit a greater degree of projection when asked about fresh workers. Our data does not bear this out. We present a series of exploratory analyses in Appendix A.3 which demonstrate that task completion time has no effect on participant predictions, their self-reported confidence, or the imputed degree of projection. We suspect that heterogeneity in task-completion times likely stemmed from inattention (e.g., doing other things online), but we have no direct evidence to back this assertion.[22]

[21]This is the percentage change between a predictor's first and third guesses. Note that this is the same non-parametric measure of projection bias considered in Section 4.3.

[22]Ex post, this finding helps justify our choice to alter tiredness in a binary way. However, we believe that varying tiredness in a more continuous way would be a viable approach in a laboratory setting.

5 Intrapersonal Projection Bias: Predictions about Own WTW

We now compare inter- and *intra*personal projection bias—the propensity for one's current state to overly influence predictions about their *own* behavior in a different state. To provide evidence for the latter in our real-effort domain, we ran an additional worker group (called "Predicting Workers"). This group was identical to our Tired-Workers group, except that participants predicted their own WTW ahead of time. This allows us to measure the extent to which fresh workers mispredicted their own WTW while tired. Specifically, after a predicting worker completed 5 mandatory rounds (out of 20), we asked them to predict how many additional rounds they would complete for a bonus of \$3 once they had finished the mandatory 20 rounds. Thus, while in the fresh state, these participants predicted their own attitude toward work in the tired state.[23] Then, after completing the mandatory 20 tasks, we asked participants how many additional tasks they would complete for a bonus of \$3. We elicited this WTW using a BDM mechanism exactly as in the Tired-Workers group.

This additional group allows us to measure the extent to which participants mispredicted their own behavior. Table 6 shows the predictions and actual WTW among predicting workers. As in the previous section, we take the difference between beliefs and real WTW as a raw metric of projection bias: on average, fresh workers overestimated their own WTW when tired by roughly 5 tasks—approximately 30% of their true WTW.

Table 6:
PREDICTING WORKERS' GUESSES AND WTW

	Prediction	Actual	Difference
WTW (# of Tasks)	22.11	17.02	5.09***
	(1.179)	(1.161)	(1.044)
Observations	298	298	298

Notes: Standard errors are in parentheses. Difference significant at $p < .001$ (Welch's two-sided t-test).

Interpreting this number, however, requires some caution. First, these mispredictions about future WTW came from workers in the fresh state who had not yet experienced the tired state. Hence, these mispredictions may have stemmed from predictors underestimating how onerous the task would become. Since this force acts in the same direction as projection, the 30% error noted above may overstate the degree of intrapersonal projection. In contrast, by monetarily incentivizing participants' predictions, we may have indirectly incentivized consistency. Namely, stating a WTW

[23]These predictions were incentivized in the same way as all other predictions in this experiment.

close to one's prediction would have increased a person's payout (relative to stating a different WTW). Since consistency acts against projection bias, the 30% error may also understate the degree of intrapersonal projection.[24]

To assess the relative magnitudes of intra- and interpersonal projection bias, we compare prediction errors among the predicting workers with those among the fresh predictors who guessed the WTW of tired workers. Recall that fresh predictors overestimated the WTW of others in the tired state by roughly 10.7 tasks (see Table 3)—approximately 50% of tired workers' true WTW—while fresh workers overestimated their *own* WTW by approximately 30%. Thus, despite significant biases among both groups, participants were much better calibrated when making predictions about themselves rather than others: our measure of the intrapersonal prediction error is substantially smaller than the interpersonal one.

The comparison above comes with a caveat worth noting. Specifically, we can directly compare our measures of intra- and interpersonal projection if we assume that the uncertainty about how onerous the task would become was similar when considering oneself and considering others. Importantly, note that Table 4 (and the surrounding discussion) suggests that the interpersonal error we observe among fresh predictors stems from both uncertainty and projection, and that their relative contributions are roughly equal. Thus, even in the limit case in which workers held well-calibrated beliefs about their own future effort cost, we would conclude that intrapersonal projection has roughly the same magnitude as interpersonal projection. Short of this (implausible) limit, we would conclude that interpersonal projection is larger.[25]

Our measure of intrapersonal projection bias falls in the range of existing estimates in the literature. These measures come from a variety of different domains and different estimation schemes; accordingly, there is no a priori reason that our results should be the same as others. Nevertheless, we find a good deal of agreement. For example, Loewenstein and Adler (1995) find that (unendowed) people underappreciate how the endowment effect will alter their selling price by about 31%. Other papers estimate models of intrapersonal projection analogous to our model in Section 3.3, where α measures the extent of projection. Conlin, O'Donoghue and Vogelsang (2007) and find $\alpha \in [0.31, 0.50]$ for cold-weather clothing catalog sales, while Augenblick and Rabin

[24]Comparing Tables 6 and 1 reveals that predicting workers were significantly less willing to work than tired workers (difference of 3.42 total tasks; significant at $p = .045$). Recall that these two groups were nearly identical except the former made predictions about their eventual WTW, and the latter did not. Hence, stating predictions seemed to have a *negative* effect on eventual effort. This finding stands in contrast to research suggesting that stated goals form a reference point (e.g., Heath, Larrick, and Wu, 1999). However, our experiment is not well-suited to draw such conclusions.

[25]We believe this limit is implausible because it requires a large, directional error when thinking about others' cost functions, but no such error when thinking about one's own. An extensive psychology literature suggests that such an asymmetry is unlikely (see van Boven et al., 2013 for a review).

(2019) find $\alpha \in [0.27, 0.53]$ in a real-effort experiment.[26,27] We thus join an emerging consensus suggesting that—across a variety of domains—intrapersonal projection bias alters predictions and behavior by about 25-50%. Moreover, we add to a literature in psychology proposing that empathy gaps stem in part from using oneself as a simulation for others (for a review, see Van Boven et al., 2013). That is, insofar as intrapersonal projection bias skews predictions, we should expect to see a similar bias in interpersonal predictions.

6 Conclusion

In this paper, we provide evidence of interpersonal projection bias. Participants in our experiment failed to fully appreciate differences in others' tiredness when guessing their WTW on a tedious task. This mistake was meaningful in magnitude and led to costly errors in predictions. Our evidence suggests that neither uncertainty about the task nor learning were the root cause of these errors. We also measured intrapersonal projection bias—mispredictions about oneself—and find that intrapersonal projection was substantial, albeit likely smaller in magnitude than interpersonal projection.

Thus far we have interpreted our results as evidence for intrapersonal projection of tastes. Although there are other forms of projection—in particular, information projection—we believe that this alternative form likely does not explain the magnitudes we observe. Information projection (à la Madarász, 2012) is the idea that people act as if others have access to their private information. In our setting, this would map to predictors acting as if their beliefs about the onerousness of the task were known by the workers about whom they were guessing. On first glance, this alternative seems to explain some of our results, particularly if participants (i) initially underestimated how onerous the task would become once tired and, (ii) subsequently revised those beliefs downward after gaining experience. However, upon closer inspection, we believe that information projection is an unlikely explanation for our findings. Specifically, this explanation implies that tired predictors believed that their additional experience would be useful information to fresh workers when

[26] Augenblick and Rabin (2019) consider a real-effort experiment similar to our domain and find that projection bias leads tired workers to commit to doing fewer tasks in the future than their fresh counterparts. The authors offer caution in the precision of their estimates of α since their estimation procedure requires strong assumptions on the effort-cost curve. Moreover, their experiment also examined present bias, and their ability to separately measure projection bias was somewhat limited by their design.

[27] Although there are a number of studies on projection bias, many—particularly early experimental studies—are not suited to estimate the degree of projection. Likewise, some recent empirical papers cannot estimate projection bias directly, but find support for its main premise. For example, Chang, Huang, and Wang (2018) find that Chinese consumers are more likely to purchase health insurance on days with high pollution and are likely to reverse this decision (during a cooling-down period) when pollution drops. Busse et al. (2015) find that people are more likely to buy a convertible car on sunny days than on overcast or rainy days. Both papers are consistent with projection bias (and other mechanisms such as salience) but cannot directly estimate the degree of projection.

those workers stated their WTW.[28] That is, a tired predictor must have believed that fresh workers would, on average, change their WTW if they were to learn *the predictor's* WTW in the tired state. Given that these workers only stated their WTW when fresh, we believe this information would be of limited utility to them. Nevertheless, future work should endeavor to better disentangle these various sources of projection. Although we believe that information projection plays (at most) a small role in our experiment, replicating our experiment in a setting with no scope for learning would further clarify the underlying mechanism.

References

AMBUEHL, S., D. BERNHEIM, AND A. OCKENFELS (2019): "Projective Paternalism," *Working paper*.

AUGENBLICK, N. AND M. RABIN (2019): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," *The Review of Economic Studies*, 86(3), 941–975.

BADGER G., BICKEL W., GIORDANO L., JACOBS E., LOEWENSTEIN, G. AND L. MARSCH (2007) "Altered States: the Impact of Immediate Craving on the Valuation of Current and Future Opioids," *Journal of Health Economics*, 26(5), 865–876.

BLANCO, M. , ENGELMANN, D. , AND H. NORMANN (2011): "A Within-subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior,* 72, 321–338.

BOHREN, A. AND D. HAUSER (2018): "Learning with Model Misspecification: Characterization and Robustness," *Working paper*.

BROWN, C. (1982): "A False Consensus Bias in 1980 Presidential Preferences," *Journal of Social Psychology*, 118, 137–138.

BUCHANAN, J. (2020): "My Reference Point, Not Yours," *Journal of Economic Behavior and Organization*, 171: 297–311.

BUSSE, B., D. POPE, J. POPE, AND J. SILVA-RISSO (2015): "The Psychological Effect of Weather on Car Purchases," *Quarterly Journal of Economics*, 130(1), 371–414.

BUTLER, J., GIULIANO, P., AND L. GUISO (2016): "The Right Amount Of Trust," *Journal of the European Economic Association*, 4(5), 1155–1180.

[28] This explanation additionally requires that participants were initially overoptimistic about how onerous the task would become. Such optimism seems limited in our data. In the self-predictions from Table 6, participants guessed that their own WTW would be 22.11 tasks when tired. However, the actual WTW among tired workers was 20.44 (Table 1). This is suggestive evidence that, while participants may have held slightly optimistic views about their future disutility of effort, these early beliefs were likely not sufficiently biased to drive the information-projection story here.

CAMERER C., LOEWENSTEIN, G., AND M. WEBER. (1989): “The Curse of Knowledge in Economic Settings: An Experimental Analysis, ”*Journal of Political Economy*, 97(5), 1232–1254.

CHANG, T., W. HUANG, AND Y. WANG (2018): “Something in the Air: Pollution and the Demand for Health Insurance,” *The Review of Economic Studies*, 85(3), 1609–1634.

CONLIN, M., T. O’DONOGHUE, AND T. VOGELSANG (2007): “Projection Bias in Catalog Orders,” *American Economic Review*, 97(4): 1217–1249.

CRONBACH, L. (1955): “Processes Affecting Scores on ‘Understanding Others’ and ‘Assumed Similarity’,”*Psychological Bulletin*, 52, 177–193.

CRUCES, G., R. PEREZ-TRUGLIA, AND M. TETAZ (2013): “Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment,” *Journal of Public Finance*, 98, 100–112.

DAWES, R. (1989): “Statistical Criteria for Establishing a Truly False Consensus Effect,” *Journal of Experimental Social Psychology*, 25(1), 1–17.

DAWES, R. (1990): “The Potential Nonfalsity of the False Consensus Effect,” in *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, edited by R. Hogarth, University of Chicago Press.

ELLINGSEN, T., JOHANNESSON, M., TORSVIK G., AND S. TJØTTA (2010): “Testing Guilt Aversion,” *Games and Economic Behavior,* 68(1), 95–107.

ENGELMANN, D. AND M. STROBEL (2000): “The False Consensus Effect Disappears if Representative Information and Monetary Incentives Are Given,” *Experimental Economics*, 3, 241–260.

ENGELMANN, D., AND M. STROBEL (2012): “Deconstruction and Reconstruction of an Anomaly,” *Games and Economic Behavior*, 76, 678–689.

FARO, D., AND Y. ROTTENSTREICH (2006): “Affect, Empathy, and Regressive Mispredictions of Others’ Preferences Under Risk,” *Management Science*, 52(4), 529–541.

FREDERICK, S. (2012): “Overestimating Others’ Willingness to Pay,” *Journal of Consumer Research*, 39(1), 1–21.

FRICK, M., R. IIJIMA, AND Y. ISHII (2020): “Misinterpreting Others and the Fragility of Social Learning,” *Econometrica*, forthcoming.

GAGNON-BARTSCH, T. (2016): “Taste Projection in Models of Social Learning,” *Working paper*.

GAGNON-BARTSCH, T., M. PAGNOZZI, AND A. ROSATO (2020): “Projection of Private Values in Auctions,” *Working paper*.

GOEREE, J., AND J. GROSSER (2007): “Welfare-Reducing Polls,” *Economic Theory*, 31(1), 51–68.

GOLDMAN, A.I. (1993): "The Psychology of Folk Psychology," *Behavioral and Brain Sciences*, 16, 15–28.

GORDON, R.M. (1986): "Folk Psychology as Simulation," *Mind and Language*, 1(2), 158–171.

HEATH, C. LARRICK, R. AND G. WU (1999): "Goals as Reference Points," *Cognitive Psychology*, 38(1), 79–109.

ICHHEISER, G. (1949): "Misunderstandings in Human Relations: a Study in False Social Perception," *American Journal of Sociology*, 55(2), 70.

KATZ, C., AND F. ALLPORT (1931): *Students' attitudes*, Syracuse, NY: Craftsman Press.

KAUFMANN, M. (2019): "Overwork Through Projection Bias," *Working paper*.

KLEIN R.A., VIANELLO M., HASSELMAN F., ET AL. (2018): " Many Labs 2: Investigating Variation in Replicability Across Samples and Settings," *Advances in Methods and Practices in Psychological Science,* 1(4), 443–490.

KRUGER, J. AND D. DUNNING (1999): "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 77(6), 1121–1134.

KURT, D. AND J. INMAN (2013): "Mispredicting Others' Valuations: Self-Other Difference in the Context of Endowment," *Journal of Consumer Research*, 40(1), 78–89.

LOEWENSTEIN, G. AND D. ADLER (1995): "A Bias in the Prediction of Tastes," *Economic Journal*, 105(431), 929–937.

LOEWENSTEIN, G., T. O'DONOGHUE AND M. RABIN (2003): "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 118(4), 1209–1248.

MADARÁSZ, K. (2012): "Information Projection: Model and Applications," *Review of Economic Studies*, 79(3), 961–985.

MARKS, G. AND N. MILLER (1987): "10 Years of Research on the False-Consensus Effect: An Empirical and Theoretical Review," *Psychological Bulletin*, 102(1), 72–90.

MORTON, A. (1980): *Frames of Mind*, Oxford: Oxford University Press.

MULLEN, B., J. ATKINS, D. CHAMPION, C. EDWARDS, D. HARDY, J. STORY, AND M. VANDERLOK (1985): "The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests," *Journal of Experimental Social Psychology*, 21(3), 262–283.

READ, D. AND B. VAN LEEUWAN (1998): "Predicting Hunger: The Effects of Appetite and Delay on Choice," *Organizational Behavior and Human Decision Processes*, 76(2), 189–205.

ROUHANA, N., A. O'DWYER, AND S. VASO (1997): "Cognitive Biases and Political Party Affiliation in Intergroup Conflict," *Journal of Applied Social Psychology*, 27(1), 37–57.

ROSS, L., D. GREENE, AND P. HOUSE (1977): “The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes,” *Journal of Experimental Social Psychology*, 13, 279–301.

SHERIF, M., AND D. HOVLAND (1961): *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*, New Haven, CT: Yale University Press.

VAN BOVEN, L., AND G. LOEWENSTEIN (2003): “Social Projection of Transient Drive States,” *Personality and Social Psychology Bulletin*, 29(9): 1159–1168.

VAN BOVEN, L., G. LOEWENSTEIN, AND D. DUNNING (2003): “Mispredicting the Endowment Effect: Underestimation of Owners’ Selling Prices by Buyers’ Agents,” *Journal of Economic Behavior and Organization*, 51: 351–365.

VAN BOVEN, L., G. LOEWENSTEIN, D. DUNNING, AND L. NORDGREN (2013): “Changing Places: A Dual Judgment Model of Empathy Gaps in Emotional Perspective Taking.” In J. M. Olson and M. P. Zanna (Eds.), *Advances in Experimental Social Psychology*, 48, 118–171. San Diego, CA: Academic Press.

A Supplemental Material

A.1 All Predictions by Group and Timing

Recall that the true WTW of fresh workers was 22.1 tasks for \$2, while that of tired workers was 20.4 tasks for \$3. Table A1 shows all predictions of these quantities for each predictor group.

Table A1: ALL PREDICTIONS (NUMBER OF TASKS)

	In Group		Out Group A		Out Group B	
	Fresh	Tired	Fresh	Tired	Fresh	Tired
State Guessing About						
Fresh (after 5 tasks)	21.62	17.40	n.a.	22.48	n.a.	16.89
	(1.209)	(1.048)		(1.410)		(1.065)
Tired (after 20 tasks)	n.a	18.95	30.65	22.62	n.a.	18.27
		(1.044)	(1.473)	(1.378)		(1.127)
Observations	223	223	221	221	222	222

Notes: Standard errors are in parentheses.

A.2 Confidence Measures

Recall that after each guess, predictors reported their confidence in that guess on a five-point scale, where 1 represents "Not at all confident" and 5 represents "Extremely confident". Table A2 reports average responses.

Table A2:
SELF-REPORTED CONFIDENCE BY GROUP

	1st Prediction	2nd Prediction	3rd Prediction
Group *I*	3.26	3.34	3.27
	(0.062)	(0.066)	(0.072)
Group *A*	3.37	3.32	3.34
	(0.060)	(0.065)	(0.065)
Group *B*	-	3.57	3.52
		(0.060)	(0.065)

Notes: Standard errors are in parentheses. Survey used five-point scale ranging from 1: "Not at all confident" to 5: "Extremely confident".

A.3 The (Non) Effect of Task-Completion Time

Here, we discuss how task completion time had no apparent effects on the decisions of predictors from Groups *I* and *A*.[29] Our analysis splits the horizon of the experiment into two parts: (i) the "early phase" is the segment prior to completing five tasks, and (ii) the "late phase" is the segment after completing five tasks yet prior to completing twenty tasks. In Table A3, we show that neither predictors' first guesses nor their first confidence ratings were affected by the amount of time it took them to complete the early phase. We then repeat this analysis for predictors' second guesses and second confidence ratings in Table A4: the time taken to complete the late phase similarly had no effects.

It's worth noting that this analysis drops some extreme outliers. There was a great deal of heterogeneity in task completion time. While the average participant took a little over six minutes to complete the early phase (mean completion time is 386 seconds), that number is inflated by outliers (median completion time is 295 seconds). Excessively long completion times likely stemmed from inattention to the experiment. In particular, eleven subjects took more than twenty minutes to complete the early phase, and ten subjects took over an hour to complete the late phase; we

[29] We focus on these two groups because we examine predictions cast when both fresh and tired; Group *B* only cast predictions when tired.

drop these subjects from the regressions in Tables A3 and A4. Throughout, we fail to find any convincing evidence suggesting that the amount of time participants took to complete the work altered their predictions or confidence.

Table A3:
EFFECT OF EARLY-PHASE COMPLETION TIME ON PREDICTIONS

Estimation Technique:	OLS		Ordered Probit
	1st Prediction	1st Prediction	1st Confidence
Work Time (Seconds)	-0.004	-	-
	(0.005)	-	-
Work Time $\times$ $\mathbb{I}\{\text{Group } I\}$	-	-0.008	-0.0005
	-	(0.008)	(0.0004)
Work Time $\times$ $\mathbb{I}\{\text{Group } A\}$	-	0.001	0.00006
	-	(0.008)	(0.0004)
$\mathbb{I}\{\text{Group } I\}$	22.72	24.30	-
	(2.293)	(2.984)	-
$\mathbb{I}\{\text{Group } A\}$	31.47	29.96	-
	(2.317)	(2.954)	-
Observations	433	433	433

Notes: Standard errors are in parentheses. Dependent variable is listed in column header. Confidence ratings coded 1 (“Not at all confident”) to 5 (“Extremely confident”). Ordered probit includes separate estimation of cuts by group.

Table A4:
EFFECT OF LATE-PHASE COMPLETION TIME ON PREDICTIONS

Estimation Technique:	OLS		Ordered Probit
	2^{nd} Prediction	2^{nd} Prediction	2^{nd} Confidence
Work Time (Seconds)	-0.001	-	-
	(0.001)	-	-
Work Time $\times$ $\mathbb{I}\{\text{Group } I\}$	-	-0.0003	-0.0007
	-	(0.002)	(0.001)
Work Time $\times$ $\mathbb{I}\{\text{Group } A\}$	-	-0.0005	-0.0001
	-	(0.002)	(0.001)
Work Time $\times$ $\mathbb{I}\{\text{Group } B\}$	-	0.003	-0.0002
	-	(0.002)	(0.001)
$\mathbb{I}\{\text{Group } I\}$	20.21	19.34	-
	(1.578)	(2.298)	-
$\mathbb{I}\{\text{Group } A\}$	23.78	29.96	-
	(1.625)	(2.103)	-
$\mathbb{I}\{\text{Group } B\}$	18.13	19.87	-
	(1.570)	(2.201)	-
Observations	433	433	433

Notes: Standard errors are in parentheses. Dependent variable is listed in column header. Confidence ratings coded 1 (“Not at all confident”) to 5 (“Extremely confident”). Ordered probit includes separate estimation of cuts by group.

B Experimental Instructions

B.1 Workers

Preliminary Instructions

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 tricks. You will do a simple task and then we will ask you about your willingness to do additional tasks. You will earn at least the fixed payment of $3. Depending on your willingness to work, you may earn more. You must complete the session 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.

Overview

The experiment is simple, but we want to make sure you understand the basic structure.

1. We will review the real-effort task and you will complete some tasks.
2. We will ask you about doing additional work for additional pay.

You will know that you have reached the end of the survey when you see a screen saying “THIS IS THE END OF THE SURVEY”. Please do not exit until you have seen this screen. This final screen includes a code that you must input into MTurk in order to get paid.

Task

The task in the experiment involves counting. You will see an image like the image below:

!	0	)	0	0	0	?	(	i	!	0	)	!!	!	0
!!	0	t	!	0	t	!	)	t	0	(	!	t	(	?
!	!	!	t	)	!	!	(	!	0	!	!	0	!	!!
!!	(	!	)	!	!	t	t	?	!	!	0	0	!	1
!	!	0	0	!	(	)	0	t	!	)	?	0	0	i
t	1	!	!	0	!	0	!!	!	t	!	!	!	!	0
t	!	(	!	!	0	!	(	(	!	!!	!	!	t	?
!	0	!	0	0	0	!	0	!	!	)	!	i	!	0
(	!	!	!	!	t	!	0	!	!	!	!	!	(	!
0	(	0	(	0	)	!	!	i	0	(	t	t	0	0

You will then be asked to count a specific character that is present in the image. The question will be phrased as: How many are in the picture?

Symbol to count: t

This means you should count how many ”t” there are in the image.

The symbol that you will count will change in each image, so pay close attention. To make the task harder (and to prevent cheating) we have included two symbols that are very close to one another: ! and !!

These are different. So if you are asked to count ! in the image above, there are 61. If you are asked to count !!, there are only 6. Do not count !! when counting !

PLEASE NOTE: You must type the exact correct answer in order to advance to the next image. Counting each image should take about 30 seconds.

This is the end of the instructions. Reminder: you will be asked questions about your willingness to complete more of this task for additional pay at the end of this initial block of work. You will complete 5 *[Alternate: 20]* tasks in this initial block of work. When you click to advance to the next slide, you will begin.

[Here, the participant completed either five or twenty tasks.]

Willingness to Work

As of right now, you have earned $3 for completing the tasks and for your overall participation in this study. In a few moments, we will ask you one question about your willingness to do additional tasks to increase your payout.

You have already sampled the task and we will ask you about your willingness to complete more of the same task. The task is not different from your sample experience, except that you would have different tables to count.

We will ask you just one question, and this question will count for real. Your choice will determine whether you must complete additional tasks and whether you might earn additional pay.

We will use a specific system to ensure you answer truthfully. The next few pages will explain this in detail.

The method we use to determine whether you will complete extra tasks 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. Here's how the system works.

First, we will ask you how many additional tasks (counting matrices) you are willing to do for a fixed amount of money.

For instance, we might ask: "What is the maximum number of extra tasks 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 slider that goes between 0 and 100 tasks. You will also see an amount of money next to the slider.

You will move the slider to indicate the maximal number of tasks you'd be willing to do for that 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.

We will then draw a random number between 0 and 100. If your answer is less than that random number, you will not do 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 $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 tasks.

The next pages have a short quiz to help clarify how this works.

Suppose you were asked ”What is the maximum number of additional tasks you are willing to do for $0.80?” and you responded 60. If the random number is 17, how many tasks will you complete? *[Four multiple-choice answers; subject must answer 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 tasks equal to the random number.

Suppose you were asked ”What is the maximum number of additional tasks you are willing to do for $0.80?” and you responded 60. If the random number is 76, how many additional tasks will you complete? *[Four multiple-choice answers; subject must answer correctly.]*

Correct. If the random number is greater than your choice, you will complete zero tasks and you will not receive an extra payment.

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, for example, you’d be willing to do 20 tasks for $0.40 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.

We will now ask you the question about your willingness to do additional tasks for additional payment. Remember, we are using the method just described, so answer honestly.

The next screen is the real question, so think carefully.

What is the maximal number of additional tasks you’re willing to complete for $2? *[]Alternate: $3]*

[Slider here.]

We’ll now draw the random number to determine if you complete additional tasks.

Since the random number was higher than the number you were willing to do, you will not complete any supplemental tasks and you will be paid any additional earnings. *[Alternate: Since you were willing to work, you will now complete supplemental tasks and you will be paid $2 / $3 additional earnings when you complete the survey]*.

Thank you for participating. This is the last screen before the MTurk code.

Your responses have been stored. The code to input into Amazon's MTurk is on the screen that follows. Payments will be processed within one week.

Please click the final button below to submit your work.

B.2 Predictors

Preliminary Instructions

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 tricks. This experiment is about your ability to predict others' behavior. You will do a simple task and you will predict how many additional tasks other people would do for additional money. You will earn at least the fixed payment of $3. Depending on your ability to guess others' behavior, you may earn more. You must complete the session 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.

Overview

The experiment is simple. First, we want to make sure you understand the basic structure.

1. We will review the real-effort task and you will complete some tasks to help you learn.
2. We will interrupt you after 5 tasks and you will make a prediction about other people.
3. You will complete 15 additional tasks.
4. You will make two other predictions about other people.

You will know that you have reached the end of the survey when you see a screen saying "THIS IS THE END OF THE SURVEY". Please do not exit until you have seen this screen. This final screen includes a code that you must input into MTurk in order to get paid.

Predictions

More than 500 people have already completed different versions of this experiment. In those other experiments, they simply completed tasks and we asked them their willingness to complete additional tasks for additional payment.

Specifically, we asked them "What is the maximum number of tasks you are willing to complete for _?" where we inserted different amounts of money into the blank spot. We asked some people this question after they had completed 5 tasks. We asked other people this question after they had completed 20 tasks.

You will try to guess the average answer to this question. That is, you will guess how many tasks they were willing to do, and you will be given a bonus if you're correct.

In order to help you guess, over the next few slides you will work through the same instructions that the other participants did. You will also complete tasks like they did. Therefore, the total amount of time for the experiment for you should be similar to the total amount of time it took others.

Task

The task in the experiment involves counting. You will see an image like the image below:

!	0	)	0	0	0	?	(	i	!	0	)	!!	!	0
!!	0	t	!	0	t	!	)	t	0	(	!	t	(	?
!	!	!	t	)	!	!	(	!	0	!	!	0	!	!!
!!	(	!	)	!	!	t	t	?	!	!	0	0	!	1
!	!	0	0	!	(	)	0	t	!	)	?	0	0	i
t	1	!	!	0	!	0	!!	!	t	!	!	!	!	0
t	!	(	!	!	0	!	(	(	!	!!	!	!	t	?
!	0	!	0	0	0	!	0	!	!	)	!	i	!	0
(	!	!	!	!	t	!	0	!	!	!	!	!	(	!
0	(	0	(	0	)	!	!	i	0	(	t	t	0	0

You will then be asked to count a specific character that is present in the image. The question will be phrased as: How many are in the picture?

Symbol to count: t

This means you should count how many "t" there are in the image.

The symbol that you will count will change in each image, so pay close attention. To make the task harder (and to prevent cheating) we have included two symbols that are very close to one another: ! and !!

These are different. So if you are asked to count ! in the image above, there are 61. If you are asked to count !!, there are only 6. Do not count !! when counting !

PLEASE NOTE: You must type the exact correct answer in order to advance to the next image. Counting each image should take about 30 seconds.

Predictions and Overview

Some participants completed only five tasks. Others completed 20. You will guess about both.

As a reminder, the steps coming up are as follows:

1. You will complete 5 tasks in this initial block of work.
2. We will ask you about your prediction about others.
3. You will complete 15 more tasks.

4. We will ask you for two additional predictions.

When you click to advance to the next slide, you will begin.

[Here, the participant completes five tasks.]

Predictions

In a moment, you will make your first prediction about others' willingness to do additional tasks.

[Alternate, Outgroup B: After you complete 15 more tasks, you will make predictions about others' willingness to do additional tasks.]

In order to give you more information about the specific questions we asked and the environment that others faced, you will work through very similar instructions to the instructions from our earlier experiments.

As a reminder: your goal will be to guess how many additional tasks a person was willing to do for some additional payment.

We will describe the method we used to ensure people in the previous experiments answered truthfully. It may seem complicated. But we'll walk through it step-by-step. The punchline: it was in their best interest to just answer truthfully.

Here's how the system works.

First, we asked them how many additional tasks (counting matrices) they were willing to do for a fixed amount of money.

Specifically, we asked questions of the form: "What is the maximum number of extra tasks you are willing to do for \$0.40?" This question meant that we would give them \$0.40 in exchange for completing some amount of additional work.

On the decision screen, they were presented with a slider that went between 0 and 100 tasks, and they also saw an amount of money next to the slider.

They would move the slider to indicate the maximal number of tasks they were willing to do for that amount of money.

That is, if they were willing to do 15 additional tasks but not 16, then they should have moved the slider to 15.

We then drew a random number between 0 and 100. If the person's answer was less than that random number, they did not do additional tasks and they received no additional payment.

However, if their answer was greater than or equal to that random number, they completed a number of additional tasks equal to the random number and received the additional payment.

Example: Suppose the person indicated they were willing to do 15 additional tasks for \$0.40. If the random number was 16 or higher, they would do no additional tasks. However, if the random number was 12, they would do 12 additional tasks.

While this may seem complicated, the punchline from this setup is that participants should have simply answered truthfully. We told them this in the same manner we have just told you.

We will now ask you to PREDICT how many additional tasks. people were willing to do—on average—for an additional payment. Please pay attention to the amount of money involved in the question. You will make three predictions in this experiment, and the amount will change.

*[Alternate, Outgroup B: You will now continue and complete 15 additional tasks. Afterwards, we will ask you to make two predictions. (Participant skips to * below)]*

You are predicting about people who also completed five (5) tasks. That is, they completed 5 tasks, read instructions very similar to those you just completed, and then we asked their willingness to do additional tasks.

The next screen is the real question, so think carefully. If your guess is within 5 tasks of the correct answer, you will receive $0.50

Think about people **who just completed five tasks.**

What do you think is the (average) maximal number of additional tasks they would be willing to complete for $2.00?: *[Alternate, Outgroup A: Think about people who just completed 20 tasks. What do you think is the (average) maximal number of tasks they would be willing to complete for $3.00?]*

[Slider here]

We're curious how confident you are about your answer on the previous screen. Your answer to this question will not affect your pay.

Not at all A tiny bit So-so Fairly confident Extremely confident

You will now continue and complete 15 additional tasks.

*[*Here, the participant completes 15 tasks.]*

Afterwards, we will ask you to make two other guesses. As of right now, you have earned $3 for completing the tasks and for your overall participation in this study. In a few moments, we will ask you to make two additional predictions about others' willingness to complete additional work for additional pay.

This time, you will make predictions about two different groups of people:

1. You will PREDICT how many additional tasks people were willing to do after they completed a total of 20 tasks *[Alternate, Outgroup A,B: 5 tasks]*. That is, they completed 20 tasks and then we asked their willingness to do additional tasks.

2. You will PREDICT how many additional tasks different people were willing to do after they completed a total of 5 tasks *[Alternate, Outgroup A,B: 20 tasks]*. That is, they completed 5 tasks and then we asked their willingness to do additional tasks.

The few screens are the real questions, so think carefully. For each prediction, if your guess is within 5 tasks of the correct answer, you will receive $0.50

Think about people who just completed twenty tasks.

What do you think is the (average) maximal number of additional tasks they would be willing to complete for \$3.00:

[Slider here]

We're curious how confident you are about your answer on the previous screen. Your answer to this question will not affect your pay.

Not at all A tiny bit So-so Fairly confident Extremely confident

Think about people who just completed five tasks.

What do you think is the (average) maximal number of additional tasks they would be willing to complete for \$2.00:

[Slider here]

We're curious how confident you are about your answer on the previous screen. Your answer to this question will not affect your pay.

Not at all A tiny bit So-so Fairly confident Extremely confident

Finally, imagine we asked you the following after completing 20 tasks. (Note that your answer to this question will not affect your pay, nor will you have to do any additional tasks).

What is the maximal number of additional tasks you would be willing to complete for \$3.00:

[Slider here]

Thank you for participating. This is the last screen before the MTurk code.

Your responses have been stored. Since others are completing this experiment at the same time as you and to avoid information becoming public, we won't tell you if you were correct at this time. Any bonus payments will be processed within one week.

Please click the final button below to submit your work.