Failures in Forecasting: An Experiment on Interpersonal Projection Bias

Benjamin Bushong
Michigan State University

Tristan Gagnon-Bartsch*
Florida State University

April 25, 2023

Abstract

Using a real-effort experiment, we show that people project their current tastes onto others when forecasting others' behavior, even when their tastes are exogenously manipulated and transparently different. In the first part of our experiment, "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 guessed the WTW of workers when they were in the same state as the workers about whom they were predicting, but, (iii) when fresh predictors were guessing about tired workers, they substantially overestimated their WTW, and (iv) when tired predictors were guessing about fresh workers, they underestimated their WTW. Using an additional treatment, we find that workers also mispredicted their own future WTW, and we compare the magnitudes of intra- and interpersonal projection bias.

JEL Classification: C91, D84, D91.
Keywords: Projection bias, Real-effort experiments, Forecasting, State-dependent preferences.

*E-mails: bbushong@msu.edu and tgagnonbartsch@fsu.edu. We thank Andy Brownback, Jon X. Eguia, Christine Exley, Laura Gee, Kristof Madarász, Kirby Nielsen, Matthew Rabin, Antonio Rosato, Joshua Schwartzstein, and audiences at the ESA World Conference in Vancouver, Chapman Univeristy, Michigan State University, Purdue 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? . . . [Y]ou shouldn't just *project your own* situation and psychology on the other."
>
> — Robert M. Gordon

Predicting others' preferences is a ubiquitous feature of economic and organizational interactions. For instance, estimating others' valuations is central to bidding in an auction, attending to a counterparty's goals enables effective negotiation, and being mindful of others' recent workloads is important for managers when allocating tasks, since fatigue will diminish workers' productivity and willingness to take on additional work. Many other aspects of organizational decision-making—e.g., motivating effort, advertising, and sales planning—similarly require some form of prediction about what others might do. But do people accurately forecast others' preferences?

In this paper, we present experimental evidence that people's predictions about others' willingness to work (WTW) on a real-effort task are biased by their own current fatigue. Our evidence supports a simple model of "interpersonal projection bias": people project their current tastes onto others, even when others face transparently different circumstances. This basic idea speaks to a recent literature showing that people—including experts—sometimes struggle to forecast how others will respond to incentives and experimental interventions (DellaVigna and Pope 2018a, 2018b, 2022; DellaVigna, Pope, and Vivalt, 2019). Our results suggest that projection bias may be a potential source of this prediction error, and a natural way to mitigate this error is to put forecasters in states similar to those of the people they are estimating. Furthermore, while projection bias has been shown among people predicting their own future preferences (see, e.g., Read and van Leeuwen, 1998; Badger et al., 2007; Conlin, O'Donoghue, and Vogelsang, 2007; Busse et al., 2015), we measure both inter- and intrapersonal projection bias in the same setting and compare their relative magnitudes. We find suggestive evidence that interpersonal projection bias may be larger: that is, a person's current tiredness distorts her predictions about others' WTW by more than it distorts predictions about her own future WTW.

Participants in our experiment worked on a tedious real-effort task. In the first part of our experiment, we elicited participants' willingness to continue working on the task for additional pay. In a second part, different participants cast incentivized predictions about the 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 WTW: 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 their initial workloads: some predictors completed five tasks before guessing the WTW of others, while others completed twenty. Our central question is whether—and to what extent—predictors abstracted from their own state (i.e., their current tiredness) when making guesses about others' preferences.

We present five stylized facts about behavior and predictions in our primary experiment on interpersonal projection bias. (i) Tired workers were less willing to work than fresh workers. (ii) Predictors (in aggregate) accurately guessed 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.

Overall, these results suggest that participants in our experiment projected their current sense of tiredness onto others. We find—through both non-parametric measures and the estimates of a reduced-form model—that our manipulation of tiredness induced large and significant errors in predictions about the choices of others. While these predictions were accurate when guesses were about others in one's own state, our estimates suggest that guesses about others in a different state were systematically distorted (in the direction of the predictor's own tiredness) by 21-50%.

To further elucidate the mechanism driving these errors, we divided our predictors into three subgroups. These groups varied in how many guesses each predictor cast and their tiredness 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 workers when they themselves were tired. Comparing predictions across these two groups allows us to measure how predictors' estimates changed as their own states changed. Finally, a third group made their first prediction—about fresh workers—when they themselves were tired; 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. When focusing solely on initial predictions, we find that 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, comparing initial guesses across groups also 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 roughly similar magnitudes.

Additionally, we analyze how predictors' guesses about fresh workers changed as the predictors went from fresh to 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) once they themselves became tired, they substantially revised their guesses downward (by approximately 19%; difference significant at $p < .001$). This was a mistake. By revising their guesses, predictors significantly decreased their accuracy and lowered their expected earnings. Since predictors became less accurate even as they gathered more first-hand experience with the task, this suggests that our results do not stem from predictors simply lacking information about the task. Indeed, this analysis—and our collection of results more broadly—offers strong support for interpersonal projection bias.

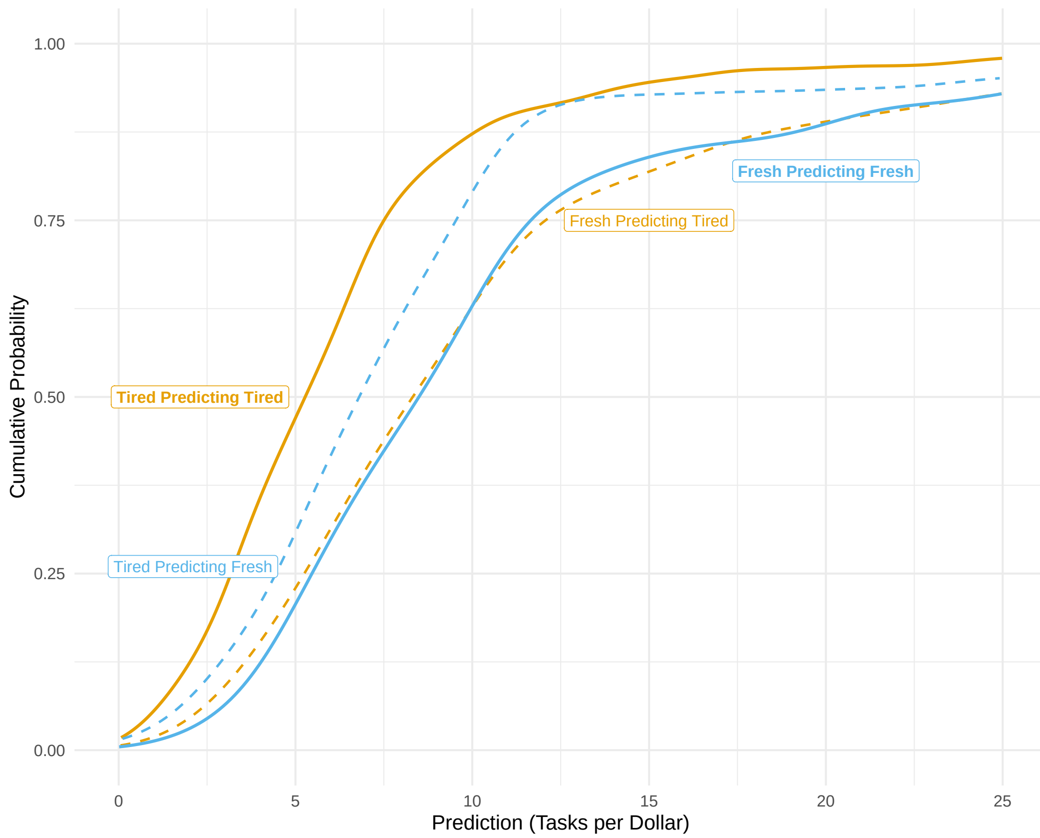

Figure 1: *Empirical (smoothed) CDFs of predictions about fresh (blue) and tired (orange) workers. Predictions are shown in units of "tasks-per-dollar"—the average number of tasks workers are willing to do for each dollar of compensation. The solid lines represent guesses cast by predictors in the same 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).*

Finally, we advance 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 the 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 the totality of errors that accumulate into interpersonal projection bias may be more severe than the intrapersonal analog.

In the conclusion, we discuss a related idea—information projection (e.g. Camerer, Loewenstein, and Weber, 1989; Madarász, 2012)—wherein people exaggerate the degree to which others share their private information (as opposed to their preferences). Some of our findings are consistent with information projection, but we discuss why information projection may not be a complete explanation for our collective evidence.

Despite its potential ubiquity, interpersonal projection bias has received relatively little attention in the economics literature. On the empirical side Van Boven, Loewenstein, and Dunning (2003) find that sellers in experimental markets project their sense of endowment onto potential buyers; Ambuehl, Bernheim, and Ockenfels (2021) 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 the related empirical evidence (including studies from psychology) in more detail, and we describe the variety of ways in which our study builds on this literature.

A few of these ways are worth emphasizing here. First, we transparently induced changes in tastes along a familiar dimension (tiredness). Second, 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. These two elements of our design reduce the possibility that biased predictions arose simply because predictors were either uncertain about other people's tiredness or were unfamiliar with the experience of being in the fresh or tired state. Third, we elicited multiple predictions from each participant, which allows us to obtain within-subject measures of how a person's guesses 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 benefit from greater engagement with the perspectives of others, particularly in domains involving effort provision and fatigue. For instance, a fresh manager who suffers from interpersonal projection bias may fail to design optimal incentive schemes because she underestimates how workers' tiredness will influence their responsiveness to incentives. Regardless of her own tiredness state, the manager will also systematically underestimate the heterogeneity in her workers' marginal disutility of effort and will therefore underestimate the value of strategically tailoring assignments to workers based on their recent workloads. Beyond the specific domain of fatigue, projection bias has many important implications for decision making in organizations. The strategies we use to motivate others will be overly-reliant on what would motivate ourselves at a given moment. The advice we give will be overly-tailored to our own tastes or circumstances. And the policies we advocate for will be overly-aligned with our own private objectives, even when we are directly concerned with satisfying others' preferences. Projection bias may also cloud our judgements of others. At times when we are relatively unburdened, we may wrongly attribute another person's limited productivity to intrinsic characteristics (e.g., low ability or laziness) rather than momentary hardship.

Our findings also speak to the emerging literature studying the accuracy of forecasts about behavioral responses to incentives and interventions (e.g., DellaVigna and Pope 2018a, 2018b, 2022; DellaVigna, Pope, and Vivalt, 2019). Those papers highlight that even though forecasters demonstrate some ability to predict others' responses to interventions, they still make significant forecasting errors. Furthermore, neither higher incentives nor added expertise substantially improve forecasters' guesses. Our experiments suggest that these predictions may improve when forecasters are themselves in a similar state as the people they are trying to predict. Given the pitfalls of projection described above, this practice may directly improve the decision-making of managers and researchers alike. In fact, researchers have recently explored this idea in the domain of policy and intervention design: Thomas et al. (2020) show that organizations can gain valuable information for designing such interventions by eliciting forecasts about their impact directly from the "local" population that the intervention intends to target. They emphasize that this is especially true when the target population may be in a different psychological state than those designing the intervention.

Finally, our evidence bolsters the empirical foundation for a growing theoretical literature emphasizing the broader implications and importance of interpersonal projection bias. For instance, projection of political preferences can generate inefficient election outcomes (Goeree and Grosser, 2007); projection of private valuations can lead to overbidding and inefficient allocations in auctions (Gagnon-Bartsch, Pagnozzi, and Rosato, 2021); and in social-learning contexts, such as the adoption of a new technology, mispredicting others' tastes can prevent people from inferring their optimal action (Gagnon-Bartsch, 2016; Bohren and Hauser, 2020; Frick, Iijima, and Ishii, 2020). Relatedly, Kaufmann (2022) shows how *intra*personal projection bias over effort can lead people to over-commit to and over-work towards projects that go unfinished 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 framework similar to that of Loewenstein, O'Donoghue, and Rabin (2003)—we derive 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 (e.g., Katz and Allport, 1931; Cronbach, 1955; Sherif and Hovland, 1961). In a seminal paper Ross, Greene, and House (1977) dub this error the "false-consensus effect".[1] Subsequent work has proposed various mechanisms that can generate a false-consensus effect, but distinguishing which factor drives this error in a given context remains elusive.

One explanation of the false-consensus effect is the idea that people project their preferences or states onto others. For instance, in Van Boven, Loewenstein, and Dunning (2003), endowed sellers overestimated the willingness to pay of potential buyers.[2] To further explore how a person's own temporary state distorts their predictions of others' preferences, 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 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.[3]

Our experiment drew inspiration from Van Boven and Loewenstein's (2003) design in that we

[1]Marks and Miller (1987) document the false-consensus effect in 45 different studies published in the decade following Ross, Greene, and House (1977). These studies generally 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 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 predicted that the bill would receive more support than those who voted against it). More recently, Selten and Ockenfels (1998) extend the false-consensus-effect paradigm to the expectation of others' behavior in a conditional gift-giving game.

[2]Buchanan (2020) similarly finds that participants neglected the effect of others' endowments when trying to predict their risk attitudes and acted as if others shared their own endowment.

[3]Relatedly, 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). We return to a discussion of intrapersonal projection in Section 5.

also exgoenously manipulated a predictor's state. Superficially, we focus on projecting fatigue 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 both 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.

Lastly, our predictors had first-hand experience with the situations experienced by the target groups: they worked on the same task as the workers, and they faced that task in both the fresh and tired states. This aspect of our design—combined with our exogenous manipulation of others' states—allows us to consider an argument put forth by Dawes (1989,1990) suggesting that projection may be rational (see also Krueger and Clement, 1994). An adaptation of this argument to our setting with state-dependent utility is as follows: a predictor's own WTW in a given state may provide information about others' WTW in that same state, and she may therefore guess that others will behave like herself simply because she lacks any further information about others' preferences (i.e., she acts as a Bayesian with a single data point, herself). This could manifest in "rational projection." We address this by (i) arming predictors with information about both states via first-hand experience; (ii) changing the taste of predictors (making them tired); and (iii) asking whether predictors use their WTW when tired to predict the WTW of fresh workers.[4] Rational projection would imply that predictors use their own WTW *when fresh* to predict the WTW of fresh workers. We explore—and reject—this notion.

In addition to the limited-information explanation discussed above, our design was intended to

[4]Furthermore, we told participants 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 interest to pay attention to and recall their sentiment toward work at these two points.

isolate the role of projection from other alternative mechanisms that are discussed in the false-consensus literature (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 sidestep these alternative channels in order to better identify whether projection distorts predictions. 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.

Finally, while several papers in the economics literature find indirect evidence of a false-consensus effect, a few specifically study this concept.[5] Ambuehl, Bernheim, and Ockenfels (2021) conducted an experiment where participants constructed choice sets on the behalf of others and were asked what they believed others would choose. The authors document a strong false-consensus effect, and the effect persisted in two settings where participants were given information on either a specific person's preferences or the distribution of other' preferences (see their Online Appendix B.4.).[6] Engelmann and Strobel (2000) find that the false-consensus effect disappears after participants are provided with signals about others' choices. However, Engelmann and Strobel (2012) suggest that the artificial nature of that environment—namely, the free acquisition of strong sig-

[5]Research focusing on separate questions has documented results akin to the false-consensus effect. For example, one line of papers suggests that people project their own social preferences or beliefs in settings involving guilt aversion, solidarity, and trust (e.g., Selten and Ockenfels, 1998; Charness and Dufwenberg, 2006; Vanberg, 2008; Ellingsen et al., 2010; Blanco et al., 2014; Butler, Giuliano, and Guiso, 2016). In a related vein, Ziegler, Romagnoli, and Offerman (2022) show that those whose morals have been eroded are overly pessimistic about the morals of others trading in the market.

[6]Our design shares a feature with that of Ambuehl, Bernheim, and Ockenfels (2021): choice Architects in their experiment also had first-hand experience making the choices faced by others. Our design goes further. Since we directly altered a predictor's own preferences by changing her state of fatigue, we can examine if and how first-hand experience with *different preferences* matters. This is akin to a Choice Architect in their experiment making choices for others when she is both patient and impatient; such an exercise is obviously infeasible in their setting.

nals about others—led to the null finding in their previous work. Our contribution relative to these papers is three-fold. First, by focusing on a setting where predictors have been "in the shoes" of the target group, we offer additional evidence of projection in a setting where uncertainty about others is reduced. Second, we transparently and exogenously manipulated tastes in this setting, allowing us to further isolate the effect of projection bias on predictions. Finally, as highlighted in the introduction, we measure both inter- and intrapersonal projection bias in the same domain.

3 Experimental Design

A total of 1,566 people participated in our experiments on Amazon's Mechanical Turk (MTurk).[7] Our experiment had two distinct, mutually exclusive parts which corresponded 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 part—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 part—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.

3.1 Workers

Participants in the first part of the experiment completed a set number of rounds of the task, and then we elicited their willingness to complete additional rounds. Workers were randomized into

[7]Participants were recruited 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. Complete experimental instructions are in Online Appendix **??**.

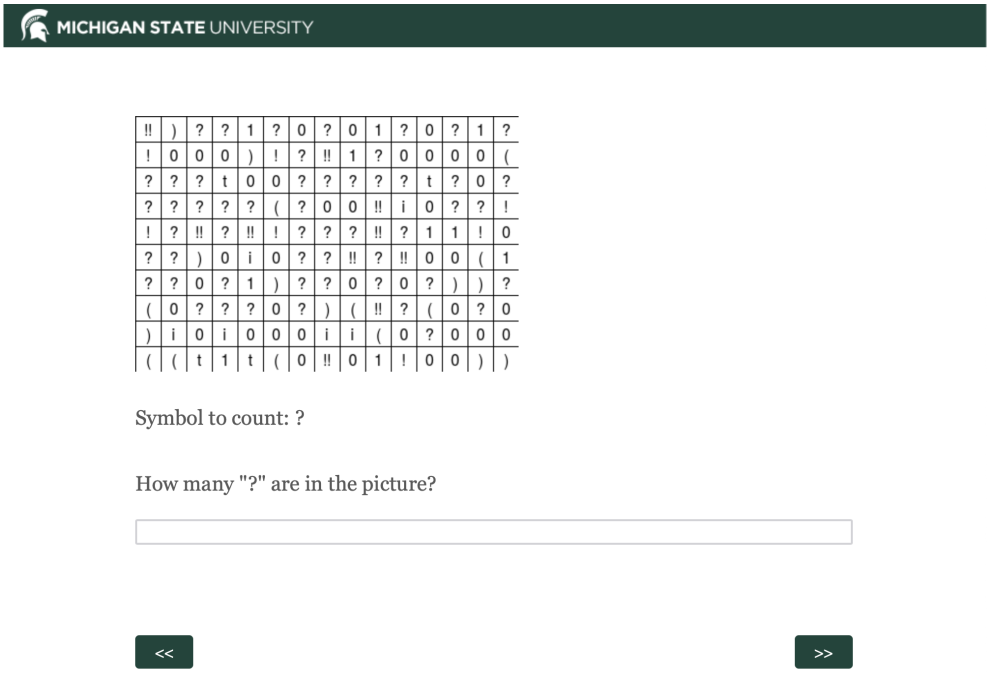

Figure 2: *Screenshot of the counting task.*

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 $\$m$. Participants used a slider to select a WTW between 0 and 100. We then randomly drew an integer z 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. We varied the bonus payment $m \in \{2, 3\}$ depending on the group (Fresh vs. Tired, respectively).[8]

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

[8]As we discuss below, these measures of WTW were the objects that predictors had to guess. We varied the monetary incentives across the two worker groups to ensure that a predictor faced distinct questions when asked about the two groups. This was intended to promote independent assessments for each prediction. Had we asked predictors the same question repeatedly, we may have introduced a consistency bias.

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 intra- and interpersonal 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" (Group I henceforth; $n = 223$) began by making predictions about others in their own state. To clarify:

> A predictor in *Group I* made 3 guesses in total: (1) after completing 5 tasks himself, he predicted $\overline{W}(\$2, F)$—the WTW of fresh workers; (2) after completing 20 tasks, he predicted $\overline{W}(\$3, T)$—the WTW of tired workers; and (3) immediately after the second prediction, he again predicted the WTW of fresh workers, $\overline{W}(\$2, F)$. This final prediction allows us to test whether a predictor changed his view of others simply as a result of becoming tired himself.
>
> 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: Group O ($n = 221$) cast three predictions in total, made when both fresh and tired; Group O_{-1} ($n = 222$) cast two predictions, only when tired. To clarify:

> A predictor in *Group* O made three guesses 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 WTW of tired workers, $\overline{W}(\$3, T)$.
>
> Note that for the first two predictions above, the predictor guessed the WTW of workers who were in a different state as herself: 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.
>
> A predictor in *Group* O_{-1} made two guesses after completing all 20 tasks. Aside from not making a guess after completing 5 tasks, a predictor in this group made the same guesses as

Group O: after completing 20 tasks, she predicted $\overline{W}(\$2, F)$—the WTW of fresh workers; immediately after, she 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 O, we interrupted each predictor in Group O_{-1} after 5 tasks. During this pause, 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 O_{-1} 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 $g \in \{I, O, O_{-1}\}$ about workers in state $s \in \{F, T\}$ facing bonus 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.[9] 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.[10]

Predictors received no feedback after making their predictions. Their payments—which were

[9]Note that this mechanism is not incentive compatible for eliciting point estimates for beliefs. In particular, any prediction below 5 (or above 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.

[10]Adding incentives to this last question would have substantially increased the length of the experiment for predictors, since participants would have actually had to complete additional work. Given 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.

based on the accuracy of their guesses—were revealed only once the experiment was over.

3.3 Theoretical Predictions

Interpersonal projection bias implies that a person's own state (fresh vs. tired) distorts her predictions about the WTW of others. Using a simple model, we now show that projection leads fresh predictors to overestimate the WTW of tired workers and leads tired predictors to underestimate the WTW of fresh 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 $C(e; s_i, \theta_i) \equiv c(e + s_i; \theta_i) - c(s_i; \theta_i)$, 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 "type", capturing an idiosyncratic taste for the task. We assume that the cost function $c(\cdot; \theta_i)$ is strictly increasing and convex for all θ_i, and thus 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, Participant i chooses e to maximize $\int_0^e [m - C(\tilde{e}; s_i, \theta_i)] d\tilde{e}$. Participant i's optimal choice is thus implicitly defined by the solution to

$$c(e + s_i; \theta_i) - c(s_i; \theta_i) = m, \tag{1}$$

which we denote 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.

Building from Loewenstein, O'Donoghue, and Rabin's (2003) model of *intra*personal projection, we now provide a simple model of *inter*personal projection bias in our experimental setting. Suppose that Participant i wrongly believes that Participant j's cost function is

$$\begin{aligned} \widehat{C}(e; s_j, \theta_j | s_i, \theta_i) &= \alpha C(e; s_i, \theta_i) + (1 - \alpha) C(e; s_j, \theta_j) \\ &= \alpha \left[c(e + s_i; \theta_i) - c(s_i; \theta_i) \right] + (1 - \alpha) \left[c(e + s_j; \theta_j) - c(s_j; \theta_j) \right], \end{aligned} \tag{2}$$

where parameter $\alpha \in [0, 1]$ captures the degree of projection bias. That is, a projecting predictor perceives another person's cost as a convex combination of her own cost and that other person's true cost.[11] 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 $\widehat{C}(e; s_j, \theta_j | s_i, \theta_i) = m$. Hence, i's prediction about the WTW of an 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 $\alpha > 0$.[12] 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 henceforth denote these two numerical states by F and T, respectively. Additionally, our formalization of projection bias in Equation (2) implies that a projector's predictions of others' WTW is biased by her own tiredness state, s_i, *and* her own idiosyncratic taste, θ_i. However, since we can only exogenously manipulate the former, our main hypotheses will compare predictions cast by predictors in state F versus state T.

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 $g \in \{I, O, O_{-1}\}$.

Hypothesis 2: *Relative to tired predictors, fresh predictors overestimate the WTW of tired workers:* $\widehat{W}^O(\$3, T|F) > \widehat{W}^g(\$3, T|T)$ for each $g \in \{I, O, O_{-1}\}$.

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 $g \in \{I, O, O_{-1}\}$.

[11] Equation 2 is inherently a model of interpersonal projection rather than *intra*personal projection because it does not presume that Participant i mispredicts her own future costs. That said, we could additionally capture intrapersonal projection by replacing $C(e; s_i, \theta_i)$ in Equation 2 with Participant i's current misperception of her cost function. Following Kaufmann (2022), an appropriate specification of one's own misperceived costs due to intrapersonal projection bias in our context would be $\widetilde{C}(e; s_i, \theta_i) = \tilde{\alpha} c'(s_i; \theta_i) e + (1 - \tilde{\alpha})[c(e + s_i; \theta_i) - c(s_i; \theta_i)]$, where $\tilde{\alpha} \in [0, 1]$ captures the degree of this bias. Although we present reduced-form evidence of intrapersonal projection in Section 5, our results in that section could be interpreted using this newly stated model.

[12] 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. We assume a predictor is Bayesian aside from the misspecified model of costs presented in (2). Thus, $\widehat{W}(m, s_j | s_i, \theta_i)$ is the expected value of effort that maximizes (2) given Predictor i's beliefs over θ_j.

A variety of biases in belief formation *independent* of projection bias could jeopardize Hypothesis 1 (see, e.g., Benjamin, 2019 for a review).[13] Hypotheses 2 and 3 are robust to such biases because they compare predictions across groups rather than compare the predictions of one group to the truth. Moreover, if Hypothesis 1 does hold, then we have two immediate corollaries of Hypotheses 2 and 3:

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

Hypothesis 3A: *Tired predictors underestimate the WTW of fresh workers:* $\widehat{W}^g(\$2, F|T) < \overline{W}(\$2, F)$ for each $g \in \{I, O, O_{-1}\}$.

3.4 Discussion

Some of our predictions above are potentially generated 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. Since fresh predictors had not yet completed 20 tasks themselves, they were unfamiliar with the state in which tired workers made decisions. Therefore, we are cautious to interpret an outcome in which $\widehat{W}^O(\$3, T|F) > \overline{W}(\$3, T)$ as stemming purely from projection bias. Note, however, that this limited-information channel is less relevant for *tired* predictors who guessed about the behavior of fresh workers—tired predictors had already experienced the fresh state. Accordingly, predictions cast by tired workers potentially provide a cleaner measure of projection bias. Furthermore, comparing the prediction errors across fresh and tired predictors sheds light on the extent to which this form of limited information distorted initial predictions. We present this analysis in Section 4.2.

A variant of the informational confound just discussed could still emerge, however, if tired

[13]Two recent experiments highlight how errors distinct from projection bias can warp interpersonal predictions. Frederick (2012) shows that people generally overestimate others' willingness to pay for goods, and Kurt and Inman (2013) demonstrate that both endowed and unendowed participants are inaccurate in their predictions about others in their *same* state.

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 O were interrupted after 5 tasks (i.e., while they were in the fresh state) in order to make predictions. 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 O to pause and make predictions when fresh likely provided them with useful information about that state, predictors in Group O_{-1} did not make such predictions. Hence, they may have been less familiar with workers' sentiment in the fresh state (relative to Groups I and O). As mentioned above, we tried to mitigate this concern by briefly interrupting participants in Group O_{-1} 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.

Furthermore, Group O_{-1} provided an important degree of control that Groups O and I lacked. Namely, participants in groups O 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 O_{-1}—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 O to control for informational concerns and Group O_{-1} to control for order effects.

4 Results on Interpersonal Projection Bias

In this section, we present our main findings. 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 also use within-subject variation in predictions to show that participants erroneously decreased their predictions about fresh workers as they themselves became tired and provide estimates of the degree of interpersonal projection bias. We conclude with a few additional analyses that further support 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 versus 6.8 tasks per dollar when tired (difference significant at $p < .001$; Welch's two-sided t-test used throughout discussions of results despite our directional hypotheses).[14] This suggests our tiredness manipulation succeeded at generating a meaningful change in participants' attitude toward work.[15]

Since tired workers faced a higher monetary bonus than fresh workers, it is worth noting that this would mechanically reduce their WTW in tasks-per-dollar if their effort costs were convex—even if our tiredness manipulation had no effect. Simple back-of-the-envelope calculations help

[14]Our basic finding that there is a difference in WTW is a necessary step to identify projection bias. In a different context—gym attendance—März (2019; commenting on Acland and Levy, 2015) notes that a similar "first stage" effect is notably absent once the appropriate estimation technique is applied.

[15]We observe small differences in attrition across treatment groups. 55 workers failed to complete the experiment in the "fresh" treatment, while 88 failed to complete the experiment in the "tired" treatment (and 87 in the additional tired treatment in which they had to predict about themselves; see Section 5). Conditional on staying in the experiment for 60 seconds (and thus plausibly reaching the screen which announced the treatment assignment), 89.8% (81.1%) of workers assigned to the fresh (tired) treatment completed the experiment.

demonstrate that our observed reduction in WTW did not stem entirely from this convexity effect. Toward a contradiction, suppose that the tiredness state had no effect on WTW and the effort-cost function was merely $C(e;s) = \psi e^2$ for some scaling parameter ψ.[16] If a representative (fresh) participant with this cost function exerts the WTW reported in Table 1, then Equation 1 implies $\psi = \frac{2}{21.29^2} = 0.0044$. If we then use this estimate to calculate WTW for \$3, we would predict that tired workers would be willing to complete 26.11 tasks, or 8.70 tasks per dollar. However, tired workers in our experiment state an average WTW of 20.44 tasks for \$3; i.e., 6.81 tasks per dollar.[17] This difference suggests that even if workers faced substantially convex effort costs, our tiredness manipulation materially changed their WTW.

Table 1:
AVERAGE WILLINGNESS TO WORK

	Workers' State		
	Fresh (5 tasks; \$2)	Tired (20 tasks; \$3)	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 predictors' guesses about workers' WTW and evaluate each of our three enumerated hypotheses from Section 3.3. Throughout this subsection, we continue to normalize WTW (and predictions thereof) in terms of tasks per dollar. This is purely expositional; none of our

[16]For this exercise, we ignore idiosyncratic differences in the cost function and hence drop the θ_i term introduced in Section 3.3

[17]Following our setup in Section 3, a more realistic model might be that e additional tasks after having already completed s yields a cost of $C(e;s) = \psi(e+s)^\gamma - \psi(s)^\gamma$. Fitting this specification to match the average behavior of fresh and tired workers, we find $\psi = 0.01$ and $\gamma = 1.66$.

results rely on this normalization, and unnormalized predictions are presented in Appendix A.

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

As highlighted by Table 2, when predictors were 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 guessed 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 $p = 0.704$ top to bottom; Welch's two-sided t-test).

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

We now test whether fresh predictors overestimated the WTW of tired workers. As described in Section 3.3, we control for potential (state-independent) biases in predictions by comparing guesses of fresh predictors to those cast by tired predictors.

We present this information in Table 3. Note that only Group O provided predictions about tired workers while fresh, yet all three predictor groups did so while tired. Hence, the top-right cell of Table 3 shows $\widehat{W}^O(\$3, T|F)$, and the bottom-right cell shows $\widehat{W}^g(\$3, T|T)$ averaged over all members of groups $g \in \{I, O, O_{-1}\}$. (Below, we further discuss features of this data disaggregated across groups.) On average, fresh predictors guessed that tired workers had a WTW of 10.22 tasks per dollar, while tired predictors guessed 6.65. Thus, we find that guesses cast by fresh predictors were more than 50% higher than those cast by tired predictors (difference of 3.57 tasks per dollar or 10.71 total tasks; $p < .001$). Moreover, given that tired predictors accurately estimated the WTW of tired workers, this result implies that fresh predictors significantly overestimated the true WTW of tired workers (they estimated 10.22 while the truth, from Table 1, was 6.81).[18]

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

The "Fresh" column of Table 3 confirms this result. The top-left cell shows $\widehat{W}^I(\$2, F|F)$ for Group I (since only they made predictions about fresh workers while fresh) and the bottom-left cell shows $\widehat{W}^g(\$2, F|T)$ averaged over all members of groups $g \in \{I, O, O_{-1}\}$. On average, tired predictors guessed that fresh workers have a WTW of 9.46 (tasks per dollar), while fresh predictors guessed 10.81. Hence, tired predictors provide significantly lower estimates than fresh predictors (difference of 1.36 tasks per dollar or 2.72 total tasks; $p = .049$). This additionally implies that tired predictors underestimated the true WTW of fresh workers (they estimated 9.46 while the truth, from Table 1, was 10.64).

The "Difference" row of Table 3 shows the discrepancies in guesses cast by predictors in different states when holding the target group fixed. Notice that this difference is smaller for guesses about fresh workers than for guesses about tired workers (1.36 tasks per dollar versus 3.57), imply-

[18]The first row of Table 3 also highlights that predictors were aware of the different monetary incentives across groups. This is most apparent when considering predictions in terms of the absolute number of tasks rather than tasks per dollar. Fresh predictors guessed that the WTW of fresh workers for \$2 is 20.62 tasks and that the WTW of tired workers for \$3 is 30.66 tasks. Thus, predictors clearly expect that higher monetary payments will increase workers' WTW.

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: Only specific groups made predictions reported in the "Fresh" row (the top-left cell reports average predictions among Group I, and the top-right cell reports average predictions among Group O). The "Tired" row reports predictions averaged across all groups. Sample sizes are (clockwise from top-left): 223, 221, 666, 666. Standard errors are in parentheses. Decreased variance in tired row reflects the fact that all predictors made two guesses when tired. Differences significant at $p = .049$ and $p < 0.001$ (left to right; Welch's two-sided t-test).

ing 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 from Section 3.4: since fresh predictors faced uncertainty about how onerous the task would become, their guesses about tired workers may have been inaccurate not only due to projection bias, but also limited experience. By contrast, tired predictors had first-hand experience with the fresh state. Hence, their guesses about fresh workers involved less uncertainty, and the prediction error made by tired predictors therefore provides a cleaner measure of projection bias. As we discuss next, however, Table 3 understates the magnitude of this error because it aggregates data across predictor groups.

Analysis of Initial Predictions by Group

We now utilize the group-level variation in *initial* guesses. As we emphasize here, disaggregating the data down to the group level allows us to control for any potential order effects by exclusively comparing initial guesses across groups.[19]

To see why it may be important to control for order effects, we first consider Group O. This group first guessed about tired workers when they themselves were fresh (average guess of 10.22 tasks per dollar; see Table 3). They then guessed about fresh workers when they themselves were tired. The average of this second guess was relatively high (11.24 tasks per dollar; see Table A1 in Appendix A), and accordingly, did not exhibit the underestimation we'd expect under Hypothesis 3 since the true WTW of fresh workers was 10.64 (see Table 1). We suspect that these elevated second predictions stemmed from order effects, e.g., anchoring or consistency bias. Namely, since their first guesses were very high—perhaps due to projection—their subsequent guesses may have been shifted upward as well. This would partially obfuscate our ability to detect projection amongst Group O's second predictions.

Fortunately, our experiment was specifically designed to allow us to sidestep such order effects by analyzing only the first predictions cast by each group. Indeed, predictors in Group O_{-1}, who first guessed about fresh workers when they themselves were tired, did significantly underestimate

[19] See Table A1 in Appendix A for all disaggregated predictions.

this WTW consistent with Hypothesis 3. While predictors in Group I made accurate first guesses (as shown in Table 2), Table 4 shows that the first guesses among predictors in both Out Groups were systematically biased. Group O's first guess—about tired workers when they themselves were fresh—was far too high on average (average guess of 10.22 versus a truth of 6.81), while Group O_{-1}'s first guess—about fresh workers when they themselves were tired—was too low (average guess of 8.44 versus a truth of 10.64).[20] Furthermore, when focusing only on first guesses, the projection error among tired predictors guessing about fresh workers is now more pronounced than it was in Table 3 (2.20 tasks per dollar in Table 4 versus 1.36 in Table 3). The previous estimate in Table 3 understates this degree of projection because it averaged over all groups and thus included the second guesses of Group O, which were potentially shifted upwards by order effects.[21]

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

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

Notes: Predictions about tired workers (top row) are from Group O. Predictions about fresh workers (bottom row) are from Group O_{-1}. Standard errors are in parentheses. Differences significant at $p < .001$ and $p = .012$ (top to bottom; Welch's two-sided t-test).

In addition to being free of any potential order effects, the initial guesses made by Group O_{-1} also involve a diminished role of uncertainty, as noted in our discussion of Hypothesis 3: tired

[20]Note that Figure 1 in the introduction depicts the the distribution of responses underlying Table 4.

[21]Group I was similar to Group O_{-1} in that they also underestimated the WTW of fresh workers when they themselves were tired. On average, they guessed 8.7 tasks per dollar, while the truth was 10.64 (see Table A1).

predictors guessing about fresh workers did not face the same degree of uncertainty that fresh predictors faced when guessing about tired workers. Thus, we believe the prediction error in the initial guesses from Group O_{-1} reflects our cleanest measure of the effect on projection bias on predictions. The magnitude of this error is 2.2 tasks per dollar, representing an underestimate of approximately 21% relative to the truth. This measure also allows us to loosely approximate the effect that uncertainty had on Group O's initial guesses. For this exercise, we assume that projection bias similarly led Group O to overestimate the WTW of tired workers by 21%. Since this WTW is 6.8 tasks per dollar in truth, projection would then induce a prediction error of 1.4 tasks per dollar. But the observed prediction error is 3.4 (or 50% of the true WTW). The additional portion of this observed error, $3.4 - 1.4 = 2$ tasks per dollar (or 29% of the true WTW), could then be a result of uncertainty about the task. In the next section, we quantify the degree of interpersonal projection bias using alternative approaches and find similar magnitudes.

4.3 Within-Subject Measures of Interpersonal Projection Bias

We now consider some within-subject measures of projection bias, leveraging the fact that each predictor made several guesses over time. This section proceeds as follows. We first examine how a predictor's guesses about a fixed target group changed as the predictor moved from fresh to tired. We then estimate a simple, reduced-form variant of our model, allowing us to additionally account for how a predictor's guesses depended on their own stated WTW. We utilize both approaches to quantify the overall bias.

We begin by noting a simple fact: predictors in Group I changed their guesses about fresh workers after they completed additional tasks and became tired. In our opinion, this represents one of our strongest indicators of projection bias. Note that Group I's first guess about fresh workers was made while they themselves were fresh. They therefore had the same information—and tiredness—as the workers did when stating their WTW. Thus, additional exposure to the task should not have led these predictors to change their guesses about fresh workers. Despite this, we find that Group I predictors significantly lowered their guesses about fresh workers once they themselves became tired: $\widehat{W}^I(\$2, F|F) - \widehat{W}^I(\$2, F|T) = 4.22$ total tasks ($s.e. = 0.858$, $p < .001$ for difference). Although the predictors' initial guesses about fresh workers were well calibrated (see Table 2), their revised guesses were approximately 19% lower than their first, and they ultimately underestimated the true WTW of fresh workers by about 21%. As we show below, these revisions significantly reduced the expected earnings for Group I.

We likewise note that predictors in Group O changed their guesses about tired workers once they became tired. Recall that Group O first guessed about tired workers while fresh. At that time, they potentially lacked information about how it felt to be tired. We may therefore expect a relatively large revision in their predictions, stemming from both projection and the resolution of this uncertainty. Indeed, the average revision in guesses about tired workers was $\widehat{W}^O(\$3, T|F) - \widehat{W}^O(\$3, T|T) = 8.02$ total tasks ($s.e. = 1.018$, $p < .001$ for difference). On average, their revised guesses were approximately 26% lower than their first.

Recall that we also collected predictors' own (hypothetical) WTW while tired. As a final approach to measuring projection bias, we incorporate this data to estimate a simple reduced-form model motivated by our theoretical framework in Section 3.3. Our specification in that section (and other common models of projection bias) assume that the parameter capturing projection bias, α, enters as a convex combination of utilities across states (see Equation 2). However, since we instead observe willingness to work, we estimate a parameter that captures a convex combination of the optimal WTW across states. More specifically, we assume a projector's prediction about a worker's WTW is a convex combination of her own WTW in her current state and a her unbiased estimate of a worker's WTW in the target state. As above, 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 WTW of a worker in state s facing payment m is then

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

where $\mathbb{E}_\theta[\cdot|\theta_i]$ denotes Predictor i's expectation over θ conditional on herself having type θ_i, and

s_i is her state when casting this prediction. Parameter ρ therefore measures the extent to which the predictor's estimate is biased toward her own current WTW.

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 (3), we can write these two predictions as

$$\widehat{W}(\$2, F|T, \theta_i) = \rho W(\$2, T|\theta_i) + (1-\rho)\mathbb{E}_\theta[W(\$2, F|\theta)|\theta_i], \tag{4}$$

and

$$\widehat{W}(\$3, T|T, \theta_i) = \rho W(\$3, T|\theta_i) + (1-\rho)\mathbb{E}_\theta[W(\$3, T|\theta)|\theta_i]. \tag{5}$$

Recall that we elicited predictors' own (hypothetical) WTW for \$3 when tired but not for \$2; thus we measure $W(\$3, T|\theta_i)$ but not $W(\$2, T|\theta_i)$. In order to obtain a conservative estimate of ρ with this limited data, we impose a linear effort-cost function, implying that $W(\$2, T|\theta_i) = \frac{2}{3}W(\$3, T|\theta_i)$. Substituting this value into Equation (4) and then differencing Equations (4) and (5) yields

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

We estimate this—given the data we observe—with the following econometric model:

$$\text{Diff}_i = \beta_0 + \beta_1\left(\frac{1}{3}\text{WTW}_i\right) + \epsilon_i, \tag{6}$$

where Diff_i is the observed difference in Predictor i's guesses cast while tired, WTW_i is her own willingness to work, and β_1 corresponds to ρ. When pooling all predictors and estimating via OLS, this estimation yields $\rho = 0.23$ ($s.e. = 0.061$).[22] However—as suggested above—treating costs as linear yields a conservative estimate of ρ.[23] Thus, although our design was not optimized for

[22] Allowing the intercept to vary for each of the three groups (I, O, and O_{-1}) leads us to estimate $\rho = 0.22$ ($s.e. = 0.063$) and thus does not substantively alter the results.

[23] To illustrate this claim, suppose each tired Predictor i had a normalized cost of additional effort given by $\theta_i(e)^\gamma$. Our observations $W(\$3, T|\theta_i)$ and $W(\$2, T|\theta_i)$ would solve $\theta_i W(\$3, T|\theta_i)^\gamma = 3$ and $\theta_i W(\$2, T|\theta_i)^\gamma = 2$, respectively. It then follows that $W(\$2, T|\theta_i) = \left(\frac{2}{3}\right)^{1/\gamma} W(\$3, T|\theta_i)$, and the regressor in Equation (6)—which captures

this parametric approach, it still provides estimates of projection bias that are consistent with our non-parametric results.

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 the 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 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 became slightly more accurate with time, on average. However, this improvement was 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, $p = 0.032$).

Furthermore, Group I helps show that any improvements in accuracy stemming from experience must have been relatively small compared to the reduction in accuracy associated with being in a different state than the target group. Recall that Group I's first two guesses were about workers in their same state (fresh and tired, respectively). Table 5 suggests that Group-I predictors' second guess was more accurate than their first, and hence they may have improved in their ability to estimate others in their same state. Group I's third guess, however, was about fresh workers

$W(\$3, T|\theta_i) - W(\$2, T|\theta_i)$—would then be $\left(1 - \left(\frac{2}{3}\right)^{1/\gamma}\right) W(\$3, T|\theta_i)$. Our estimation assumes $\gamma = 1$. As γ increases above 1, the multiplier term $\left(1 - \left(\frac{2}{3}\right)^{1/\gamma}\right)$ decreases; this mechanically increases the estimate of β_1 (and hence ρ). For instance, if we assume costs are quadratic instead of linear (i.e., $\gamma = 2$), then we would estimate $\rho = 0.41$ ($s.e. = 0.111$).

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

	Group I	Group O	Group O_{-1}
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.

while they themselves were tired, and was less accurate than either of their initial guesses. Thus, experience with guessing about those in their same state was not sufficiently beneficial to overcome the reduction in accuracy stemming from projection bias (i.e., misalignment between predictors' and workers' states). Moreover, while the reduction in Group I's accuracy between their first and third guess does not appear significant in Table 5, it indeed 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% ($p = .002$ for difference).

Overall, we believe that these limited improvements in accuracy—along with our results on confidence, below—suggest that there may have been some learning from experience, 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 **??** in Online Appendix A. We first consider the confidence of Groups I and O, as these groups made predictions while both fresh and tired. As shown in Table **??**, average confidence did not increase with experience—neither in going from the first prediction to the second (and thus accumulating more experience with the task) nor 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 explore whether this correlation indeed stems from projection, we pooled predictors from Groups I and O, and we split them into two new 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.[24] 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 ($p = 0.032$ for difference). 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 post, we believed that those who took longer to complete the tasks might be more fatigued. 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 Online Appendix **??** which demonstrate that task completion time has no effect on participant predictions or their self-reported confidence. We suspect this null result may stem from the fact that much of the heterogeneity in task-completion times arose due to inattention (e.g.,

[24]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.

doing other things online), but we have no direct evidence of this.

5 Comparison to Intrapersonal Projection Bias

*Intra*personal projection bias—the propensity for one's current state to overly influence predictions about their *own* behavior in a different state—is well-documented (see below for examples and discussion). To provide evidence for this in our 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 once 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.[25] 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 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 the predicted and actual 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. Perhaps more dramatically, 93 out of 298 participants overestimated their WTW in a costly way: their prediction was more than 5 tasks higher than their true WTW, which prevented them from earning the bonus.[26]

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.

[25] These predictions were incentivized in the same way as other predictions in this experiment: participants earned the bonus if their prediction of their own WTW was within 5 of their subsequent stated WTW.

[26] Figure **??** in Online Appendix A shows the distribution of individual differences in predictions versus actual WTW. The distribution is skewed toward positive values. As noted, 93 subjects out of 298 overestimated their subsequent WTW to a degree that reduced their payoffs, while a substantially smaller fraction (23 out of 298) underestimated their WTW in a similarly costly way.

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).

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 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.[27]

To assess the relative magnitudes of intra- and interpersonal projection bias, we compare prediction errors of the predicting workers with those of 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 10.25 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 better calibrated when making predictions about themselves rather than about others.

This finding suggests that the intrapersonal prediction error is substantially smaller than the interpersonal one. However, this interpretation comes with some critical caveats. The difference

[27]Comparing Table 6 with Table 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 motivational reference point (e.g., Heath, Larrick, and Wu, 1999). However, our experiment is not well-suited to draw such conclusions.

in the prediction errors noted above (30% vs 50%) captures the difference in the degree of intra- and interpersonal projection bias only if we assume that uncertainty about how onerous the task would become was similar when considering oneself and considering others.[28] Importantly, 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 similar in magnitude. If the 30% intrapersonal error we document here stems from the same composition of uncertainty and projection as the interpersonal error, then we would conclude that interpersonal projection is stronger. This simple arithmetic allows us to consider alternative assumptions on the relative composition of these effects. For instance, it is likely that participants have less uncertainty about themselves than about others. If we observed a small difference in uncertainty across self and other, we would still conclude that interpersonal projection bias is stronger than intrapersonal. But if this difference was sufficiently large (e.g., in the extreme case where participants faced *no* uncertainty about themselves) then we would instead conclude that interpersonal projection is slightly weaker than intrapersonal projection. As an additional caveat, any comparison across the intra- and interpersonal domains should be interpreted with care due to the potential for different measurement errors. Since the prediction error in the intrapersonal case is smaller, any measurement error would loom larger.

That said, it is intuitive that the interpersonal prediction error may be larger than the intrapersonal one. In the interpersonal case, there are more factors that might be projected: people could, for instance, project their intrinsic taste for the task, captured by θ_i. By contrast, in the intrapersonal case, there is limited scope for one's stable, intrinsic taste to distort their predictions. Moreover, past work on perspective taking (see, e.g., Van Boven and Loewenstein, 2005) suggests that interpersonal predictions across states entail two distinct judgements: (i) what one's own choice would be in an alternative state, and (ii) how similar another person's choice would be to their own. Both steps in this dual-judgment process are prone to mistakes. Table 6 reveals an error in the first step (i.e., the intrapersonal one), and thus interpersonal prediction errors may be larger since they

[28] An extensive psychology literature suggests that such a symmetry is likely (see Van Boven et al., 2013 for a review).

additionally reflect any error in the second step.

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 structurally estimate the extent of intrapersonal projection. Conlin, O'Donoghue and Vogelsang (2007) find $\alpha \in [0.31, 0.50]$ for cold-weather clothing-catalog sales, while Augenblick and Rabin (2019) find $\alpha \in [0.27, 0.53]$ in a real-effort experiment.[29,30] Our measures thus accord with an emerging consensus on the magnitude of projection bias observed across a variety of domains.

6 Conclusion

In this paper, we provide evidence that interpersonal projection bias leads to substantial and costly errors in forecasting others' behavior. Specifically, we find that predictors correctly guessed the behavior of others in their own state, but fresh predictors systematically guessed that tired workers would behave as if they too were fresh, and tired predictors guessed that fresh workers would behave as if tired. Additionally, we find evidence for intrapersonal projection in the same domain, and this error is likely smaller in magnitude than interpersonal projection. Our evidence suggests that neither uncertainty about the task nor learning were the root cause of these errors. Rather, our

[29] 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 function. Moreover, their experiment also examines present bias, and their ability to separately measure projection bias is somewhat limited by their design.

[30] Although there are a number of related studies, many—particularly early experimental studies—are not suited to estimate the degree of projection. Likewise, some recent empirical papers do not parametrically estimate projection bias directly, but find support for its main premise (though other possible explanations are also offered, such as salience and over-weighting situational cues). 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.

results accord with a simple model of interpersonal projection bias.

Our results also suggest that projection bias may be one potential source of inaccuracy documented among forecasters attempting to predict others' response to incentives (e.g., DellaVigna and Pope 2018a, 2018b, 2022). To illustrate how such forecasts may go awry, consider a projecting manager who is forecasting her workers' behavior. Such a manager will underestimate how workers' marginal disutility of effort differs across states and fail to anticipate how the motivating effect of incentives changes over time. We find suggestive evidence of this misperceived gap in responsiveness to incentives. As reported in Table 1, the true difference in the WTW between fresh and tired workers was 3.83 tasks per dollar. Tired predictors, for instance, underestimated this difference by 27% (see Table 3). Future work could be tailored to more directly test how these perceived differences depend on a predictor's current state.[31]

There are several other avenues for further research. Our evidence shows that people project "tiredness", which is in line with the literature on projection of visceral states (see, e.g., Van Boven and Loewestein, 2003 for a discussion on this point). However, it may be that the degree of projection decreases in contexts where differences in preferences are driven by less visceral factors. Nevertheless, other barriers to perspective taking may still prevent accurate forecasts.[32] Better understanding when these various barriers emerge and how to mitigate them may greatly improve how we predict the behavior of others who face differing perspectives or circumstances.

Additionally, we have primarily modeled and interpreted our results as stemming from the projection of preference-relevant states. We've adopted this interpretation following the large psychology and economics literature that does the same (see, e.g., Van Boven et al., 2013; Loewenstein, O'Donoghue, and Rabin, 2003). However, other forms of projection—namely information projection (as in, e.g., Camerer, Loewenstein, and Weber 1989; Madarász, 2012; Danz, Madarász, and Wang, 2018)—could also play a role. In Madarász's (2012) model of information projection, an

[31]Our experiment was not designed to precisely estimate this difference-in-differences across groups. However, the main hypotheses that we test and confirm (Section 4.2) imply this result.

[32]Epley and Caruso (2009) describe three forms of barriers: (i) failing to even think about differing perspectives; (ii) failing to fully adjust away from one's own current perspective; and (iii) holding an inaccurate sense of what perspectives others may hold. Our evidence is most in line with the second barrier, yet the first may be more prominent when salient visceral or emotional factors are absent.

individual wrongly acts as if others might know his private information. In our setting, information projection may be relevant for tired predictors if they gained information about the onerousness of the task when doing extra work. For example, as predictors worked more (that is, became tired), they may have come to learn the task was more onerous than expected and projected this new information onto others. As such, tired predictors may have underestimated fresh workers' WTW not only because they wrongly treated them as more tired than they were (as in our interpretation), but also because they wrongly thought these fresh workers had additional information about the task that they didn't yet have. Such information projection may be a partial explanation for our effect, but given that mispredictions about others seem to be significantly more biased than participants' mispredictions about their own future attitudes toward the task, we suspect there is scope for both forms of projection.[33] Furthermore, information projection has difficulty explaining the systematic pattern of mispredictions made by fresh predictors because they don't have additional information to project. But it is inherently difficult to fully distinguish between state-based preference projection and information projection when the information at hand concerns a person's marginal utility. Future work could further illuminate when and how these two specific forms of interpersonal projection differentially distort predictions in order to, for instance, inform the design of strategies to debias forecasters. If an intervention was limited to highlighting either others' states *or* their information, the optimal approach would depend on the relative magnitude of these effects.

References

ACLAND, D. AND M. LEVY (2015): "Naiveté, Projection Bias, and Habit Formation in Gym Attendance," *Management Science*, 61(1), 146–160.

[33] In our setting, the magnitude of bias stemming from information projection depends on the degree to which fresh workers were initially miscalibrated about the eventual onerousness of the task. If, for example, this miscalibration was relatively small, then gaining access to the knowledge of a tired individual should have little effect on a fresh worker's WTW. We can use the self-predictions data (reported in Table 6) to get a sense of any such miscalibration in our setting: fresh workers guessed that their own WTW would be 22.11 tasks when tired, while the actual WTW among tired workers was 20.44 (Table 1). Thus, miscalibration among fresh workers appears small relative to the magnitude of prediction errors we observe. Recall that fresh predictors guessed that the WTW of tired workers was 30.65—considerably higher than guesses about their own WTW when tired.

AMBUEHL, S., D. BERNHEIM, AND A. OCKENFELS (2021): "What Motivates Paternalism? An Experimental Study," *The American Economic Review*, 111(3), 787–830.

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., W. BICKEL, L. GIORDANO, E. JACOBS, G. LOEWENSTEIN, 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.

BENJAMIN, D. (2019): "Errors in Probabilistic Reasoning and Judgment Biases," in *Handbook of Behavioral Economics: Applications and Foundations*, Elsevier, 2, 69–186.

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

BOHREN, A. AND D. HAUSER (2020): "Learning with Model Misspecification: Characterization and Robustness," *Econometrica*, 89, 3025–3077.

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

CAMERER C., G. LOEWENSTEIN, 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.

CHARNESS, G. AND M. DUFWENBERG (2006): "Promises and Partnership," *Econometrica*, 74: 1579-1601

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.

DANZ, D., K. MADARÁSZ, AND S. WANG (2018): "The Biases of Others: Projection Equilibrium in an Agency Setting," *Working paper*.

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.

DELLAVIGNA S. AND D. POPE (2018a): “Predicting Experimental Results: Who Knows What?” *Journal of Political Economy*, 126(6), 2410–2456.

DELLAVIGNA S. AND D. POPE (2018b): “What Motivates Effort? Evidence and Expert Forecasts,” *Review of Economic Studies*, 85(2), 1029–1069.

DELLAVIGNA, S. AND D. POPE (2022): “Stability of Experimental Results: Forecasts and Evidence,” *American Economic Journal: Microeconomics*, 14(3), 889-925.

DELLAVIGNA S., D. POPE AND E. VIVALT (2019): “Predict Science to Improve Science,” SCIENCE, 366(6464), 428–429.

ELLINGSEN, T., M. JOHANNESSON, 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.

EPLEY, N., AND E. CARUSO (2009): “Perspective Taking: Misstepping into Others’ Shoes,” in K. D. Markman, W. M. P. Klein, and J. A. Suhr (Eds.), *Handbook of Imagination and Mental Simulation*, 295–309. Psychology Press.

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*, 88, 2281–2328.

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

GAGNON-BARTSCH, T., M. PAGNOZZI, AND A. ROSATO (2021): “Projection of Private Values in Auctions,” *American Economic Review*, 111, 3256–3298..

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

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

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

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

KAUFMANN, M. (2022): “Projection Bias in Effort Choices,” *Games and Economic Behavior*, 135, 368–393.

KRUEGER, J. AND R. CLEMENT (1994): “The Truly False Consensus Effect – An Ineradicable and Egocentric Bias in Social-Perception,” *Journal of Personality and Social Psychology*, 67(4), 596–610.

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.

MÄRZ, O. (2019): “Comment on “Naiveté, Projection Bias, and Habit Formation in Gym Attendance,” *Management Science*, 65(5): 2442–2443.

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

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.

SELTEN, R. AND A. OCKENFELS (1998): “An Experimental Solidarity Game,” *Journal of Economic Behavior and Organization*, 34(4), 517–539.

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

THOMAS, C., N. OTIS, J. ABRAHAM, H. MARKUS, AND G. WALTON (2020): “Toward a Science of Delivering Aid with Dignity: Experimental Evidence and Local Forecasts From Kenya,” *Proceedings of the National Academy of Science*, 117(27): 15546–15553.

VANBERG, C. (2008): “Why Do People Keep Their Promises? An Experimental Test of Two Explanations,” *Econometrica*, 76: 1467–1480.

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. AND G. LOEWENSTEIN (2005): “Empathy Gaps in Emotional Perspective Taking,” in B. F. Malle and S. D. Hodges (Eds.), *Other Minds: How Humans Bridge the Divide Between Self and Others*, 284–297. New York: Guilford Press.

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.

ZIEGLER, A., G. ROMAGNOLI, AND T. OFFERMAN (2022): “Morals in Multi-Unit Markets,” *Working paper*.

Appendix

A All Predictions by Group and Timing

Recall that the true WTW of fresh workers was 21.3 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 O		Out Group O_{-1}	
	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.