Working Hard by Hardly Working?

Self-Other Asymmetries in Perceptions of Hard Work

Abstract

Society values hard work, but what that constitutes is ambiguous: It can entail toiling for much time or with great intensity. Three studies show that people (undergraduates, American adults) define hard work differently when evaluating their own vs. others' efforts. Studies 1a/1b used a yoked design in which workers supplied information about a recently completed project. In judging how hard workers worked, workers leaned relatively more on their work intensity, whereas observers put more stock in time worked. Study 2 replicated this pattern in recollections of participants' own vs. others' prototypical hard-work experiences. In Study 3, workers completed two in-lab work tasks—that required one to work for longer or with more intensity. Compared to observers, workers characterized the intense but shorter task as requiring harder work. Defying observer expectations, workers chose to repeat that "harder" task. Workers thus define hard work in a way that matches their work-style preferences.

Keywords: self-other differences, procrastination, self-enhancement, motivated reasoning

Statement of Relevance

Social and clinical psychologists have long understood people's fundamental desire to think well of themselves. Assisting with this goal, people define ambiguous traits and abilities in ways that play to their own strengths. We move to a qualitatively different domain—characterizations of work—and find that people use qualitatively different criteria in judging themselves versus other. Whether because people are simply optimistic about how much time they will devote to a project or because people simply prefer shorter, intense bouts of work (we find evidence for both), the self can see itself as a hard worker even as it judges (equivalently working) others as falling short. This double standard is not only a new finding in the study of self-enhancement, but should be of interest to organizational psychologists with interests in how work is structured and evaluated. Managers may mis-assess workers' work-style preferences and be hypocritically predisposed toward perceiving indolence.

Working Hard by Hardly Working?

Self-Other Asymmetries in Perceptions of Hard Work

A strong worth ethic is a core value in many countries (Furnham et al., 1993). Central to this ethos is a valuation of mastery and self-control (Furhnham, 1984; McClelland, 1961). In Weber's (1905) seminal work on the Protestant Work Ethic, he noted Benjamin Franklin both described and extolled the local norm to "be prudent, diligent…do not be idle, for time is money." Of course, social values cannot always be lived up to. Working without idling is often an ideal that is not easily attained. For example, 80% to 95% of college students tend to procrastinate, a self-regulatory shortcoming that continues into adulthood (Steel, 2007). People begin projects with the best of intentions to put in steady efforts only to delay, and delay more (with a nagging awareness that they are worse off for it), thereby finding themselves rushing to complete tasks last minute. Such procrastination is associated with decreased well-being and poorer mental health (Rozental & Carlbring, 2014; Stead, Shanahan, & Neufeld, 2010; Tice & Baumeister, 1997).

How do people resolve this tension between societal prescriptions and their actual behavior? Do they see themselves as procrastinators who fail to live up to society's ideals to be hard workers, or as hard workers who are in denial about their degree of procrastination? We conducted an exploratory pretest (N = 197 American undergraduates, 6 failed an attention check; see Supplemental Materials). Our participants thought they worked harder than the average student in their class (M=60th percentile, SD=20.30), $t(190)$=6.61, $p<.001$, d=0.48. Curiously, they also admitted to symptoms of procrastinating. They thought they were more likely than others to be optimistic about how much time they would devote to work, thereby finding themselves putting in intense last-minute efforts to meet deadlines (M=59th percentile,

SD=19.97), *t*(190)=6.33, *p*<.001, *d*=0.46. They also saw themselves as less likely than others to devote less-frantic, long and careful attention to projects (*M*=44th percentile, *SD*=19.53), *t*(190)=-4.23, *p*<.001, *d*=-0.31. In other words, people see themselves as above-average workers, but ones who are spending less time working on projects than their peers. How can we explain this seeming contradiction?

One possibility is simply that people see themselves as having more tasks to complete than others—explaining why they are hard workers who never have as much time as others to complete any specific task. After all, modern life is busy, and people may be egocentric in failing to appreciate that they are not unique in confronting this challenge. But in this paper, we focus on an additional possibility: We hypothesize people display a self-other asymmetry in how they define hard work, one that turns their own shortcomings (lack of sustained, prolonged effort) into a strength (ability to work intensely, whether by choice or necessity). Although past research on work ethic (Blau & Ryan, 1997, Rau & Durand, 2000), achievement motivation (Helmreich & Spence, 1978), learning (Naquin & Holton III, 2002), performance (Rynes, Gerhart, & Parks, 2005), flow (Csikszentmihalyi, 2014), and workaholism (Schaufeli, Shimazu, & Taris, 2009) have examined the contributors to and consequences of *working hard*, a read of those literatures also make clear the multidimensionality and thus fundamental ambiguity of the construct's definition. This can even be seen in a simple dictionary definition of Hard Work (n.d.) as "a great deal of effort or endurance." This emphasizes both qualitative *intensity* (effort) and the actual *time spent* (endurance). This split is similarly observed in the literature. Working hard—defined as engrossing intensity—can induce feelings of flow (Csikszentmihalyi, 2014).Working hard—instead defined as working a large number of hours—predicts sobriety and better grades (Rau & Durand, 2000).

Although previous research has found people define traits in self-serving ways (see Critcher, Helzer, & Dunning, 2011, for a review), we go one step further and suggest people define hard work differently for themselves and for others. Self-judgments frequently look inward, prioritizing private internal experience (Pronin, 2008). For example, the self underestimates how soon it will complete tasks in part because it sets aside its (observable) performance history and focuses on its own lofty intentions (Buehler, Griffin, & Ross, 1994). Social judgments prioritize these attributes in reverse (Helzer & Dunning, 2012). The self believes it is best reflected not in how others see it today, but in its potential that only the self truly appreciates (Williams, Gilovich, & Dunning, 2012). In contrast, social judgments can be less charitable because they prioritize what they observe others doing, not what they intend (Kruger & Gilovich, 2004). In an example closest to our focus, people judge the quality of others' work by how long they worked on it (Kruger, Wirtz, Van Boven, & Altermatt, 2004).

This research tradition makes plausible that work intensity and time worked—as an internal, more private experience and an externally observable behavioral feature—may play a relatively stronger role in self and social conceptions of hard work, respectively. This double standard may help to paint the self in a positive light: One who dawdles for most of the race only to have to sprint at the end may wish to focus on the strength displayed in that mad dash.

Study 1a and 1b

Studies 1a and 1b used a yoked design to test whether self and social perceptions of hard work put relatively more weight on qualitative work intensity or time spent working, respectively. Workers supplied descriptions of a recent work project as well as information about how much they worked on it as well as how intensely they worked while actually working. Both workers and observers provided judgments of how hard workers worked. We expected that

observers' and workers' ratings would prioritize workers' time worked and intensity of that work, respectively.

Method

Sample-size determination. Given the novelty of the research question, we did not have *a priori* knowledge of the sample size we would need to achieve good statistical power. This is of course a common problem in research, and it led us to focus on achieving as large of sample sizes as resources permitted. When a study was run in the lab, at a minimum we recruited as many participants as we could in a single academic semester subject to the size of the subject pool. We deviate from this rule only when running multi-wave studies, which always drew on the same underlying subject pool but were run across multiple semesters. When a study was run on Amazon Mechanical Turk (AMT), we took the funding lab's monthly budget and divided it among the studies run that month on AMT. This allowed us to far exceed Simmons, Nelson, and Simonsohn's (2013) rough minimum threshold (of 50 participants per condition) below which they argue sample sizes require additional justification.

Participants and design. Participants (N = 650) were undergraduates drawn from a subject pool at an American university. The first waves of subjects ($n_{1a} = 144$, $n_{1b} = 215$) were *workers* who reported on their own experience. The second waves of subjects ($n_{1a} = 83$, $n_{1b} = 208$) were *observers* who judged workers' efforts. What differentiated whether a participant was in the first or second wave was whether participants completed the study before or after we added an exploratory measure that we supplied to some yoked observers (about workers' degree of stress about their project) that had no effect on the effects reported below (see Supplemental Materials for more details). At that point, we also increased the number of workers that each

observer rated from four to seven. The primary result described below was statistically

significant in each sample alone.

Procedure. Workers were asked to identify "a recent, time-consuming school assignment

that you worked on...over the course of one or more days." Such work tasks were supposed to

satisfy two criteria. First, workers were to have had at least two weeks to work on the project.

Second, the task was to be one that was assigned on a specific date and due on a specific,

already-passed date. We noted that a term paper, a class presentation, or (preparation for) an

exam all had the potential to satisfy these criteria.

Workers began by describing the assignment. Informed that their descriptions might be

shared with others (i.e., observers), workers were encouraged to write with sufficient clarity that

others would understand the nature of the work task. Workers provided several pieces of

information about their work. First, workers supplied the exact calendar dates of both when the

task was assigned as well as when it was due. From this we could calculate the *assignment length*

in days ($M = 46.31$, $SD = 44.13$). Second, workers indicated how many distinct days they spent

working on the project as well as the average number of hours they spent working on those dates.

From these two questions, we could calculate *time worked* in total hours ($M = 25.86$, SD $=$

64.10). Third, workers indicated their *work intensity* by responding to a single question: "When

you *were* working on the project, how intensely were you working?" Participants responded on a

9-point Likert-type scale ($M = 6.68$, $SD = 1.50$) anchored at 1 (*not at all intensely)* and 9

(*extremely intensely*).

Observers saw the prompts that workers responded to as well as their actual responses.

Each observer considered multiple workers: four in the first wave and seven in the second.

Crucially, workers did not see observers' responses to one additional *worked hard* question that

workers answered: "Overall, how hard would you say you worked on this task/project?" from

1(*not very hard*) to 9(*very hard*). Observers answered this same question, but about the yoked

worker instead of themselves. Unlike workers—who answered the worked hard item either just

after providing the task description or at the very end—observers always completed this item

last. Because the counterbalancing did not influence results, it will not be discussed further.

Results

     To examine differences in how workers and observers define hard work, we used a mixed

model predicting the *worked hard* ratings. The predictor variables were the participant's role (+1:

worker, -1: observer), assignment length (log-transformed and standardized), time worked (log-

transformed and standardized), and work intensity (standardized). We defined two interactions

relevant to our hypotheses: Role × Time Worked and Role × Work Intensity. In our first model,

we included the sum of the two interaction terms as well as the difference of the two interaction

terms. The difference term would allow us to test in a single model whether workers and

observers showed a different relative reliance on time worked vs. work intensity in determining

whether the worker had worked hard. In our second model, we instead merely included the two

interaction terms. In both models, we included random intercepts for the participant (given

observers rated several projects) and for the project described (given workers and observers rated

the same set of projects). Full output are presented in Table 1.

     In the first model, we observed a main effect of role, $b = 0.36$, $SE = 0.05$, $t(1033.00) =$

7.24, $p < .001$. Workers perceived themselves to have worked harder ($M_{adjusted} = 6.63$, $SE = 0.08$)

on their projects than social observers judged ($M_{adjusted} = 5.92$, $SE = 0.06$). This reflects self-

enhancement (or other-derogation). But crucially the difference score term (Role × Work

Intensity - Role × Time Worked) was significant, $b = 0.11$, $SE = 0.04$, $t(1753.78) = 3.01$, $p = .003$

Table 1

Regression Output from Two Primary Models Predicting Worked Hard Ratings (Studies 1a/1b)

	Model 1		Model 2	
Predictors	b (SE)	t	b (SE)	t
Intercept	6.27 (0.05)	120.66***	6.27 (0.05)	120.66***
Role	0.36 (0.05)	7.24***	0.36 (0.05)	7.24***
Assignment Length	-0.28 (0.04)	-7.56***	-0.28 (0.04)	-7.56***
Work Intensity	0.53 (0.05)	10.78***	0.53 (0.05)	10.78***
Time Worked	0.58 (0.05)	11.78***	0.58 (0.05)	11.78***
Int1 – Int2	0.11 (0.04)	3.01**		
Int1 + Int2	-0.00 (0.03)	-0.16		
Int1: Role ✕ Work Intensity			0.10 (0.05)	2.24*
Int2: Role ✕ Time Worked			-0.11 (0.04)	-2.48*

*Note.* Role (+1: worker, -1 = observer). Assignment Length and Time Worked were log-transformed and then standardized. Work Intensity was standardized. ***$p < .001$

(see Figure 1). This reflects an omnibus test confirming our central hypothesis: Observers—compared to workers—defined hard work by putting relatively more weight on workers' time worked (instead of their work intensity).

To more formally unpack this test, we proceeded to our second model. First, we observed a Role × Time Worked interaction, b = -0.11, SE = 0.04, t(1763.30) = -2.48, p = .013. In determining whether workers worked hard, observers gave more weight (b = 0.69, SE = 0.04) than workers (b = 0.47, SE = 0.08) to the amount of time that workers put in. We also observed an oppositely signed Role × Work Intensity interaction, b = 0.10, SE = 0.05, t(1756.38) = 2.24, p = .025. This reflected that workers (b = 0.63, SE = 0.09) gave more weight to their work intensity than did observers (b = 0.42, SE = 0.04). In short, perceptions of the intensity of work and the amount of time worked were relatively more important to workers' and observers' perceptions, respectively, of what it meant to work hard.

Figure 1

The Independent Predictive Strength of Work Intensity and Time Worked on Perceptions of Hard

Work, by Role (Studies 1a/1b)
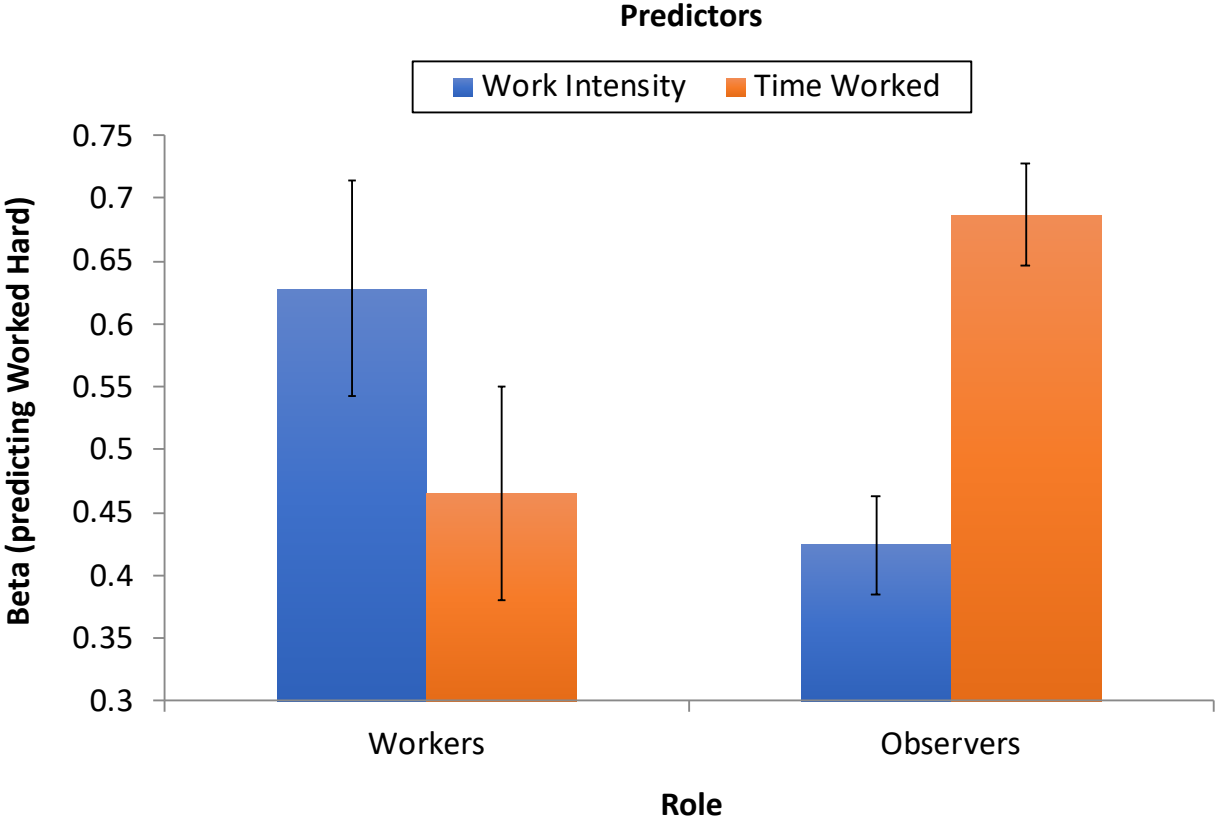
Predictors

Note. Each bar reflects the beta associated with the independent effect of work intensity or time

worked on perceptions that the worker worked hard, controlling for the other predictor.

Observers were yoked to workers, meaning that the betas reflect contributors to perceptions of

the same set of work projects. Error bars are ± 1 SE.

Study 2

Study 1a-1b provided initial evidence that self and social definitions of hard work differ.

One strength of the design—that workers and observers characterized the exact same work

episodes—also introduced a potential problem: Workers themselves supplied all information about which the judgments were made. Study 2 provided a complementary test of our account. Participants identified recent examples in which they themselves or someone they knew worked especially hard. We expected that these prototypical hard-work examples of the self or an other would prioritize work intensity or time worked, respectively.

Method

Participants and Design. Participants (N = 480) were Americans recruited from Amazon's Mechanical Turk in exchange for monetary compensation. Participants were randomly assigned to role conditions: *worker* or *observer*. Eleven participants were excluded for not following directions (e.g., providing nonsensical response to free response questions or listing end dates that followed the start date for their work projects). This left 469 participants in all analyses reported below.

Procedure. All participants were asked to recall projects on which someone had worked "*very hard*." Such projects were supposed to satisfy four criteria. First, the project should have been worked on over the course of one or more days during the last year. Second, the person completing the project was supposed to have had at least two weeks to work on the project. Third, the task was to be one that was assigned on a specific date and due on a specific, already-passed date. We noted that projects completed for an employer, family member, social or religious organization, or as part of a formal educational program all had the potential to satisfy these criteria.

The fourth criterion differed by participants' role condition. Workers were supposed to identify three assignments or tasks that they themselves had worked very hard on. Observers were asked to identify three tasks that someone they knew worked very hard on. For observers,

these three projects could have been completed all by the same person or by two or three different people.

After participants had provided a brief description of each of the very-hard-work projects, we returned to each task one-by-one. Participants characterized each task using the same measures used in Study 1. Participants indicated when the project was assigned and when it was due, from which we could calculate *assignment length*. Participants indicated how many days the project was actually worked on as well as the average number of hours of work on those days; we multiplied these responses to calculate the total *time worked.* In addition, participants indicated the *work intensity* while actually working on the project. We realistically conceded that participants may not have exact answers to each question, but that they should do their best to provide estimates.

Finally, we had participants complete the *worked hard* measure. Note that because we asked participants to constrain their recollections to times in which the self or another "worked very hard," this item was included not because we wished to explain variation in it. Instead, we wished to confirm that participants followed instructions and thought of prototypical working-hard experiences. For example, participants' recalled experiences were more representative of hard work than those recalled for Study 1 (when such instructions were not present). Analyses on this measure, as well as presentation of untransformed descriptive statistics of our other measures, are presented in the Supplemental Materials.

Results

All participants were asked to think of tasks that were characterized by especially hard work. But did the features of prototypically hard-work tasks display a different pattern when identified by the workers themselves as opposed to by social observers? We began by performing

certain data transformations that would make answering this question possible. Like in Study 1,
we log-transformed both the time worked and the assignment length variables to address positive
skew. Next, we standardized these two variables as well as the work intensity measure, so that all
measures were placed on the same (standardized) scale. Unlike in Study 1 workers and observers
considered different projects (and observers had a much broader set of potential projects from
which to draw), so this standardization was also important so that we could examine participants
in each role condition (worker or observer) in their relative prioritization of one hard-work
attribute over the other (i.e., time worked vs. work intensity). This is because comparisons
between workers and observers on the same attribute reflect a combination of our hypothesized
effects and actual differences between the set of recallable work episodes.

 To test our main hypothesis, we constructed a mixed model predicting participants'
ratings. The predictors were role (worker or observer), index (time worked or work intensity),
and crucially their interaction. We also included two random intercepts: one for the specific
project and one for the participant. Confirming our key prediction, this model returned a
significant Role × Index interaction, $F(1, 1321) = 11.93$, $p < .001$.

 We decomposed the interaction to determine whether observers' (compared to workers'
own) prototypical examples of hard work were especially characterized by time worked
(compared to work intensity). In short, this was the case. Others' prototypical hard-work
episodes were more likely to be characterized by more time worked ($M_{adjusted} = 0.11$, SE $= .05$)
than workers' examples from their own lives ($M_{adjusted} = -0.10$, $SE = .05$), $t(787.29) = 3.31$, p
$= .001$. This difference went away (and non-significantly reversed) when looking at the work
intensity metric: observers ($M_{adjusted} = -0.01$, SE $= .05$) vs. workers ($M_{adjusted} = 0.01$, $SE = .05$), $t <$

1. As discussed earlier, workers and observers recalled different events, which does not undermine the key interaction, but does potentially distort these comparisons.

Given the ratings were standardized, we can also perform pairwise comparisons of the indices within each role condition (thus holding constant the set of work episodes available for recall). This will allow us to understand the indices' relative prioritization in defining workers' own or social observers' identified hard-work experiences. Observers' prototypical hard-work episodes showed a relative prioritization of time worked over intensity, $t(1321) = -2.49$, $p = .013$. Workers' prototypical hard-work episodes showed a relative prioritization of intensity over time worked, $t(1321) = 2.39$, $p = .017$.

Finally, we wanted to make certain that differences in what characterized the self's own (i.e., workers') vs. others' (i.e., observers') hard-work episodes could not be attributable to differences in the scope or length of the recalled task. We regressed both of our key indices— time worked and work intensity—on log-transformed assignment length and saved the residuals. We then submitted these residuals to our initial model. If anything, we observed a stronger Role X Index interaction, $F(1, 1321) = 15.03$, $p < .001$. Observers' very-hard-work recollections prioritized time worked over intensity, $t(1321) = -2.80$, $p = .005$. Workers' very-hard-work recollections featured intensity over time worked, $t(1321) = 2.69$, $p = .007$ (Table 2).

The three sets of analyses converge toward the same conclusion, one that complements the findings of Study 1. When thinking about times that others displayed hard work, people identified times that others worked longer (but no more intensely) than the self. Because participants could identify recent examples of hard work from anyone else (instead of one particular person), one worry is that participants merely had access to more extreme examples of others' work projects. By analogy, one's own top three accomplishments for the year are unlikely

Table 2

Features of Workers' Own and Observers' Recalled Prototypical Hard-Work Experiences (Study 2)

Role	Standardized Features of Work Projects, Z (SE)		Test: P2 > P1
	F1: Work Intensity	F2: Time Worked	
R1: Workers	0.01 (0.05)	-0.11 (0.05)	$t = 2.69, p = .007$
R2: Observers	-0.01 (0.05)	0.12 (0.05)	$t = -2.80, p = .005$
Test: R2 > R1	$t < 1$	$t = 3.49, p = .001$	

Note. The table provides standardized means (and standard errors) to permit comparisons across the two features. Values are standardized residuals (with standard errors) that account for variation in log-transformed assignment length. Although workers had to recall three of their own hard-work experiences, observers could recall three hard-work experiences performed by anyone they knew. Given this offered observers a wider range of hard-work experiences to draw from, this makes direct comparisons between R1 and R2 suspect. This should not interfere with comparisons between F1 and F2 or the crucial Feature X Role interaction.

to be as impressive as the top three accomplishments from one's entire social network. Although this feature likely led to higher work intensity and time worked ratings from observers than from workers, this feature cannot account for why we observed a Role × Index interaction displaying the prioritization of time over intensity in people's social (as opposed to self) examples of hard work. In other words, if both components (time and intensity) had been equally important to self and social conceptions of hard work, then the larger pool of social work examples would simply have allowed participants to identify social examples that were higher on both dimensions. The interaction provides crucial support for our hypotheses that self and social definitions of hard work differ.

Study 3

In Study 3, workers completed two tasks—one that was quick but intense (high-intensity/low-time), one that was longer and less taxing (low-intensity/high-time). Observers merely learned what the tasks were but did not complete them themselves. Our first prediction was that workers (compared to observers) would think the quick, intense task required much more hard work than the longer but laxer one.

We then gave workers a choice of which type of task they would want to complete for a third and final task. Recall that we opened this manuscript with a discussion of the prevalence of procrastination—a work style that ends up being characterized by relatively intense but shortened bouts of work. We were curious whether workers—even without a deadline that sneaks up on them—actually prefer this relatively efficient (even if vigorous) style of work. We operationalized our tasks in a way that the high-intensity/low-time task would seem especially difficult to everyone (even if especially more so to workers). In this way, we thought that observers would expect workers would want to take the path of least resistance and repeat the low-intensity/high-time task. Suspecting that workers might be inflating how much hard work is required of the style of work they actually prefer (not merely the type of work poor time management often foists upon them), we predicted that the workers themselves might surprise observers by choosing the high-intensity/low-time task.

Method

Participants and design. Two hundred five participants were undergraduates drawn from a subject pool at an American university, who took part in exchange for course credit. Participants were randomly assigned to one of two roles: worker or observer.

Procedure. We begin by describing the procedure that workers completed. Then we describe how observers completed a similar procedure that required them to make judgments about workers. At that point, we provide more detail about the materials—namely, the "executive function tasks" that workers completed and observers judged.

Workers. After arriving to our lab and providing informed consent, workers were truthfully informed they would complete three executive function tasks. These tasks took one of two forms—a relatively intense task that could be completed in a short amount of time (*high-intense/low-time*) or a relatively long task that was simpler to complete (*low-intense/high-time*). After completing these two tasks in a counterbalanced order, workers were informed they would complete one final task. In this case, they could decide whether they wanted to complete a task of one form or the other—i.e., a high-intense/low-time or a low-intense/high-time task.

Before participants made this choice, we measured how much hard work each task was perceived to entail. For each form, participants indicated how hard they would "be expected to" and would "actually" work on the task of that form. Responses were provided on 1 (*not very hard*) to 9 (*very hard*) scales. These responses were tightly correlated, $r = .76$. As such, we averaged them to form *working hard* composites—one for each task. Finally, participants chose which form they wanted their third task to take. To avoid deception, we then had workers complete that task.

Observers. Observers went through a similar sequence, but simply learned about the tasks a worker would face instead of actually completing the tasks themselves. That is, observers read the detailed instructions that explained how to complete the opening two tasks—the high-intensity/low-time and the low-intensity/high-time tasks. These instructions communicated both the duration of each task and the specific instructions that made the task more or less intense. At
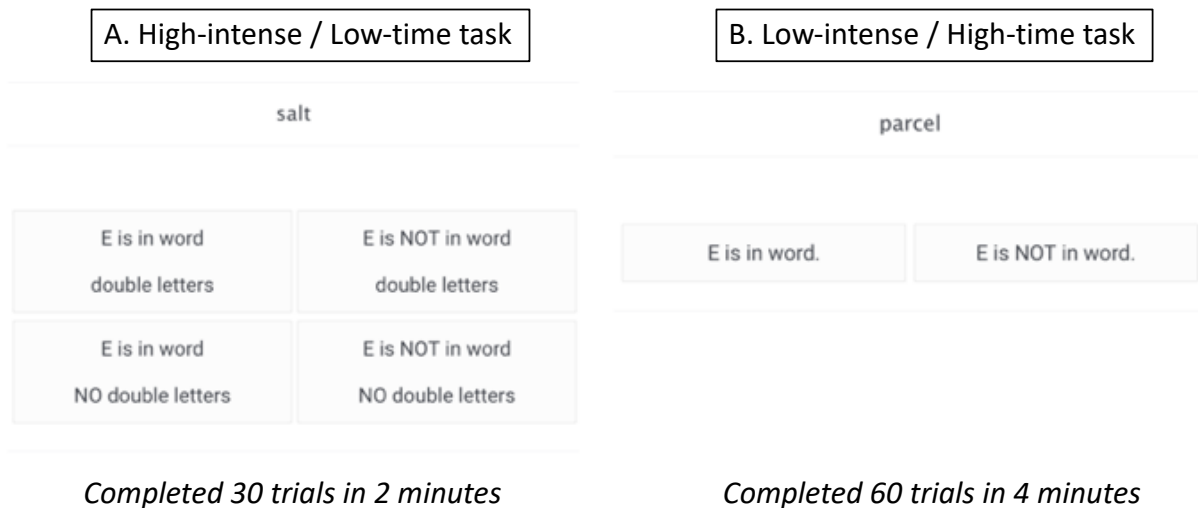
that point, observers completed the same *working hard* items as workers, but in this case by indicating how much *workers* would be expected to and actually would work on a final task of each form. As with workers, these items were averaged to form a *working hard* composite, *r* = .68. Observers then guessed what form of task the worker (about whom no individuating information was provided) would choose for their third task.

Materials. Both executive functioning tasks—the high-intensity/low-time and the low-intensity/high-time form—were similar. They each required participants to consider a list of words—presented one at a time—and make timed judgments about each word. The high-intensity/low-time task took *2 minutes* whereas the low-intensity/high-time form took *4 minutes*. Although participants in both tasks had 4 seconds to make each judgment, the judgments were more complex in the quicker, high-intensity than the longer, low-intensity format.

In the low-intensity form, participants merely had to indicate whether the letter 'E' was included in each word shown. Participants would click one of two options to indicate the letter's presence or absence. In the high-intensity form, participants had to identify not merely whether 'E' was present, but also whether the word contained consecutive identical letters (e.g., as in "apple" or "peel"). Participants would click one of four options to indicate whether or not the letter 'E' was present as well as whether or not a double letter was as well (see Figure 2). For both workers and observers, we counterbalanced the order in which they learned about (and for workers, completed) the two tasks. For workers, we counterbalanced which word list was used for which form of the task.

Results

Worked hard. We began by testing whether workers and observers had different perceptions of how much hard work each test required. To do so, we submitted the working hard

Figure 2

Example Trials from Both Tasks Used in Study 3

A. High-intense / Low-time task	B. Low-intense / High-time task
salt	parcel

E is in word	E is NOT in word
double letters	double letters

E is in word	E is NOT in word
NO double letters	NO double letters

E is in word.	E is NOT in word.

Completed 30 trials in 2 minutes   *Completed 60 trials in 4 minutes*

Note. Regardless of the task, participants had 4 seconds to complete each trial. The A task (high-intensity, low-time task) was completed in half the time that the B task (low-intense/high-time) was. Whereas the B task merely required participants to indicate whether the letter 'E' appeared in the target word, the A task also required participants to indicate whether the same letter appeared consecutively in the word.

ratings to a 2 (Role: worker vs. observer) × 2 (Task: low-intense/high-time vs. high-intense/low-time) repeated measures ANOVA. The model revealed a main effect of role, $F(1, 203) = 15.95$, $p < .001$, $\eta_p^2 = .073$, and a main effect of task, $F(1, 203) = 223.83$, $p < .001$, $\eta_p^2 = .524$. The latter identified the pattern that the high-intense/low-time task ($M_{adj} = 5.84$, $SE = 0.11$) was generally perceived as involving harder work than the low-intense/high-time task ($M_{adj} = 3.61$, $SE = 0.13$). Of course, our critical hypothesis deals with whether workers and observers differed in these perceptions. But this main effect is important, because it will allow us to test whether workers

actually (and perhaps surprisingly to observers) prefer the task that is universally perceived to require more hard work.

Of primary importance, there was also a Role × Task interaction effect, $F(1, 203) = 7.59$, $p = .006$, $\eta_p^2 = .036$ (see Figure 3). This interaction reflected that observers perceived the low-intense/high-time task ($M_{adj} = 4.19$, $SE = 0.18$) as requiring harder work than did the workers themselves ($M_{adj} = 3.02$, $SE = 0.18$; $t(203) = 4.61$, $p < .001$, $d = 0.64$). But once the intensity rose (and the time to complete the task dropped), workers saw the task as requiring as much hard work ($M_{adj} = 5.67$, $SE = 0.16$) as observers did ($M_{adj} = 6.01$, $SE = 0.16$), $t(203) = 1.52$, $p = .130$, $d = 0.21$. These findings support our main hypothesis: Observers (compared to workers) gave relatively more weight to a task's time instead of its intensity in deciding what constituted hard work.

Task choice. When given the choice of which work task they wanted for the third round, which did workers choose? Keep in mind that both workers and observers agreed that the high-intensity, low-time task was the one that required more work. Presumably this is why observers thought that only 25% of workers would actually select to repeat it. Workers—even more than observers—saw this task as requiring relatively more hard work. But 87% of such workers chose to repeat the high-intensity, low-time task. This large gap was statistically significant, $\chi^2(1, N = 205) = 64.87$, $p < .001$. In other words, workers were especially adamant that the high-intensity, low-time task was the one that reflected harder work, but they preferred that very route. Apparently, they wanted to get their work over with quickly (but intensely) and take credit for doing more work as well.

We probed the nature of workers' choices and observers' expectations further. For each participant, we created a *work chosen* (difference) *score.* For a worker, this was the hard work

Figure 3

Workers' and Observers' Judgments, Decisions, and Predictions about Study 3's Tasks
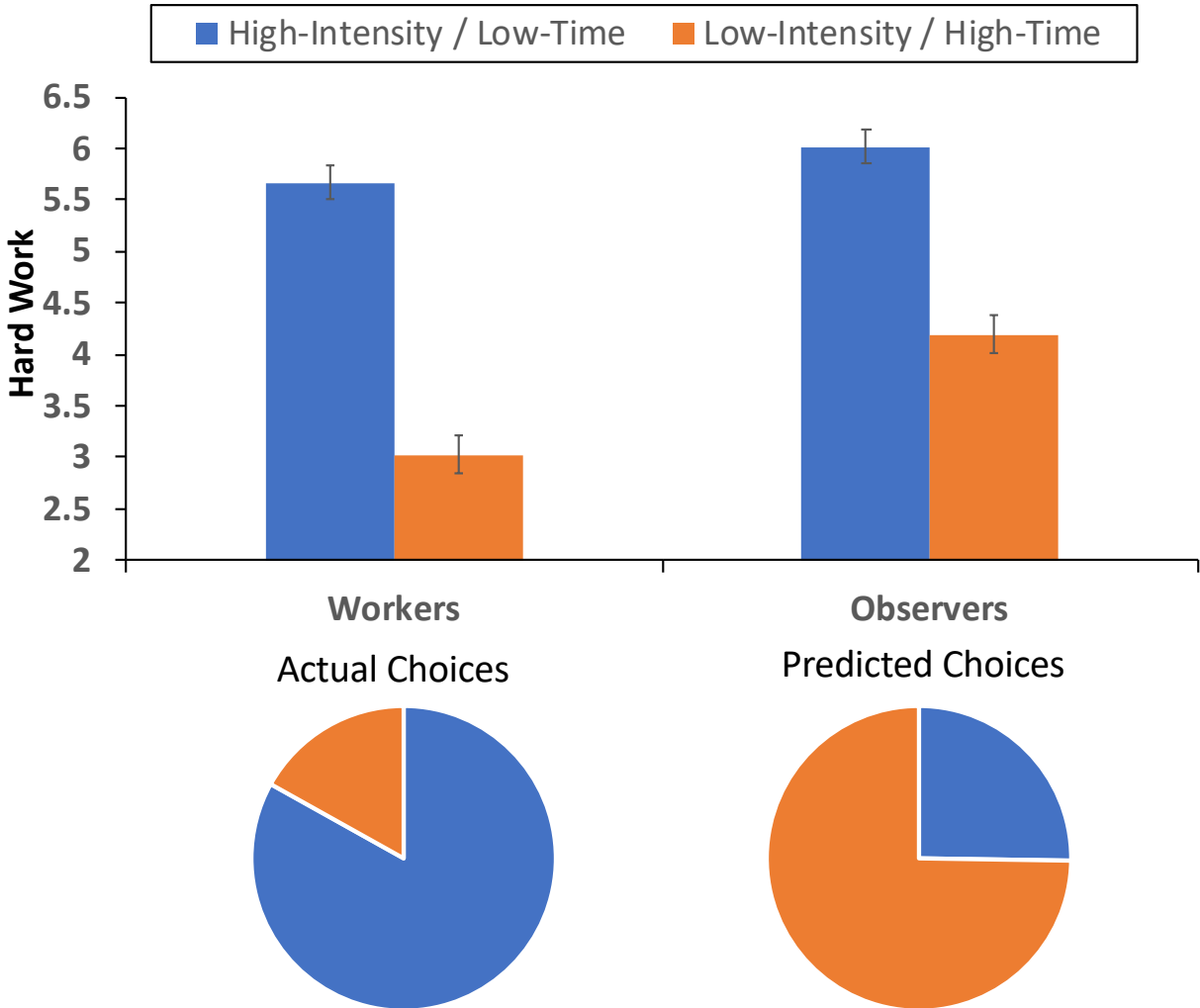
Note. Although workers and observers agreed that the high-intense/low-time task entailed more hard work than the low-intense/high-time task, workers were stronger in this relative belief. Most observers predicted that a worker would choose the low-time/high-intense task. Instead, the vast majority of workers chose to complete the high-intense/low-time task, whose hard-work requirements they (compared to observers) exaggerated. Error bars are ± 1 SE.

ratings of the task that that worker chose to repeat minus the hard work rating of the task the participant did not choose. For observers, this difference score reflected the hard work ratings of the task that that observer expected the worker to choose minus the hard work ratings of the task the observer thought the worker would not choose.

Observers expected the worker would select the task that required less hard work (M = -1.07, SD = 2.51, $t(102)$ = -4.34, p < .001). In contrast, workers tended to choose the task they perceived to require more hard work (M = 2.12, SD = 2.74, $t(101)$ = 7.83, p < .001). This difference was statistically significant, $t(203)$ = 8.71, p < .001, d = 1.22. In other words, workers surprised observers by seeking out the task they saw as harder work. What more, recall workers (compared to observers) exaggerated just how much more work that task was. In combination, this allows workers to pursue work in a manner that they apparently prefer (working intensely over less time) while telling themselves (even more than observers believe) that this approach reflects especially hard work. The fact that these results are consistent with those observed across a wide range of actual work experiences (in Studies 1 and 2) bolsters preliminary confidence in the generality of these results.

General Discussion

In deciding what it means to work hard, people seem to use qualitatively different criteria for themselves than they do for others. Workers' judgments of how hard they themselves had worked were more sensitive to how intensely they had worked, whereas yoked observers who considered the same episodes were more sensitive to how many hours the workers had put in (Study 1). Similarly, people's own recollections of times they (as opposed to others) had worked particularly hard showed a similar divergence (Study 2). Finally, in-lab workers who were in the middle of work tasks—compared to observers who merely saw what workers were having to

complete—showed similar tendencies to think that a quick-but-intense work task required much more work than a long-but-easier task (Study 3). Notably, workers preferred that style of work: They choose to repeat the very task that they (compared to observers) thought required much more hard work.

Especially in light of this last finding—that workers showed a surprising (to observers) preference for repeating a task that would be short but quite intense—it may be useful to consider the present findings in light of Dunning and colleagues' classic work on the role of idiosyncratic trait definitions as contributors to self-enhancement (Dunning & McElwee, 1995; Dunning, Meyerowitz, & Holzberg, 1989). That research highlights that people define desirable traits and competencies by playing to their own strengths. In the present case, workers seem to be doing something similar: In defining what it means to work hard, they seem to prioritize the component that they themselves naturally seek out (i.e., intensity over duration). Where the present findings differ from the past work is that participants actually do not apply this self-serving definition to judgments of others. Instead, they seem to switch things up, deciding that others' hard work is actually revealed by a different pattern.

This naturally raises the question of why these self and social definitions differ. One possibility foreshadowed above is that it is motivated: Workers can feel like solid workers who are living up to Weber's ideal by defining work in this way. But a second possibility relates to the difference between working long hours compared to working intensely. The former is a statistic. The latter is an experience. Knowing someone spent an extra ten hours in the office is easily interpretable. Appreciating the worker's phenomenological state during that time—the adrenaline, the struggle to maintain focus, the exhaustion of facing (and finally overcoming!) each hurdle—is hard for anyone else to fully appreciate. And even if observers can understand it

in the abstract, it is simply unlikely to carry the same phenomenological weight. This suggests a possible constraint on our effects' generality. When observers are closer to workers—either because they are working alongside them or simply hearing a narration of the emotional roller coaster that was the task—they may better understand the intensity behind the work and give it more weight as well. Testing the extent to which providing this information attenuates our effects could give one hint to whether they are cognitively or motivationally determined.

Employees, students, and other underlings often want their superiors to see them as hard workers. Whether people are able to set aside their personal conceptions of hard work to understand how others see evidence of it is a question for future research. For example, graduate students may impress their advisors more by emphasizing the number of hours they put into a revision than by trying to capture the intensity with which they attacked the draft. But whether workers appropriately override their own personal conceptions of what constitutes hard work in order to manage observers' impressions is a question for future research (see Leary & Kowalski, 1990).

For all of our discussion of how workers and observers define hard work differently, this naturally raises the question of whose definition is right. Of course, there is not a perfect way to adjudicate, but one approach may be to ask which work input—time or intensity—actually leads to better-quality work. Past research suggests that both time (Rau & Durand, 2000) and intensity (Engeser & Rheinberg, 2008; Jackson, Thomas, Marsh, & Smethurst, 2010) correlate with objective markers of work success. Better understanding when one or the other input best leads to better performance would be useful not merely for knowing how workers may best approach looming tasks, but may help to determine when workers' typically inflated self-views are actually justified.

References

Blau, G., & Ryan, J. (1997). On measuring work ethic: A neglected work commitment facet. *Journal of Vocational Behavior*, 51(3), 435-448.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), 366–381. https://doi.org/10.1037/0022-3514.67.3.366

Critcher, C. R., Helzer, E. G., & Dunning, D. (2011). Self-Enhancement via redefinition: Defining social concepts to ensure positive views of self. In M. D. Alicke, & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 69-91). New York, NY: The Guilford Press.

Csikszentmihalyi M. (2014) *Toward a Psychology of Optimal Experience. In: Flow and the Foundations of Positive Psychology*. Springer, Dordrecht.

Dunning, D., & McElwee, R. O. (1995). Idiosyncratic trait definitions: Implications for self-description and social judgment. *Journal of Personality and Social Psychology*, 68(5), 936–946. https://doi.org/10.1037//0022-3514.68.5.936

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, 57(6), 1082–1090. https://doi.org/10.1037/0022-3514.57.6.1082

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158–172

Furnham, A. (1984). The Protestant Worth Ethic: A review of the psychological literature. *European Journal of Social Psychology, 14*, 87-104.

Furnham, A., Bond, M., Heaven, P., Hilton, D., Lobel, T., Masters, J., … Van Daalen, H. (1993).

A comparison of protestant work ethic beliefs in thirteen nations. *Journal of Social

Psychology*, 133(2), 185–197. https://doi.org/10.1080/00224545.1993.9712136

Hard Work | English Oxford Dictionaries. (n.d.). Retrieved from

https://en.oxforddictionaries.com/definition/us/hard_work

Helmreich, R. L., & Spence, J. T. (1978). The Work and Family Orientation Questionnaire: An

objective instrument to assess components of achievement motivation and attitudes

toward family and career. *JSAS Catalog of Selected Documents in Psychology*. 8, 35.

Helzer, E. G., & Dunning, D. (2012). Why and when peer prediction is superior to self-

prediction: The weight given to future aspiration versus past achievement. *Journal of

Personality and Social Psychology, 103*, 38-53.

Jackson, S. A., Thomas, P., Marsh, H., & Smethurst, C. J. (2010). Relationships between flow,

self-concept, psychological skills, and performance. *Journal of Applied Sport

Psychology, 13*, 154-178.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-

enhancement is paved with good intentions. *Personality and Social Psychology Bulletin,

30*(3), 328–339.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of

Experimental Social Psychology,* 40(1), 91–98. https://doi.org/10.1016/S0022-

1031(03)00065-9

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-

component model. *Psychological Bulletin, 107*(1), 34–47. doi:10.1037/0033-

2909.107.1.34.

McClelland, D. (1961). The Achieving Society, D. Van Nostrand, New York.

Naquin, S. S., & Holton III, E. F. (2002). The effects of personality, affectivity, and work commitment on motivation to improve work through learning. *Human Resource Development Quarterly*, 13(4), 357-376.

Pronin, E. (2008). How we see ourselves and how we see others. *Science, 320,* 1177-1180.

Rau, W., & Durand, A. (2000). The Academic Ethic and College Grades: Does Hard Work Help Students to "Make the Grade"? *Sociology of Education*, 73(1), 19. https://doi.org/10.2307/2673197

Rozental, A., & Carlbring, P. (2014). Understanding and treating procrastination: A review of a common self-regulatory failure. *Psychology*, 5(13), 1488–1502. https://doi.org/http://dx.doi.org/10.4236/psych.2014.513160

Rynes, S. L., Gerhart, B., & Parks, L. (2005). Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology*, 56, 571-600.

Schaufeli, W. B., Shimazu, A., & Taris, T. W. (2009). Being driven to work excessively hard: The evaluation of a two-factor measure of workaholism in the Netherlands and Japan. *Cross-Cultural Research*, 43(4), 320-348.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). *Life after p-hacking.* Retrieved from the SSRN website: https:// ssrn.com/abstract=2205186

Stead, R., Shanahan, M. J., & Neufeld, R. W. J. (2010). I'll go to therapy, eventually: Procrastination, stress and mental health. *Personality and Individual Differences*, 49, 175-180. http://dx.doi.org/10.1016/j.paid.2010.03.028

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133(1), 65.

Tice, D. M., & Baumeister, R. F. (1997). Longitudinal Study of Procrastination, Performance,

    Stress and, Health: The Costs and Benefits of Dawdling. *Psychological Science*, 8, 454-

    458. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00460.x

Weber, M. (1905). *The Protestant Ethic and the Spirit of Capitalism*, Scribners, New York.

Williams, E. F., Gilovich, T., & Dunning, D. (2012). Being all that you can be: The weighting of

    potential in assessments of self and others. *Personality and Social Psychology Bulletin,*

    38, 143-154.

Supplemental Materials

Pretest

One hundred ninety-seven undergraduates enrolled in the same course at an American university took part in the study for course credit. Six participants who failed an attention check ("This question is simply to ensure you are paying attention…Please select 80 on the slider") were excluded from the analyses.

All participants began by seeing general instructions describing the purpose of the study: "We are interested in how you think about how you go about doing work (i.e. homework, reading, studying, projects, etc.) as a student at [university, masked for review]."

Hard work. Participants began by indicating in what percentile they thought they fell in terms of being a hard worker: "How hard do you work as a student at [university, masked for review]?" To make sure participants understood how the percentile scale worked, we labeled 0 (*I work less hard than every other student my year at [university, masked for review]*), 100 (*I work harder than every other student my year at [university, masked for review]*), and 50 (*I am right in the middle: I work harder than half of [university, masked for review] students my year, and less hard than the other half of [university, masked for review] students my year*). Crucially, we did not define for students what it meant to be a hard worker. Participants responded on a slider scale that defaulted to 50, meaning participants had to drag the slider up or down to indicate they saw themselves as a harder or less-hard worker than average.

Low-Time, High-intensity work. We wanted to know to what extent students found themselves—compared to their peers—as especially likely to find themselves working less time on projects than they first intended or expected (low-time), meaning they ended up working more intensely on projects than they first intended or expected (high-intensity): "Compared to

other students your year at [university, masked for review], are you more or less likely to work more intensely on a project than you first thought you would because you ended up having less time to work on it than you expected you would?" Participants offered a response on a 0-to-100 slider scale anchored at 0(*This is more true of every [university, masked for review] student my year than of me*), 100 (*This is more true of me than of every other [university, masked for review] student my year*), and 50 (*I am right in the middle: This is more true of half of [university, masked for review] students my year than of me, and this is more true of me than the other half of [university, masked for review] students my year.*) Participants responded on a slider scale by dragging the slider from the default response of 50.

High-Time, Low-intensity work. The final item was meant to reflect the opposite pattern. That is, we wanted to understand whether students saw themselves—compared to how they saw their peers—as one who worked less intensely over more time on projects than they had initially intended or expected. More specifically, participants saw, "Compared to other students your year at [university, masked for review], are you more or less likely to work less intensely than you could have on a project because you ended up having more time to work on it than you thought you would?" This item used the same question format and labels as the low-time, high-intensity work question.

Study 1

What differentiated Wave 1 from Wave 2 is we added two additional questions for workers: "When you were working on the project, how stressed were you about it?" and "When you were not working on the project, how stressed were you about it?" Participants responded to each question on a 9-point scale anchored at 1(*not at all stressed*) and 9(*extremely stressed*). We

either showed these responses to observers or had observers make the same judgments about the workers. We shared the stress information with only half of yoked observers.

For workers, how much stress they experienced working, but not while *not working*, predicted their own reports of how hard they worked. The same pattern was observed for observers who received workers' ratings. For observers who estimated workers' stress patterns, both stress ratings—stress experienced while working and while not working—predicted their working hard ratings. Although we find these patterns intriguing, we do not discuss them further given they do not relate to the reliance on time worked vs. work intensity as a predictor of hard work. But also, inclusion of (or exposure to) this measure did not shift our focal effects of interest.

At the end of the survey, we also included a slate of demographic questions that we never connected with participants' responses and thus never analyzed. These questions asked for participants' age, gender, race/ethnicity, political orientation, and educational attainment. These items were just a standard slate of questions included in many studies run by the lab.

Study 2

Worked Hard. Study 1 asked workers to reflect on any projects that met certain criteria, which allowed us to explain variation in how much those varied work experiences were seen (by oneself or others) to reflect hard work. Study 2 instead examined the characteristics of one's own or others' prototypical hard work experiences—i.e., examples that did indeed reflect others' hard work. There are several implications of this shift in measurement approach. First, we should find that participants recalled episodes that reflected harder work than they did in Study 1.

Second, whereas Study 1 held constant the content of work projects being judged, Study 2—as part of asking a research question that was complementary but not identical to that asked

in Study 1—did not. Note also that observers could draw from a much broader set of recollections (i.e., anyone else's work projects) than workers could. And indeed, we see some evidence of the consequence of that. Even though Study 1 found that when judging the exact same project, workers thought it reflected much harder work than did observers (Ms = 6.63 and 5.92, respectively; t = 7.24, p < .001), in Study 2 this gap was entirely eliminated (and even non-significantly reversed). That is, workers recalled slightly (though not significantly) less strong hard work experiences (M = 7.01) than did observers (M = 7.08), $t(451.57)$ = -0.60, p = .548. (Note also the increase in means, reflective of our instructions to everyone to recall examples in which the self or another had worked very hard.) Given we know from Study 1 that workers are much more charitable about their own degree of hard work than about others' when considering the exact same work experience, the fact that this gap is entirely eliminated is consistent with observers being able to draw on what might be more objectively demanding hard-work experiences than workers themselves could. Although this does not undermine the validity of our key interaction prediction (that observers will prioritize time worked over work intensity compared to workers), note that it does make comparisons between workers and observers on each index suspect.

Descriptive Statistics. Whereas the above analyses as well as the analyses reported in the main text included random effects for participant and/or project, we also present in Table S1 raw descriptive statistics for work intensity, time worked (in hours) and assignment length.

Table S1

Raw, Untransformed Features of Workers' Own and Observers' Recalled Prototypical Hard-Work

Experiences.

	Raw, Unadjusted Features of Work Projects, Mean (Median)		
Role	Work Intensity	Time Worked (hours)	Asgn. Length (Days)
Workers	6.97 (7)	98.08 (17)	123.53 (26)
Observers	6.94 (7)	114.59 (24)	165.28 (25)