Working Longer but Not Harder: The Effects of Input- and Output-Based Pay Allocations on the Duration and Intensity of Effort

Eric W. Chan
Assistant Professor
The University of Texas at Austin
eric.chan@mccombs.utexas.edu

Steven J. Kachelmeier
Randal McDonald Chair in Accounting
The University of Texas at Austin
kach@mail.utexas.edu

Xinyu Zhang
Doctoral Candidate
The University of Texas at Austin
xinyu.zhang@utexas.edu

August 2017

We are grateful for research funding from the Department of Accounting, McCombs School of Business, University of Texas at Austin. We acknowledge helpful comments from two 2017 ABO Research Conference reviewers, workshop participants at the University of Texas at Austin, and from Skylar Deture. The second author gratefully acknowledges support from the Randal B. McDonald Chair in Accounting.

Working Longer but Not Harder: The Effects of Input- and Output-Based Pay Allocations on the Duration and Intensity of Effort

ABSTRACT: Paired participants in an incentivized experiment solve anagram puzzles during six-minute work periods, generating a compensation pool that is allocated based on the input of the time each individual works on the task or on the output of each individual's correct puzzle solutions. To operationalize aversion to effort duration, participants can redeem unused time for individual payments in addition to their pay allocations from the production task. Relative to output-based pay allocations, we find that input-based allocations increase the amount of time participants devote to the task. Nevertheless, input-based pay also significantly lowers the *intensity* of participants' productive effort, even though participants have no meaningful alternatives for the time they willingly spend on the task. When pairs are highly heterogeneous in ability, we find some evidence of impaired total productivity under an input-based pay allocation scheme despite more time spent on the task. Our results suggest that the popularity of input-based incentives in practice can lead to the appearance of increased effort even if total productivity decreases.

Key words: Effort, compensation, productivity, inputs, outputs, experimental economics.

Working Longer but Not Harder: The Effects of Input- and Output-Based Pay Allocations on the Duration and Intensity of Effort

I. INTRODUCTION

A core assumption underlying the agency-theoretic perspective of management accounting is that effort is not directly observable, thus necessitating output-based compensation contracts to motivate desired actions. However, as has been observed analytically (e.g., Lazear 1986; Baker 2002) and empirically (e.g., Bell and Freeman 2001), effort often *is* observable and contractible, at least in terms of basic inputs such as the time that workers spend on productive activities. From an agency perspective, input-based compensation provides a more direct incentive that avoids the "second-best" risk sharing that agents must assume under outcome-based contracts. However, if differences in ability lead some workers to be more productive than others for the same levels of effort, input-based determinations of relative pay could lead to dysfunctional social dynamics within firms. To test these potentially opposing forces, we design an experiment to compare the effects of input-based and output-based pay allocations on effort duration, effort intensity, and worker productivity.

Contrasting the effects of input- and output-based pay is important because real-world compensation contracts and performance evaluations often reward hours worked as a key performance indicator (Landers, Rebitzer and Taylor 1996; Sousa-Poza and Ziegler 2003; Kruger, Wirtz, Van Boven, and Alternatt 2004; Yeung and Soman 2007). Effort, however, is a multifaceted construct that incorporates both *duration* and *intensity* (Bonner and Sprinkle 2002). Accordingly, even if effort duration is observable and contractible, effort intensity can vary in ways that could jeopardize the effectiveness of a compensation contract based on time spent. Of primary importance to our study is the possibility that input-based pay allocations could seem unfair if workers who put in the same amount of time receive the same pay even if they achieve different levels of productivity. If so, relative to output-based pay allocations, input-based allocations could lead employees to work longer but not as hard.

In our experiment, paired participants work individually on anagram puzzles, unscrambling letters to form ordinary words. This task is advantageous for our research objective because it is sensitive to natural variation in ability, which we measure from an uncompensated training period. After the training period, participants complete four work periods of six minutes each, generating a compensation pool for each pair during each period that is based on adding the number of anagram puzzles each participant correctly solves. That is, the task involves individual production, but the total pay available to each work pair is determined jointly, capturing the sense of a compensation base. We allocate this compensation based either on the input of relative effort duration, defined as the time each paired participant spends on the task, or on the output of the number of anagram puzzles each paired participant correctly solves.

To capture a sense of aversion to effort *duration*, participants are allowed to end each six-minute work period early, redeeming any unused time in ten-second increments for individual pay that is incremental to the pay allocation from the work period. This feature operationalizes a tradeoff between individual consumption and productive effort. However, for the time that participants willingly work on the anagram task, there is no alternative activity available and hence no meaningful opportunity to convert work time to leisure. Given that the total compensation pool from the production task depends only on output, participants would have little reason from an economic perspective to be averse to effort *intensity* during the few minutes they agree to work, insofar as both paired participants' pay strictly increases in the size of the compensation pool. Nevertheless, as we develop further below, we theorize behavioral reasons based on perceptions of fairness for why participants might not work as hard under an input-based pay allocation scheme.

Experimental findings indicate that basing pay allocations on the input of time spent significantly increases the time participants devote to the productive task. However, when paired workers are highly heterogeneous in ability, input-based pay also systematically decreases effort intensity per unit of time, controlling for ability. The net effect of these offsetting forces leads to some evidence of lower total productivity under input-based pay in highly heterogeneous pairs, despite more time spent. That is, allocating pay based on effort duration appears to induce participants to work longer but not as hard, with the latter effect potentially dominating the former. This finding provides an important caveat to the resurgence of input-based compensation schemes in practice that we discuss below, as incentivizing workers to put in more time does not necessarily imply that they will be more productive.

Two other recent papers have examined different facets of inputs and outputs in production.[1] Arnold, Hannan, and Tafkov (2016) conduct a team-based experiment in which participants do not literally exert effort, but rather choose a cost that is analogous to effort. The cost each participant chooses is multiplied by a privately revealed productivity factor to generate outputs. Findings indicate that teams are more productive when they know each other's inputs (i.e., cost decisions) than when they know both inputs and outputs or only outputs. In a different team-based experiment, Thomas and Thornock (2017) find that input-based feedback is less effective than output-based feedback in motivating participants to spend more time on a productive task. Importantly, both studies focus on inputs and outputs as *feedback*, whereas we

[1] A third recent experiment by Brüggen, Feichter, and Williamson (2017) also examines input and output measures, but is only tangentially related to our study. Brüggen et al. (2017) find that specifying uncompensated input or output targets on a routine task helps their experimental participants to achieve closure on that task and hence improve performance on a separate, creative task. Our study does not address the tradeoff between routine and creative performance.

focus on the agency-theoretic fundamentals of how participants are compensated, providing complete feedback to all participants. A compensation focus provides unique insights. Specifically, in contrast to Arnold et al.'s (2016) finding that input-based feedback facilitates increased effort, we find that input-based pay significantly lowers effort *intensity*. In contrast to Thomas and Thornock's (2017) finding that input-based feedback leads participants to spend less time on a productive task, we find that input-based pay significantly increases the time participants willingly spend on a productive task while simultaneously decreasing the intensity of that time. To our knowledge, ours is the first experiment to contrast the effectiveness of compensating inputs rather than outputs, which is of timely importance given the increasing prevalence of input-based rewards in practice, as we discuss shortly.

More broadly, our study relates indirectly to the experimental literature in management accounting on pay allocations (e.g., Fisher, Maines, Peffer, and Sprinkle 2005; Bailey, Hecht, and Towry 2011; Arnold et al. 2016), insofar as our primary manipulation involves how a compensation pool is allocated among workers. A key difference, however, is that prior studies have focused on the employer-employee dynamics that emerge when pay allocations are discretionary. In contrast, we examine the more primal question of whether pay allocations based on the input of time spent versus the output of items produced can lead to fairness-based productivity effects even in the absence of employer discretion.

Section II discusses the institutional and theoretical background for our study as the basis for developing the hypotheses we test. Section III describes our experimental procedure and design. Sections IV and V present our primary results and supplemental analyses, respectively, and Section VI concludes.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Input-Based Compensation in Practice

Notwithstanding the agency-theoretic perspective that most rewards should be outcome-based, substantial empirical evidence indicates that employees and executives who put in more hours enjoy better evaluations and increased compensation (Landers et al. 1996; Bell and Freeman 2001; Sousa-Poza and Ziegler 2003; Kruger et al. 2004; Yeung and Soman 2007; Gicheva 2013; Reid 2015). Experimental and field-based research indicates that superiors reward greater "face time" even when objective performance measures are held constant (Elsbach, Cable, and Sherman 2010; Yam, Fehr, and Barnes 2014). The rewards conferred to those who spend more time on the job has likely led to increased work hours over time, particularly in the United States (Bell and Freeman 2001; Mishel, Bivens, Gould, and Shierholz 2012; Saad 2014).

Moreover, at least anecdotally, evidence suggests that the use of input-based rewards in the workforce is on the rise. For example, having experimented with output-based, flexible work arrangements, companies such as Yahoo, Zappos, and the Bank of America (Miller and Rampell 2013) as well as Best Buy (Lee 2013) have returned in recent years to more rigid workplace policies that monitor and reward time spent at the workplace. Input-based rewards are likely facilitated by trends towards more open physical work environments (Konnikova 2014; Newman, Stikeleather, and Waddoups 2016) as well as by technological innovations that enable firms to monitor effort duration more effectively (Reid 2015).

The Effects of Rewarding Inputs versus Outputs on Effort Duration

If an organization rewards time spent on the job, simple economic reasoning suggests that workers should respond by increasing effort duration. In contrast to what agency theory characterizes as the "second-best" solution of rewarding outputs, an input-based pay scheme does not force employees to bear the risk of uncertainty that links inputs to outputs. In short, outcome-based pay provides only an indirect incentive to work, as effort duration leads to productive output with uncertainty. The noise in this linkage depresses its incentive effect (Hwang, Erkens, and Evans 2009). In contrast, input-based pay provides a direct incentive to work. Given that workers generally respond most strongly to performance measures that are rewarded directly (Baker, Jensen, and Murphy 1988; Milgrom and Roberts 1992; Prendergast 1999), we expect participants to work longer when total compensation is allocated based on the time each paired participant spends on the productive task than when it is allocated based on the output each participant produces.

> **H1:** Effort *duration* will be greater under an input-based pay allocation scheme that rewards effort duration than under an output-based pay allocation scheme that rewards productivity.

The Effects of Rewarding Inputs versus Outputs on Effort Intensity

As Bonner and Sprinkle (2002) explain, effort depends not only on duration, or how long people work, but also on intensity, or how hard people work. In our experiment, we intentionally minimize the potential for aversion to effort intensity, given the chosen level of effort duration. First, our experiment limits work periods to no more than six minutes, such that our task is unlikely to be compromised by the effects of fatigue that can impair effort intensity when work periods take several hours (e.g., Crouch 2015). Second, given that the total compensation pool available to paired participants in our study depends only on productive outputs, participants would have little economic reason not to work as hard as possible during the brief amount of time that they willingly devote to the productive task. Notwithstanding this reasoning, below we develop arguments of a more *behavioral* nature, rooted in fairness, on why an input-based pay allocation scheme could threaten effort intensity.

In addition to effort duration and intensity, productive success in our anagram-based task is also likely to depend on *ability*, as we confirm in our findings. An input-based pay allocation system ignores ability, treating all time spent as equivalent for compensation purposes. Such a system suggests a potential to be exploited by low-ability workers, who might be frustrated by the high level of task difficulty they perceive. Given the relatively low returns to effort intensity for a low-ability worker in the presence of a higher-ability coworker, working longer but not harder could seem an attractive strategy under an input-based pay allocation scheme. That is, even if task characteristics do not make effort intensity particularly aversive, effort intensity might not seem particularly beneficial either to a low-ability worker if an input-based pay allocation scheme presents an opportunity to free-ride on a higher-ability counterpart.

For different reasons, we posit that the effort intensity of *high-ability* workers is also likely to be jeopardized by an input-based pay allocation scheme. Evidence dating back at least as far as early experimental work by Adams and Rosenbaum (1962) indicates that perceptions of inequity have harmful effects on performance. In our setting, high-ability workers produce more but do not earn more than low-ability workers for the same amount of time spent under an input-based pay allocation scheme. In contrast, output-based pay rewards high-ability workers for their superior performance. As a consequence, we predict that different perceptions of fairness will lead to lower effort intensity under input-based pay.

Importantly, the above reasoning does not necessarily imply conscious decisions to reduce effort intensity. As noted above, under either input- or output-based pay allocations in our task setting, both low- and high-ability participants have the economic incentive to work hard in order to maximize the group compensation pool over the course of the brief work periods they willingly allot to the task. Nevertheless, psychology and management research finds that the distractions induced by emotion-laden moods can impair performance, even subconsciously (e.g., Kane and Montgomery 1998; Montgomery, Kane, and Vance 2004). If an input-based compensation scheme that rewards different levels of time spent rather than productivity achieved generates such moods, we should find support for our second hypothesis:

> **H2:** Effort intensity will be *lower* under an input-based pay allocation scheme that rewards effort duration than under an output-based pay allocation scheme that rewards productivity.

The Moderating Role of Heterogeneous Abilities

If our theoretical rationale is valid, support for H2 should be strongest when workers are of widely different abilities, thus increasing the potential for free-riding by low-ability workers and resentment by high-ability workers. That is, if support for H2 comes from behavioral forces (i.e., concerns for fairness) rather than economic forces (i.e., effort aversion), we should observe the strongest effect for relatively heterogeneous work pairs. This reasoning implies an interaction between pay allocation and worker heterogeneity, which we test as our third hypothesis:

> **H3:** The extent to which effort intensity is lower under an input-based pay allocation scheme than under an output-based pay allocation scheme will be most pronounced when workers are highly heterogeneous in ability.

The Effects of Rewarding Inputs versus Outputs on Production

If allocating pay based on time spent increases effort duration but lowers effort intensity, a natural question arises as to what net effect these offsetting forces would have on total productivity. Experiments are ill-suited to calibration questions that rely on the relative magnitudes of different forces. Nevertheless, we examine the effects of input- versus output pay allocations on total productivity as a nondirectional research question to provide evidence of the potential for input-based pay to lower total productivity even if it increases effort duration.

> **RQ:** What effect will input- versus output-based pay allocations have on total productivity?

III. METHOD AND DESIGN

Participants and Task

We recruit 96 undergraduate business student volunteers to participate in one of eight 75-minute experimental sessions at the dedicated business research laboratory of a large public university. Participants face computer stations (not each other) during the experiment and complete all tasks using z-Tree software (Fischbacher 2007). Each participant receives a show-up fee of $5 in addition to the performance-contingent compensation described below.

Participants read instructions that describe the task of solving anagram puzzles by unscrambling letters to form ordinary words, such as the instructional examples of NONIO = ONION and BYETRTA = BATTERY. After reading the instructions, participants take a computerized comprehension quiz, programmed such that incorrect answers lead to remedial instructions. This process ensures that all participants have an accurate understanding of the task and rules before proceeding. As the instructions explain, the experiment begins with an uncompensated training period during which all participants solve as many anagram puzzles as possible in three minutes. In addition to providing a practice opportunity, the training period provides us with a measure of individual ability that we use as a covariate.

After the training period, we pair participants to form two-person "firms." We program z-Tree to assign the top performer in the top half of the training period performance distribution with the top performer in the bottom half, and so forth until the bottom performer in the top half is paired with the bottom performer in the bottom half. While the intent of this process is to ensure *some* heterogeneity in worker ability within each pair, it does not establish *equal* heterogeneity across pairs. Accordingly, we construct a measured factor for analysis purposes based on a median split of observed heterogeneity, defining "high heterogeneity" as work pairs that differ by three (the median) or more correct puzzle solutions during the training period. This factor allows us to test the moderating role of heterogeneous abilities in H3 and to condition on heterogeneity in other analyses. Participants are unaware of the process used to pair workers, as the instructions only inform participants (truthfully) that "you will be anonymously matched with another participant to form a two-worker firm." While avoiding deception, this approach guards against the potential for participants to engage in strategic behavior during the training period. Pairings are anonymous because participants cannot track their paired coworkers to specific computer stations.

Paired participants engage in four work periods in which they solve anagram puzzles for up to six minutes (360 seconds). Before each work period, the instructions inform participants that that they can end the work period earlier than the full six minutes if desired, earning an individual payment of 60 lira (our experimental currency, as discussed shortly) for each unused block of ten seconds. The instructions provide the example that if a participant chooses to work 240 seconds on the anagram task during any of the four work periods instead of the full 360 seconds, that participant would earn an additional payment of 60 lira × 12 ten-second blocks for the 120 seconds of unused time = 720 lira. This feature has the effect of operationalizing an aversion to effort duration, given that the payment for unused time establishes an opportunity cost. Absent this feature, it is unlikely that participants would be strongly averse to a few minutes of experimental time, given that they have already volunteered for the experiment and have no alternative activities in which they can engage during the experimental session.

Work production by both paired participants determines a compensation pool of 400 lira for each puzzle solved correctly by either coworker. Thus, across conditions, the total money available for distribution is based only on each pair's joint output, capturing the reality that only outputs are salable in practice. We manipulate how the output-based compensation pool is allocated, however, as described shortly.

Across conditions, each pair of participants receives feedback after each work period regarding the amount of time each participant spent on the anagram task and the number of puzzles each participant correctly solved. We also provide output feedback after the training period, such that participants can gauge their relative abilities. Thus, we hold constant any behavioral influence of input and output feedback that other studies have examined as a separate issue (Arnold et al. 2016; Thomas and Thornock 2017).

Experimental Design

We manipulate whether the outcome-based compensation pool each pair generates is allocated based on the input of time spent by each participant or on the output of puzzles solved correctly by each participant. The experimental instructions describe these allocations in percentage terms, meaning that each paired participant's percentage of the total time spent or total output produced determines that participant's percentage of the total pay for that pair under the input- and output-based conditions, respectively.

Thus, for each work period, each participant earns the sum of the pay for unused time (60 lira for each unused ten seconds), that participant's input- or output-based share of the total compensation pool of 400 lira multiplied by the combined number of anagrams solved correctly by either paired participant, and a fixed-pay component of 1,000 lira that we hold constant across conditions. At the end of the experiment, we aggregate this pay across the four work periods and convert it to U.S. dollars at the rate of 1,000 lira = $1.00, as indicated in the instructions. Including the $5.00 show-up fee, total pay averages to $22.36 and ranges from $15.01 to $31.96.

IV. PRIMARY RESULTS

Descriptive Statistics

Table 1 provides means and standard deviations by condition for effort duration, effort intensity, and productivity across each of the four work periods and overall. We define effort duration as the number of seconds a participant agrees to work on the anagram task in each work period instead of redeeming for individual payment as unused time. We define effort intensity as the number of puzzles solved correctly within each 100 seconds a participant agrees to work.[2] Thus, our measure of effort intensity controls for different levels of effort duration. We acknowledge that the number of puzzles solved per 100 seconds clearly reflects ability in addition to effort intensity. For that reason, our measure is not as clean as a more direct measure of effort intensity, such as Hecht, Rotaru, Schulz, Towry, and Webb's (2016) proposal of a biometric measure of effort intensity based on eye pupil dilation. That being said, we control for individual ability from the training period as a covariate in all hypothesis tests, such that the number of puzzles solved per 100 seconds, controlling for ability, should serve as a reasonable proxy for effort intensity. Finally, we define productivity as the total number of anagram puzzles each participant solves correctly within each work period.

As Table 1 indicates, within each work period, average effort duration is greater in the input-based pay allocation condition than under output-based pay, while average effort intensity is *lower* under input-based pay. Simple *t*-tests of overall averages across periods that do not control for heterogeneity or ability indicate that both differences are statistically significant in the directions predicted by hypotheses H1 and H2 ($t = 1.70$; one-tailed $p = 0.05$ for effort duration and $t = -1.96$; one-tailed $p = 0.03$ for effort intensity). In contrast, an overall *t*-test does not detect a statistically significant difference between input- and output-based pay allocations for total

[2] Arithmetically, if a participant solves Y puzzles in X seconds, our measure of effort intensity is 100(Y/X).

productivity (t = -0.82; two-tailed p = 0.42), although we report evidence later of a marginally significant productivity difference in favor of output-based pay when pairs are characterized by highly heterogeneous abilities. Overall, Table 1 provides preliminary support for both H1 and H2, which we corroborate in the more refined tests reported below. We collapse across periods in our primary analyses and comment later on period-specific differences as a supplemental analysis.

We are reasonably confident that our experiment achieves random assignment, as we detect no significant treatment differences (all $p \geq 0.49$) in age, gender, risk preferences, or social-value orientation measures that we obtain from a post-experimental questionnaire.

Results for Effort Duration

Table 2 reports means and standard deviations in Panel A and an ANCOVA in Panel B for effort duration, defined as the number of seconds each participant works each period on the anagram task, averaged across the four work periods. The ANCOVA factors are *pay allocation scheme*, reflecting our manipulation of input- versus output-based pay allocations, and *heterogeneity*, reflecting whether an observation is from a participant in a pair that is at or above the median ability difference of three puzzles solved during the training period. Although our theoretical development of H1 gives us no reason to suspect sensitivity to heterogeneous abilities in our test of effort duration, we posit such a difference in H3 for our test of effort intensity, and hence include the heterogeneity factor in the effort duration analysis for completeness. The covariate in our ANCOVA is *individual ability*, which we classify as high, medium, or low on an ordinal scale depending on whether a participant's performance in the training period is above, at, or below the median of three puzzles solved.[3] We do not interact the ability measure with our

[3] All reported results are robust to an alternative measure of *individual ability* based on the total number of puzzles completed during the training period, with the exception that we obtain weaker results (F = 1.40; two-tailed p = 0.24) for our test of the simple effect of input-based pay allocation on total productivity for high-heterogeneity

treatment factors in our primary analyses, but we separate high- and low-ability participants in supplemental analyses.

Figure 1 depicts our findings. For both high-heterogeneity and low-heterogeneity pairs, Figure 1 indicates that participants devote more time to the anagram task (rather than redeeming time for individual payment) under input-based pay than under output-based pay. ANCOVA results in Panel B of Table 2 confirm that the difference is statistically significant ($F = 3.06$; one-tailed $p = 0.04$), supporting our directional prediction in H1. Neither the heterogeneity factor nor the pay allocation scheme × heterogeneity factor interaction are significant, indicating that the effort duration results are largely consistent across pairs of more and less heterogeneous abilities. The ability covariate itself is also not significant, indicating that individual ability does not appear to condition participants' willingness to work on the production task. Overall, findings for effort duration support our H1 prediction that workers will spend more time on a task when the pay allocation scheme rewards that time.

Results for Effort Intensity

Table 3 reports and Figure 2 depicts our findings for effort intensity, defined as the number of puzzles solved correctly by each participant per 100 seconds of allotted time, averaged across the four work periods. Although the main effect of pay allocation is marginally significant in the direction predicted by H2 ($F = 2.24$; one-tailed $p = 0.07$), Figure 2 shows that this main effect is driven entirely by high-heterogeneity work pairs. Consistent with H3, Panel B of Table 3 indicates that the pay allocation scheme × heterogeneity interaction is significant ($F = 6.68$; $p = 0.01$), as is the simple effect of pay allocation within high-heterogeneity pairs in Panel C ($F = 9.48$; one-tailed $p < 0.01$). As Figure 2 illustrates, participants in high-heterogeneity

pairs. However, as reported in our supplemental analyses, this test becomes *stronger* ($F = 3.76$; two-tailed $p = 0.06$) if we include high or low relative ability as a factor in an ANOVA within the high-heterogeneity pairs.

pairs are about one puzzle less productive per 100-second interval (i.e., lower effort intensity) under input-based pay allocation than under output-based pay. The fact that this difference is manifest only for high-heterogeneity pairs suggests that it reflects the fairness-based reasoning we use to motivate H2 and H3 rather than simple effort aversion to working hard under input-based pay. For low-heterogeneity pairs, effort intensity is actually slightly *higher* under input-based pay, although this difference is not statistically significant ($F = 0.53$; two-tailed $p = 0.47$).

The individual ability covariate in the Table 3, Panel B ANCOVA is highly significant ($F = 32.25$; $p < 0.01$), suggesting that our measure of effort intensity also captures a significant element of inherent ability. However, as explained earlier, our inclusion of ability from the training period as a covariate in the ANCOVA should be able to extract the ability effect from our measure of effort intensity for assessing treatment effects.

We conclude that input-based pay leads our participants to work longer but not as hard as they work under output-based pay. Given that work periods last a maximum of six minutes, it is unlikely that effort aversion alone or worker fatigue can entirely explain this finding. Rather, the fact that we discern an effort intensity effect only for pairs that are highly heterogeneous in ability suggests a more social interpretation, namely that low-ability workers may tend to free ride and/or high-ability workers build resentment when a compensation scheme rewards time spent but not output produced. We pursue this interpretation further in supplemental tests we report in Section V.

Results for Total Productivity

Table 4 reports and Figure 3 depicts our findings for average total productivity per work period, addressing our nondirectional research question on the net effect of the opposing forces we find in support of H1 and H2 for effort duration and effort intensity, respectively. As noted previously in our discussion of Table 1, we find no *overall* treatment difference for total productivity. However, after splitting the sample on heterogeneity and controlling for individual ability, Panel B of Table 4 indicates that the effect of pay allocation on productivity interacts significantly with heterogeneity (F = 4.03; p = 0.05). Following up on this interaction, the simple effects tested in Panel C indicate at least a marginally significant treatment effect in favor of greater productivity under output-based pay for participants in high-heterogeneity pairs (F = 2.72; two-tailed p = 0.10).[4] Workers in these pairs solve about 1½ more puzzles under output-based pay (i.e., means of 7.41 and 8.99 for input- and output-based pay allocations, respectively), despite the fact that they work longer under input-based pay, as reported previously in our test of H1. For low-heterogeneity pairs, the difference in productivity is in the opposite direction (see Figure 3), but the simple effect of pay allocation is not statistically significant for such pairs (F = 1.49; two-tailed p = 0.23).

We hesitate to overinterpret our findings for productivity, given that laboratory experiments are not well-suited for testing the calibration of relative magnitudes of different forces. However, we surmise from these findings that it is at least *possible* that impaired effort intensity could potentially more than offset any gains in effort duration under an input-based pay allocation scheme. That is, companies that reward work hours may well induce their employees to put in more hours, but if our findings generalize, they could also plausibly get lower productivity, particularly if employees differ in inherent abilities.

[4] This analysis measures productivity for each individual worker. An analysis of total productivity for each *pair* of workers would reduce our statistical power considerably by halving our cell sizes. Nevertheless, an analysis of paired productivity achieves similar results, albeit at somewhat weaker levels of statistical significance. Specifically, we find a marginally significant interaction between pay allocation and heterogeneity on total paired productivity (F = 3.68; p = 0.06). The F-statistic for the simple effect of the pay allocation treatment within high-heterogeneity pairs falls from 2.72 to 2.38 (two-tailed p = 0.13).

V. SUPPLEMENTAL ANALYSES

In this section, we report a series of supplemental tests intended to provide further support for our reported findings and corroborate our interpretations of those findings.

Intensity and Productivity Differences for High- and Low-Ability Workers

Our findings in support of H3 indicate that input-based pay lowers effort intensity only for highly heterogeneous work pairs. By construction, such pairs have a relatively high-ability worker and a relatively low-ability worker, prompting the follow-up question of whether our H3 result applies to both ability levels. Table 5 addresses this question with descriptive statistics in Panel A and an ANOVA in Panel B for participants of relatively high- and low-abilities *within* the high-heterogeneity classification. Given that we treat relative ability within the high-heterogeneity pairs as a factor in this ANOVA, we no longer include a covariate for individual ability, as doing so would induce collinearity from two measures of ability in the same analysis.[5]

Panel A of Table 5 indicates that effort intensity is lower under input-based pay for both high- *and* low-ability workers. The ANOVA in Panel B confirms the H3 finding that input-based pay decreases effort intensity within the high-heterogeneity pairs ($F = 9.60$; $p < 0.01$). The relative ability factor is also significant ($F = 20.15$; $p < 0.01$), which stands to reason given that high-ability workers, by definition, can solve more anagram puzzles in any given amount of time. However, the pay allocation scheme × relative ability interaction is not significant ($F = 1.24$; $p = 0.27$). Given that the same directional difference characterizes both ability levels, we conclude that the support for H3 is driven more by *heterogeneous* abilities than by abilities *per se*. That is, both high- and low-skilled participants appear to cut back on effort intensity when the compensation pool is allocated based on time spent on the task. This finding is

[5] Nevertheless, we reach the same conclusions regarding treatment effects from the supplemental analyses reported in this section even if we include a covariate for individual ability.

consistent with our reasoning that, under input-based pay, low-ability workers are likely to exert less effort intensity because they can free-ride on their high-ability counterparts, while high-ability workers are also likely to exert less effort intensity because they sense inequity from the pay allocation scheme.

Table 6 reports a similar ANOVA for the effects of pay allocation and relative ability within the high-heterogeneity pairs, but with total productivity instead of effort intensity as the dependent variable. Results are somewhat stronger than the overall productivity analysis in Table 4, as the productivity difference in favor of output-based pay comes close to the conventional 0.05 level ($F = 3.76$; two-tailed $p = 0.06$). The relative ability factor is highly significant, confirming that high-ability workers are more productive than low-ability workers across conditions ($F = 18.74$; $p < 0.01$). As with our supplemental test of effort intensity, there is no discernable interaction between the pay allocation scheme and relative ability ($F = 0.01$; $p = 0.92$), indicating that the productivity differences we observe characterize both high- and low-ability workers. As Panel A of Table 6 indicates, although high-ability workers are clearly more productive than their low-ability counterparts, workers of both types generate about 1½ more correct puzzle solutions within the high-heterogeneity pairs when pay is allocated based on puzzles solved rather than time spent. Moreover, this effect occurs despite the fact that output-based pay leads workers of both types to put in *less* time, as reported earlier in our test of H1.

Perceptions of Fairness

To the extent that our findings reflect participant perceptions that it is unfair to provide the same time-based pay to workers of different productivity, we should observe evidence of this interpretation from our post-experimental questionnaire. Specifically, the first post-experimental question asks participants, “To what extent do you believe the way in which the bonus pool was allocated between you and your co-worker was fair?" The Likert-based response scale is anchored at 1 = "Not at all fair," 4 = "Moderately fair," and 7 = "Very fair." Mean responses are tallied in Table 7 by heterogeneity classification and by relative ability level within each classification. For both high- and low-ability workers in both high- and low-heterogeneity pairs, we discern essentially the same significant difference (all $p < 0.01$) in favor of the output-based pay allocation scheme being perceived as more fair (overall average = 6.17) than is the input-based scheme (overall average = 3.43). It is interesting that even low-ability workers share this view, given that they arguably benefit the most from a system that rewards time spent rather than output produced. We conclude that differences in perceived fairness are consistent with the effort intensity effect we observe in support of H2 and H3.

Preferences for an Output-Based Pay Allocation Scheme

Another post-experimental question asks participants to indicate their preferences between an input-based and output-based pay allocation. Table 8 shows the percentages of participants who indicate a preference for an output-based scheme, categorized by treatment condition, heterogeneity levels, and relative ability levels. With one exception, participants overwhelmingly indicate that they would prefer the bonus pool to be allocated based on outputs. Interestingly, the notable exception is that only 23 percent of low-ability workers in high-heterogeneity pairs working under input-based pay indicate a preference for output-based pay. It appears that, even though low-ability workers in highly heterogeneous pairs perceive an input-based pay allocation scheme to be unfair (Table 7, as discussed previously), they still recognize that they benefit under such a scheme and express a preference for it when assigned to that condition.

Is the Effort Intensity Effect Intentional?

Although Table 3 provides evidence that participants in high-heterogeneity pairs exert lower effort intensity under input-based pay than under output-based pay, this result does not necessarily imply that participants knowingly work less hard, as effort effects can operate in the subconscious even without willful intent (Kane and Montgomery 1998; Montgomery, Kane, and Vance 2004). To shed light on this question, a post-experimental question asks participants to indicate their level of agreement with the statement, "My strategy is to maximize the amount of output I produce in the anagram task," with Likert-scale responses anchored at 1 = "Do not agree at all" and 7 = "Very much agree."

Table 9 analyzes average responses to this question within high-heterogeneity pairs, with descriptive statistics in Panel A, an ANOVA that interacts the pay allocation scheme factor with relative ability in Panel B, and simple-effect follow-up tests for high- and low-ability workers in Panel C. The only significant finding from the ANOVA is a marginally significant pay allocation scheme × relative ability interaction ($F = 3.06$; $p = 0.09$). Panel C indicates that low-ability workers perceive that they did not try as hard to maximize output under input-based pay (average = 4.92) than under output-based pay (average = 6.00), for a difference of 1.08 ($F = 4.72$; two-tailed $p = 0.03$). In contrast, high-ability workers indicate nearly the same intent to maximize output under either input-based pay (average = 6.08) or output-based pay (average = 5.93), for an insignificant difference of only 0.15 ($F = 0.09$; $p = 0.77$). We conclude from this analysis that the effort intensity effect for low-ability workers may have been intentional, as they perceive that they did not work as hard under an input-based pay allocation scheme. However, the similar if not stronger effort intensity effect we discern for high-ability workers appears to reflect a subconscious, unintentional reaction, as these participants perceive that they worked just as hard under both pay allocation schemes even though their effort intensity measures indicate otherwise.

Effects Over Time

All analyses to this point have collapsed across work periods by using overall averages as our dependent measures. To assess the potential for different effects over time, we also conduct expanded, period-specific analyses as supplemental tests. The only time-sensitive result we detect is for effort intensity among high-heterogeneity pairs. As Table 10, Panel A reports, the pay allocation scheme × period interaction for effort intensity is marginally significant at $p = 0.07$ if we justify a one-tailed test under the rationale that the fairness-related pressures driving the effort intensity effect should become more acute over time. Panel B of Table 10 reports simple effects within each work period, indicating that the significance level becomes stronger with each period, from an insignificant effect of pay allocation on effort intensity in the first work period (one-tailed $p = 0.24$) to a significant effect in the second work period (one-tailed $p = 0.04$) and a highly significant effect by the third and fourth periods (both one-tailed $p < 0.01$). We interpret these findings as lending further support to our interpretation that growing perceptions of inequity impair effort intensity under an input-based pay allocation scheme among workers in high-heterogeneity firms.

Effects on Total Payoffs

Given that participants benefit from sharing in the compensation pool *and* from redeeming unused time for individual payment, productivity alone does not capture the entirety of total participant payoffs for the experiment. Accordingly, Table 11 tests whether participants are better off or worse off under an input-based pay allocation scheme relative to an output-based scheme by reporting an ANOVA with total payoffs by participant as the dependent variable.

Consistent with our approach for other supplemental tests, we restrict this analysis to high-heterogeneity pairs for which we find the strongest effects of different pay allocation schemes on effort intensity. Table 11, Panel B indicates a significant effect of pay allocation on participant payoffs ($F = 6.78$; $p = 0.01$), but this effect must be interpreted in the context of a significant interaction between the pay allocation scheme and participants' relative ability levels ($F = 6.97$ $p = 0.01$). As the means reported in Panel A show, high-ability workers receive lower average payoffs from the anagram task under input-based pay ($17.99) than under output-based pay ($22.73),[6] a difference that is statistically significant ($F = 13.75$; $p < 0.01$). In contrast, low-ability workers generate nearly identical average payoffs under input-based ($16.86) and output-based ($16.82) pay ($F = 0.00$; $p = 0.98$).

These findings suggest that workers with the highest abilities have the most to lose from an input-based pay allocation scheme, largely because they do not work as hard for the time they agree to spend on the task (Table 5). In contrast, low-ability workers are shielded somewhat from the dysfunctional effects of input-based pay on effort intensity, as such a scheme allows low-ability workers to "free-ride" on their high-productivity coworkers. That being said, input-based pay does not appear to make low-ability workers significantly *better off* either. Put differently, input-based pay appears to lower the welfare of high-ability workers without improving the welfare of low-ability workers. Thus, from the perspective of Pareto optimality, our results appear to favor output-based pay.

Compensating both Input and Output

In our primary experiment, we operationalize a strong manipulation by allocating the compensation pool based entirely on the input of time spent or on the output of correct solutions. This design maximizes power for testing our theory-based predictions. In practice, however,

[6] These averages do not include the $5.00 fixed show-up fee paid to all participants.

many firms are likely to reward both inputs *and* outputs. Thus, for completeness, we report in this subsection a supplemental condition with 46 additional participants in which half of the available total compensation for each pair is allocated based on time spent and the other half is allocated based on puzzles solved.

Perhaps unsurprisingly, results from this "composite" input-and-output condition (untabulated) lie in between the purely input-based and output-based conditions. Specifically, within the high-heterogeneity pairs for which we find the strongest results in our primary experiment, the composite condition that allocates pay based on both input and output results in average effort duration of 313 seconds, average effort intensity of 2.87 puzzles per 100 seconds, and average productivity of 8.33 puzzles. These averages lie in between the corresponding input- and output-based averages of 318 and 302 seconds for effort duration, 2.28 and 3.16 puzzles per 100 seconds for effort intensity, and 7.41 and 8.99 puzzles solved. However, with the exception of a marginally significant difference in effort intensity between the composite and input-based conditions for high-heterogeneity pairs ($F = 2.40$; one-tailed $p = 0.06$), we do not have sufficient power to detect statistically significant differences at conventional levels between the composite condition and either the input- or the output-based conditions for any of our dependent variables, even though the input- and output-based conditions significantly differ from each other. Thus, we cannot offer strong conclusions regarding how a composite input-and-output-based pay allocation scheme would differ behaviorally from an exclusively input-based or output-based scheme, except to say that it would mostly likely lie somewhere in between these extremes.

VI. CONCLUSIONS

In principle, economic reasoning suggests that the ability to contract directly on effort should have the effect of inducing greater effort than could be achieved under indirect incentives that reward output instead, *ceteris paribus*. However, effort hinges on both duration and intensity, such that incentives for one of these dimensions may not carry over to the other (Bonner and Sprinkle 2002). In an incentivized experiment, we find that, relative to allocating total pay based on outputs produced, an allocation based on effort duration does indeed lead our participants to spend more time on a productive task instead of redeeming unused time for individual consumption. However, the gains we observe in effort duration under an input-based scheme are offset but losses in effort intensity, despite the fact that participants should arguably have little aversion to working hard for the few minutes they agree to spend on a laboratory task, especially given that the total compensation pool is based only on productive output. We find some evidence that, among the most heterogeneous work pairs in terms of relative ability differences, the net effect of these offsetting forces can lower total productivity under an input-based pay allocation scheme despite its gains in effort duration.

Our findings have implications for the popularity of input-based rewards in practice. Of particular interest is the possibility that firms such as Yahoo, Zappos, Bank of America and Best Buy that have shifted away from flexible, output-based rewards towards more insistence on "face time" at the workplace (Miller and Rampell 2013; Lee 2013) may get a sense of false confirmation from observable increases in time spent on the job even if less visible indicators of effort intensity decline. That is, our results suggest that input-based incentives are likely to give the appearance of increased effort on the surface, in terms of hours worked. In the laboratory, we are able to tease out the offsetting loss in effort intensity, but that loss may be more easily masked in real-world settings, with the result that input-based incentives may appear to be effective even if the opposite is true.

An important feature of our experiment is that the *total* compensation our paired participants share is based solely on the outputs they produce. Thus, our only manipulation is the input- or output-based allocation of total pay. This feature is realistic in the sense that only outputs generate salable cash flows that firms can use to pay their employees. Moreover, to the extent that we find lower effort intensity under an input-based pay allocation scheme even if the total available compensation depends only on outputs, it seems reasonable to surmise that our results would be even stronger if both the amount of the available compensation and its allocation were to depend on productive inputs versus outputs. That is, we view this feature as working against our prediction that input-based pay impairs effort intensity, thus strengthening the inferences we can draw from observing such a finding. Nevertheless, our conclusions are potentially limited by our examination of input- versus output-based *allocations* of compensation, as opposed to input- versus output-based *sources* of compensation.

Our conclusions are also limited by other stylized features that we employ in the traditions of experimental economics. Specifically, our participants are not literally employees, nor do they literally work for "firms." There are no residual claimants in our study besides the participants themselves, such that our participants effectively serve as both employees and shareholders. Moreover, solving anagram puzzles in six-minute work periods is likely to differ in multiple ways from real-world production tasks that can take hours, days, weeks, or months. Still, if we observe effort intensity effects in a laboratory environment in which the task takes just a few minutes and workers have no meaningful alternative activities for the time they have agreed to spend on the task, it seems likely that real-world effort intensity effects could also arise when employees can show up for work but divert that time to other activities.

We encourage additional research that challenges the fundamental agency-theoretic assumption that effort is unobservable. Whether directed to feedback (Arnold et al. 2016; Thomas and Thornock 2017), setting production targets (Brüggen et al. 2017), or our focus on pay allocation schemes, employee inputs and outputs provide viable performance measures that can be used in different ways for different purposes. As the tradeoff between effort duration and effort intensity in our study illustrates, utilizing the input of effort duration can lead to some consequences that appear beneficial, while other consequences could be unintended.

REFERENCES

Adams, J. S., and W. B. Rosenbaum. 1962. The relationship of worker productivity to cognitive dissonance about wage inequities. *Journal of Applied Psychology* 46 (3): 161–164.

Arnold, M.C., Hannan R.L., and I. Tafkov. 2016. Mutual monitoring and team member communication in teams. Working paper, University of Bern, Tulane University, and Georgia State University.

Bailey, W. J., G. Hecht, and K. L. Towry. 2011. Dividing the pie: The influence of managerial discretion extent on bonus pool allocation. *Contemporary Accounting Research* 28 (5): 1562-1584.

Baker, G. 2002. Distortion and risk in optimal incentive contracts. *The Journal of Human Resources* 37 (4): 728.

Baker, G. P., M. C. Jensen, and K. J. Murphy. 1988. Compensation and incentives: Practice vs. theory. *The Journal of Finance* 43 (3): 593.

Bell, L. A., and R. B. Freeman. 2001. The incentive for working hard: explaining hours worked differences in the US and Germany. *Labor Economics* 8 (2): 181–202.

Bonner, S.E., and G. B. Sprinkle. 2002. The effect of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27(May): 303-345.

Brüggen, A., C. Feichter, and M. G. Williamson. 2017. The effect of input and output targets for routine tasks on creative task performance. *The Accounting Review* 92: in press.

Crooch, D. 2015. Efficiency up, turnover down: Sweden experiments with six-hour working day. *The Guardian* (September 17). Available at https://www.theguardian.com/world/2015/sep/17/efficiency-up-turnover-down-sweden-experiments-with-six-hour-working-day.

Elsbach, K. D., D. M. Cable, and J. W. Sherman. 2010. How passive 'face time' affects perceptions of employees: Evidence of spontaneous trait inference. *Human Relations* 63 (6): 735–760.

Fischbacher, U. Z-Tree Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2007) 171-178.

Fisher, J. G., L. A. Maines, S. A. Peffer, and G. B. Sprinkle. 2005. An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review* 80 (2): 563-583.

Gicheva, D. 2013. Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics* 31 (4): 785–824.

Hecht, G., K. Rotaru, A. Schulz, K. L. Towry, and A. Webb. 2016. Decoding effort: Toward a measure of effort intensity in management accounting research. Working paper, University of Illinois at Urbana-Champaign, Monash Univrsity, La Trobe University, Emory University, and University of Waterloo.

Hwang, Y., D. H. Erkens, and J. H. Evans. 2009. Knowledge sharing and incentive design in production environments: Theory and evidence. *The Accounting Review* 84 (4): 1145–1170.

Kane, K., and K. Montgomery. 1998. A framework for understanding disempowerment in organizations. *Human Resource Management* 37 (3-4): 263.

Konnikova, M. 2014. The Open Office Trap. *The New Yorker*. Available at http://www.newyorker.com/currency-tag/the-open-office-trap.

Kruger, J., Wirtz, D., Van Boven, L., and T. W. Altermatt. 2004. The effort heuristic. *Journal of Experimental Social Psychology* 40(1): 91-98.

Landers, R. M., Rebitzer, J. B., & Taylor, L. J. 1996. Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review* 86(3): 329-348.

Lazear, E. P. 1986. Salaries and Piece Rates. *The Journal of Business* 59 (3): 405.

Lee, T. 2013. Best Buy ends flexible work program for its corporate employees. Minneapolis, MN *Star Tribune*, December 13. Available at http://www.startribune.com/no-13-best-buy-ends-flexible-work-program-for-its-corporate-employees/195156871/.

Milgrom, P. R., and J. Roberts. 1992. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Miller, C. C., and C. Rampell. 2013. Yahoo orders home workers back to the office. *The New York Times*, February 13.

Mishel, L., Bivens, J., Gould, E., and H. Shierholz. 2012. *The State of Working America*. Cornell University Press.

Montgomery, K., K. Kane, and C. M. Vance. 2004. Accounting for differences in norms of respect: A study of assessments of incivility through the lenses of race and gender. *Group and Organization Management* 29 (2): 248–268.

Newman, A. H., Stikeleather, B., and N.Waddoups. 2016. Organizational openness, relative performance information, and employee performance. Working paper, University of South Carolina.

Reid, E. 2015. Embracing, passing, revealing, and the ideal worker image: How people navigate expected and experienced professional identities. *Organization Science* 26 (4): 997–1017.

Saad, L. 2014. The ‘40-hour’ workweek is actually longer—by seven hours. *Gallup.* Available at http://www.gallup.com/poll/175286/hour-workweekactually-longer-seven-hours.

Sousa-Poza, A., and A. Ziegler. 2003. Asymmetric information about workers' productivity as a cause for inefficient long working hours. *Labor Economics* 10 (6): 727–747.

Thomas T.F., and T. A. Thornock. 2017. Me vs. we: The effect of incomplete team member feedback on cooperation of self-regarding individuals. Working paper, University of Wisconsin – Madison and the University of Nebraska.

Yam, K. C., R. Fehr, and C. M. Barnes. 2014. Morning employees are perceived as better employees: Employees’ start times influence supervisor performance ratings. *Journal of Applied Psychology* 99 (6): 1288–1299.

Yeung, C. W., and D. Soman. 2007. The duration heuristic. J*ournal of Consumer Research* 34(3): 315-326.

FIGURE 1

Effort Duration

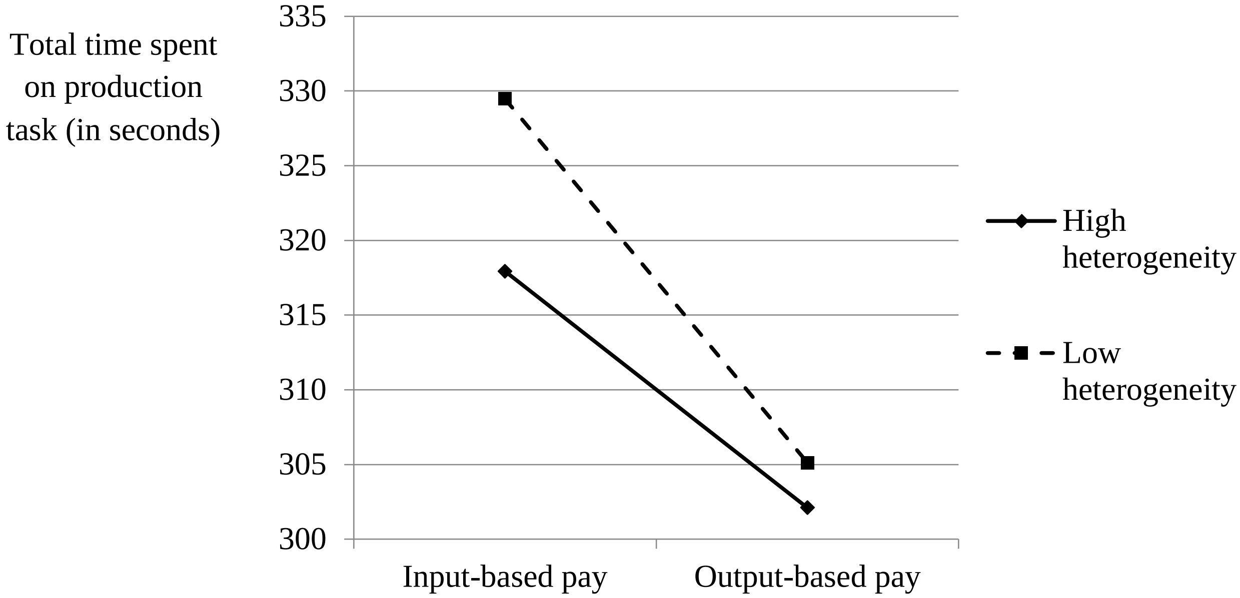

FIGURE 2

Effort Intensity

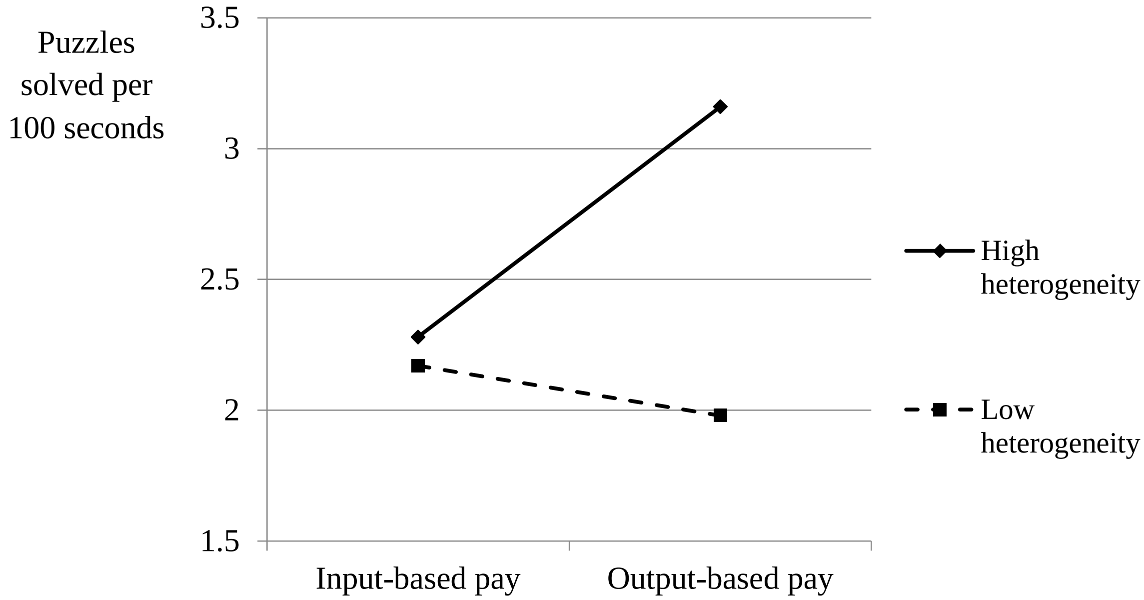

FIGURE 3

Total Productivity

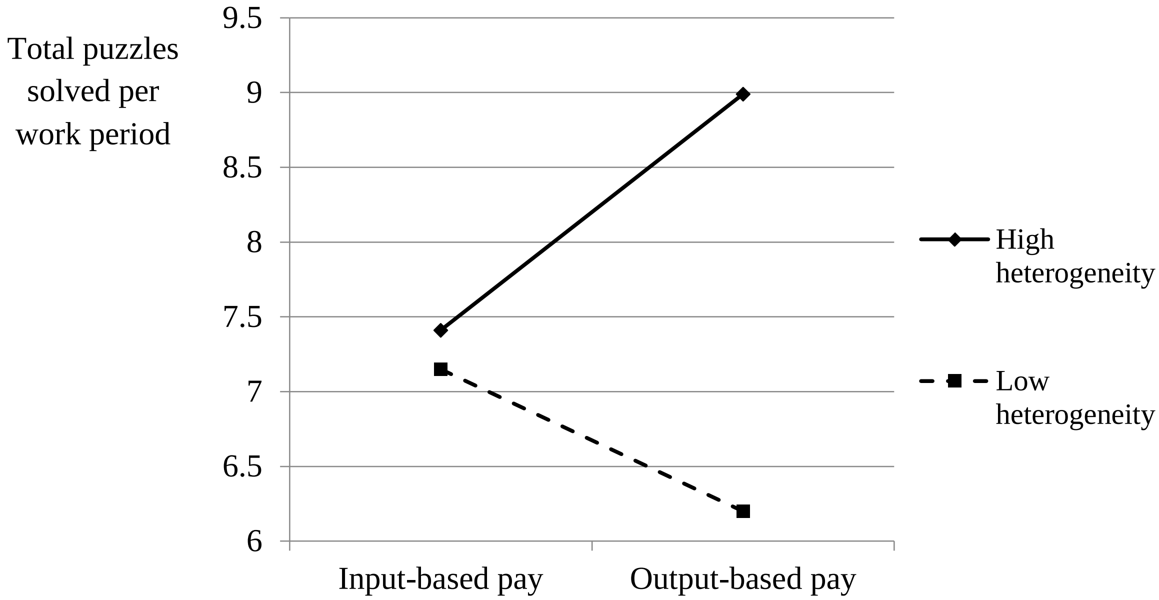

TABLE 1

Period-Specific Means and (Standard Deviations) by Pay Allocation Condition

	Work Period 1		Work Period 2		Work Period 3		Work Period 4		Overall Averages		Overall Difference between
	Input-Based[a] (n=48)	Output-Based[a] (n=48)	Input-Based (n=48)	Output-Based (n=48)	Input-Based (n=48)	Output-Based (n=48)	Input-Based (n=48)	Output-Based (n=48)	Input-based (n=48)	Output-Based (n=48)	Input-Based and Output-Based Pay Allocation
Effort duration[b]	313.42 (72.42)	305.17 (85.93)	326.16 (54.23)	305.69 (67.13)	314.52 (79.68)	298.70 (73.92)	338.78 (45.87)	303.89 (74.34)	323.22 (47.13)	303.36 (65.69)	Difference = 19.86 (t = 1.70; $\boldsymbol{p}$ **= 0.05**)[e]
Effort intensity[c]	2.17 (0.98)	2.31 (1.19)	2.27 (1.22)	2.70 (1.47)	1.68 (1.12)	2.37 (1.12)	2.81 (0.93)	3.31 (2.05)	2.23 (0.89)	2.67 (1.28)	Difference = 0.44 (t = -1.96; $\boldsymbol{p}$ **= 0.03**)
Productivity[d]	6.69 (3.28)	7.00 (3.57)	7.44 (3.98)	8.08 (3.95)	5.58 (3.91)	6.94 (3.38)	9.46 (3.42)	9.29 (3.80)	7.29 (3.14)	7.83 (3.25)	Difference = 0.54 (t = -0.82; p = 0.42)

Notes:
[a] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).
[b] *Effort duration* is the number of seconds out of 360 possible that a participant chooses to work on the anagram task instead of redeeming for individual payment.
[c] *Effort intensity* is the number of puzzles correctly solved for each 100 seconds devoted to the production task.
[d] *Productivity* is the total number of correct puzzle solutions provided by a participant during the work period.
[e] Reported p-values in **boldface** are one-tailed, conditional on hypothesized predictions. All other tabulated p-values are two-tailed.

TABLE 2

The Effects of Pay Allocation Scheme and Heterogeneity on Effort Duration

Panel A: Means (Standard Deviations) for *Effort Duration*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
High-heterogeneity pairs[c]	317.93	302.11
	(54.39)	(66.61)
	(n=26)	(n=28)
Low-heterogeneity pairs[c]	329.47	305.11
	(37.06)	(66.05)
	(n=22)	(n=20)
Overall averages	323.22	303.29
	(47.13)	(65.69)
	(n=48)	(n=48)

Panel B: Analysis of Covariance

Factor	df	Mean Square	*F*	*p*-value[e]
Pay allocation scheme[b]	1	10171.14	3.06	**0.04**
Heterogeneity[c]	1	2048.50	0.62	0.43
Pay allocation scheme × Heterogeneity	1	319.05	0.10	0.76
Covariate: Individual ability[d]	1	2665.76	0.80	0.37
Error	91	3327.75		

Notes:

[a] *Effort duration* is the number of seconds out of 360 possible that a participant chooses to work on the anagram task instead of redeeming for individual payment.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] The *heterogeneity* factor is based on a median split of relative abilities, differentiating work pairs that exhibit a difference of three or more puzzles completed during the training period (high heterogeneity) from work pairs that exhibit a difference of two or fewer puzzles completed during the training period (low heterogeneity).

[d] The *individual ability* covariate is an ordinal measure coded as high (2), medium (1), or low (0), depending on whether a participant performs above, at, or below the median of three puzzles completed during the training period.

[e] Reported *p*-values in **boldface** are one-tailed, conditional on hypothesized predictions. All other tabulated *p*-values are two-tailed.

TABLE 3

The Effects of Pay Allocation Scheme and Heterogeneity on Effort Intensity

Panel A: Means (Standard Deviations) for *Effort Intensity*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
High-heterogeneity pairs[c]	2.28 (1.01) (n=26)	3.16 (1.40) (n=28)
Low-heterogeneity pairs[c]	2.17 (0.75) (n=22)	1.98 (0.67) (n=20)

Panel B: Analysis of Covariance

Factor	df	Mean Square	F	p-value[e]
Pay allocation scheme[b]	1	1.79	2.24	**0.07**
Heterogeneity[c]	1	4.00	5.00	0.03
Pay allocation scheme × Heterogeneity	1	5.35	6.68	0.01
Covariate: Individual ability[d]	1	25.80	32.25	< 0.01
Error	91	0.80		

Panel C: Simple Effects

		df	F	p-value[e]
Effect of pay allocation scheme for high-heterogeneity pairs	1	1	9.48	**< 0.01**
Effect of pay allocation scheme for low-heterogeneity pairs	1	1	0.53	0.47

Notes:

[a] *Effort intensity* is the number of puzzles correctly solved for each 100 seconds devoted to the production task.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] The *heterogeneity* factor is based on a median split of relative abilities, differentiating work pairs that exhibit a difference of three or more puzzles completed during the training period (high heterogeneity) from work pairs that exhibit a difference of two or fewer puzzles completed during the training period (low heterogeneity).

[d] The *individual ability* covariate is an ordinal measure coded as high (2), medium (1), or low (0), depending on whether a participant performs above, at, or below the median of three puzzles completed during the training period.

[e] Reported p-values in **boldface** are one-tailed, conditional on hypothesized predictions. All other tabulated p-values are two-tailed.

TABLE 4

The Effects of Pay Allocation Scheme and Heterogeneity on Total Productivity

Panel A: Means (Standard Deviations) for *Productivity*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
High-heterogeneity pairs[c]	7.41	8.99
	(3.61)	(3.62)
	(n=26)	(n=28)
Low-heterogeneity pairs[c]	7.15	6.20
	(2.54)	(2.49)
	(n=22)	(n=20)

Panel B: Analysis of Covariance

Factor	df	Mean Square	*F*	*p*-value[e]
Pay allocation scheme[b]	1	0.22	0.03	0.86
Heterogeneity[c]	1	16.24	2.36	0.13
Pay allocation scheme × Heterogeneity	1	27.72	4.03	0.05
Covariate: Individual ability[d]	1	241.15	35.07	< 0.01
Error	91	0.68		

Panel C: Simple Effects

	df	*F*	*p*-value[e]
Effect of pay allocation scheme for high-heterogeneity pairs	1	2.72	0.10
Effect of pay allocation scheme for low-Heterogeneity pairs	1	1.49	0.23

Notes:

[a] *Productivity* is the total number of correct puzzle solutions provided by a participant during the work period.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] The *heterogeneity* factor is based on a median split of relative abilities, differentiating work pairs that exhibit a difference of three or more puzzles completed during the training period (high heterogeneity) from work pairs that exhibit a difference of two or fewer puzzles completed during the training period (low heterogeneity).

[d] The *individual ability* covariate is an ordinal measure coded as high (2), medium (1), or low (0), depending on whether a participant performs above, at, or below the median of three puzzles completed during the training period.

[e] Given that we do not make directional predictions for total productivity, all *p*-values in this table are two-tailed.

TABLE 5

Effort Intensity at Different Ability Levels within High-Heterogeneity Pairs

Panel A: Means (Standard Deviations) for *Effort Intensity*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
Higher relative ability[c]	2.76	3.95
	(1.05)	(1.46)
	(n=13)	(n=14)
Lower relative ability[c]	1.80	2.36
	(0.74)	(0.74)
	(n=13)	(n=14)

Panel B: Analysis of Variance

Factor	df	Mean Square	*F*	*p*-value[d]
Pay allocation scheme[b]	1	10.50	9.60	**< 0.01**
Relative ability[c]	1	22.04	20.15	< 0.01
Pay allocation scheme × Relative ability	1	1.35	1.24	0.27
Error	50	0.96		

Notes:

[a] *Effort intensity* is the number of puzzles correctly solved for each 100 seconds devoted to the production task.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] In this supplemental ANOVA, the *relative ability* factor differentiates the higher-ability participant within a high-heterogeneity pair from the lower-ability participant within that pair, with ability determined from the training period.

[d] Reported *p*-values in **boldface** are one-tailed, conditional on hypothesized predictions. All other tabulated *p*-values are two-tailed.

TABLE 6

Total Productivity at Different Ability Levels within High-Heterogeneity Pairs

Panel A: Means (Standard Deviations) for *Productivity*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
Higher relative ability[c]	9.21	10.71
	(3.53)	(2.90)
	(n=13)	(n=14)
Lower relative ability[c]	5.61	7.27
	(2.77)	(2.70)
	(n=13)	(n=14)

Panel B: Analysis of Variance

Factor	df	Mean Square	*F*	*p*-value[d]
Pay allocation scheme[b]	1	33.55	3.76	0.06
Relative ability[c]	1	167.16	18.74	< 0.01
Pay allocation scheme × Relative ability	1	0.08	0.01	0.92
Error	50	8.92		

Notes:

[a] *Productivity* is the number of correct puzzle solutions provided by each participant during the work period.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is are allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] In this supplemental ANOVA, the *relative ability* factor differentiates the higher-ability participant within a high-heterogeneity pair from the lower-ability participant within that pair, with ability determined from the training period.

[d] Given that we do not make directional predictions for total productivity, all *p*-values in this table are two-tailed.

TABLE 7

Perceived Fairness of Allocations[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]	t-test[e]
High-heterogeneity pairs:[c]			
Higher relative ability[d]	3.08 (0.37) (n=13)	6.07 (0.47) (n=14)	Difference = 2.99 (t = 4.94; p < 0.01)
Lower relative ability[d]	3.38 (0.31) (n=13)	6.21 (0.35) (n=14)	Difference = 2.83 (t = 6.00; p < 0.01)
Low-heterogeneity pairs:[c]			
Higher relative ability[d]	3.82 (0.55) (n=11)	6.60 (0.22) (n=10)	Difference = 2.78 (t = 4.50; p < 0.01)
Lower relative ability[d]	3.45 (0.47) (n=11)	5.80 (0.44) (n=10)	Difference = 2.35 (t = 3.60; p < 0.01)

Notes:
[a] This table reports means and (standard deviations) for participants' responses to the post-experimental question, "To what extent do you believe the way in which the bonus pool was allocated between you and your co-worker was fair?" The Likert-based response scale is anchored at 1 = "Not at all fair," 4 = "Moderately fair," and 7 = "Very fair."
[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).
[c] The *heterogeneity* factor is based on a median split of relative abilities, differentiating work pairs that exhibit a difference of three or more puzzles completed during the training period (high heterogeneity) from work pairs that exhibit a difference of two or fewer puzzles completed during the training period (low heterogeneity).
[d] *Relative ability* differentiates the higher-ability participant within any given work pair from the lower-ability participant within that pair, with ability determined from the training period.
[e] All reported *p*-values in this supplemental table are two-tailed.

TABLE 8

Percentages of Participants Who Express a Preference for Output-Based Pay[a]

	Input-Based Pay Allocation[b]	**Output-Based Pay Allocation[b]**
High-heterogeneity pairs:[c]		
Higher relative ability[d]	92% (n=13)	93% (n=14)
Lower relative ability[d]	23% (n=13)	77% (n=14)
Low-heterogeneity pairs:[c]		
Higher relative ability[d]	91% (n=11)	90% (n=10)
Lower relative ability[d]	63% (n=11)	80% (n=10)

Notes:

[a] This table reports the percentages of participants whose responses to a post-experimental question indicate a preference for an output-based allocation of the bonus pool over an input-based allocation of the bonus pool.

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] The *heterogeneity* factor is based on a median split of relative abilities, differentiating work pairs that exhibit a difference of three or more puzzles completed during the training period (high heterogeneity) from work pairs that exhibit a difference of two or fewer puzzles completed during the training period (low heterogeneity).

[d] *Relative ability* differentiates the higher-ability participant within any given work pair from the lower-ability participant within that pair, with ability determined from the training period.

TABLE 9

Self-Perceptions of Strategies to Maximize Output within High-Heterogeneity Pairs

Panel A: Means (Standard Deviations) for *MaxOutput*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
Higher relative ability[c]	6.08	5.93
	(0.38)	(0.29)
	(n=13)	(n=14)
Lower relative ability[c]	4.92	6.00
	(0.45)	(0.28)
	(n=13)	(n=14)

Panel B: Analysis of Variance

Factor	df	Mean Square	*F*	*p*-value[d]
Pay allocation scheme[b]	1	2.90	1.76	0.19
Relative ability[c]	1	3.95	2.39	0.13
Pay allocation scheme × Relative ability	1	5.06	3.06	0.09
Error	50	1.66		

Panel C: Simple Effects

	df	*F*	*p*-value[d]
Effect of pay allocation scheme for workers of higher relative ability	1	0.09	0.77
Effect of pay allocation scheme For workers of lower relative ability	1	4.72	0.03

Notes:

[a] *MaxOutput* is the extent to which participants indicate agreement in the post-experimental questionnaire with the statement, "My strategy is to maximize the amount of output I produce in the anagram task," with Likert-scale responses anchored at 1 = "Do not agree at all" and 7 = "Very much agree."

[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).

[c] The *relative ability* factor differentiates the higher-ability participant within a high-heterogeneity pair from the lower-ability participant within that pair, with ability determined from the training period.

[d] All reported *p*-values in this supplemental table are two-tailed.

TABLE 10

Analysis of Effort Intensity[a] over Periods in High-Heterogeneity Pairs

Panel A: Linear Mixed Model

	df	**Chi-Square**	***p*-value[d]**
Pay allocation scheme[b]	1	6.80	< **0.01**
Period[c]	3	37.61	< 0.01
Pay allocation scheme × Period	3	5.55	**0.07**

Panel B: Simple Effects of Pay Allocation Scheme within Each Period

	df	**Chi-Square**	***p*-value[d]**
Period 1	1	0.53	**0.24**
Period 2	1	3.36	**0.04**
Period 3	1	6.30	**< 0.01**
Period 4	1	9.14	**< 0.01**

Notes:
[a] *Effort intensity* is the number of puzzles correctly solved for each 100 seconds devoted to the production task.
[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).
[c] *Period* is an ordinal variable for the number of the work period (1, 2, 3, o4 4).
[d] **Bolded** *p*-values are one-tailed, conditional on directional predictions.

TABLE 11

Analysis of Total Payoffs within High-Heterogeneity Pairs

Panel A: Means (Standard Deviations) for *Payoff*[a]

	Input-Based Pay Allocation[b]	Output-Based Pay Allocation[b]
Higher relative ability[c]	$17.19	$22.73
	($3.96)	($3.86)
	(n=13)	(n=14)
Lower relative ability[c]	$16.86	$16.82
	($4.02)	($0.28)
	(n=13)	(n=14)

Panel B: Analysis of Covariance

Factor	df	Mean Square	*F*	*p*-value[d]
Pay allocation scheme[b]	1	101.96	6.78	0.01
Relative ability[c]	1	131.85	8.77	< 0.01
Pay allocation scheme × Relative ability	1	104.86	6.97	0.01
Error	50	15.04		

Panel C: Simple Effects

	df	*F*	*p*-value[d]
Effect of pay allocation scheme for workers of higher relative ability	1	13.75	< 0.01
Effect of pay allocation scheme For workers of lower relative ability	1	0.00	0.98

Notes:
[a] *Payoff* is the total cash each participant earns across the four work periods, including payments for unused time, but excluding the $5.00 fixed show-up fee.
[b] Our experimental manipulation is whether each pair's output-generated compensation pool is allocated based on relative time spent (input-based pay allocation) or on relative puzzles solved (output-based pay allocation).
[c] In this supplemental ANCOVA, the *relative ability* factor differentiates the higher-ability participant within a high-heterogeneity pair from the lower-ability participant within that pair, with ability determined from the training period.
[d] All reported *p*-values in this supplemental table are two-tailed.