The Role of Incentives in Sustaining High-Creativity Production Over Time

Steven J. Kachelmeier
kach@mail.utexas.edu
The University of Texas at Austin

Laura W. Wang
laurawww@illinois.edu
University of Illinois at Urbana-Champaign

Michael G. Williamson
michael.williamson@mccombs.utexas.edu
The University of Texas at Austin

September 2014

Working draft. We appreciate comments from Bill Messier, Brian White, and from workshop participants at the University of Nevada at Las Vegas.

The Role of Incentives in Sustaining High-Creativity Production Over Time

Abstract: We examine the effects of performance-based compensation incentives on high-creativity production at two points in time: (1) a first-stage experiment during which the incentive manipulation is implemented, and (2) a follow-up second-stage event ten days after the removal of the first-stage incentives. We find that experimental participants receiving quantity-based pay for their first-stage production generate more high-creativity ideas *in the second stage* than do participants receiving fixed pay, even though we observe no compensation-based difference in high-creativity production in the first-stage experiment itself. These findings support the premise that, although incentives may have no immediate beneficial effect on creativity, those operating under performance-based incentives nevertheless achieve more progress on the task, thus helping them to sustain long-term creativity.

The Role of Incentives in Sustaining High-Creativity Production Over Time

I. Introduction

As reviewed by Byron and Khazanchi (2012), several studies in psychology, management, and recently accounting (e.g., Kachelmeier, Reichert, and Williamson 2008) have used experimental methods to investigate the effects of incentives on creativity. The approaches and findings of these studies vary widely (Byron and Khazanchi 2012). A common limitation, however, is that a one-shot laboratory experiment does not capture the longer-term sense of "defocused attention" that neurobehavioral research identifies as being central to sustaining creative performance over time (Dietrich 2004). We overcome this limitation while retaining the control advantages of an experimental design by measuring creative performance several days after the experiment has ended. Our primary finding is that performance-contingent incentives that have no effect on high-creativity ideas at the time of the initial experiment exhibit a positive effect on creativity ten days later, even though the compensation scheme present in the initial experiment no longer applies.

A longer-term focus on the effectiveness of performance-contingent incentives on creativity is important because this focus has the potential to challenge the common belief expressed in the psychology and management literatures that performance-based incentives are dysfunctional for tasks that require original thought. For example, a foundational study by Deci (1971) concludes that extrinsic incentives undermine individuals' intrinsic task motivation. Prendergast (1999) argues that any such short-term demotivation simply reflects the "burn-out" effect of becoming fatigued by the pressure induced by incentives in the short-term. Accordingly, a deeper understanding of the effect of incentives on creativity would seem to require a longer-term perspective. If Dietrich (2004) is correct that creativity requires sustained but defocused attention over time, then initial incentives can provide the stimulus and initial progress that motivates this sustained attention, thereby producing a long-term

benefit even if that benefit is not discernable at the time of the initial incentive. We find such an effect, thereby challenging the conventional wisdom that incentives undermine the motivation to be creative.

To test our research question, we design a two-stage experiment. In the first stage, we conduct a simple performance-based experiment that compares quantity-based incentives to fixed pay, measuring creative production from a task based on designing "rebus puzzles," similar to Kachelmeier, Reichert, and Williamson (2008) (hereafter, KRW). To this baseline, we capture a longer-term perspective by bringing participants back for a second stage ten days later, in which we ask participants strictly on a volunteer basis for any additional creative ideas they are willing to provide. Importantly, there are no additional performance-contingent incentives in this second stage, and we do not inform participants in advance of our intent to request additional creative ideas. Indeed, from the participants' perspective, the only announced reason for even coming to the second-stage location is to collect compensation for the first stage. Thus, our treatment manipulation is implemented only in the first-stage experiment, ten days before we collect our second-stage dependent variable.

Findings show that participants with quantity-based compensation in the first-stage experiment generate more ideas than their fixed-pay counterparts during the first-stage experiment. This finding replicates a similar finding in KRW and is not surprising, as it simply reflects the total quantity effect of quantity-based incentives. For the subset of ideas that meet a "high-creativity" threshold, which we define as a creativity rating of six or better on a ten-point scale, we observe no difference in high-creativity production between quantity-incentivized and fixed-pay participants in the first-stage experiment. This finding too is consistent with KRW's conclusion that quantity-based pay neither helps nor hurts high-creativity production in the short-run.

What is new in our study occurs in the second stage, in which participants arrive ten days after the first stage to collect their compensation. Before paying participants for the amounts earned in the first stage, we ask them, strictly on a volunteer basis, for any additional creative ideas they are willing to provide. The responses differ between the first-stage incentive conditions. Specifically, unlike the "dead heat" between quantity-based pay and fixed pay that we observe for high-creativity production in the first stage, participants with first-stage quantity incentives are approximately twice as creative as their fixed-pay counterparts in the second stage, as measured by the number of high-creativity puzzles submitted. This finding suggests that, relative to our fixed-pay control condition, quantity-based pay motivates a sustained interest in the creative task long after the performance-based incentives are removed.

We interpret this result as reflecting the "progress principle" (Amabile and Kramer 2010; 2011), which holds that making initial progress on a thought-intensive task is pivotal to sustaining motivation for that task over time. In our setting, even though quantity-based pay does not lead to increased high-creativity production at the time of the first-stage experiment itself, quantity incentives serve to "prime the pump" with more practice at producing ideas of varying creativity. This progress appears to sustain the longer-term attention that Dietrich (2004) argues is central to creativity, leading to relatively more high-creativity ideas ten days later.

Our results suggest important implications for the use of performance-based pay in organizations that benefit from creativity. To our knowledge, ours is the first study to show that quantity-based performance incentives can benefit high-creativity production after the incentives have been removed. This finding helps to explain why performance-based incentive schemes are common even in creativity-dependent firms (Grabner 2014). Second, our results challenge the perspective often advanced in the psychology and management

literatures that extrinsic incentives undermine intrinsic motivation for creative tasks. Even if this phenomenon can be demonstrated in the experimental laboratory in the short-term, our study suggests that a longer-term focus is essential to a more complete picture.

We review the literature and develop hypotheses in the next section, followed by describing our experimental method and design in Section III, presenting results in Section IV, and concluding in Section V.

II. Literature Review, Theory, and Hypotheses

The Effect of Performance-Contingent Incentives on Creativity

The question of how incentives influence creativity has long been of interest in psychology and management (Byron and Khazanchi 2012), and has recently been extended to management accounting, including experiments by KRW (2008), Kachelmeier and Williamson (2010), Chen, Williamson, and Zhou (2012), and a recent survey-based study by Grabner (2014).[1] A common theme from the accounting studies is that, contrary to the view often expressed in psychology and management, creativity and performance-contingent incentives can be compatible. KRW find that a simple quantity-based incentive scheme outperforms fixed pay in terms of both quantity and creativity-weighted productivity, albeit with no better (and no worse) performance in terms of *high-creativity* output. Grabner's (2014) survey-based field evidence finds that creativity-intensive firms tend to utilize a greater degree of performance-contingent incentives than do firms that depend less on creativity, although her study is not designed to measure the extent to which incentives actually improve creativity.

While these studies show that performance-based pay and creativity can be compatible, they stop short of demonstrating that performance incentives can actually

[1] Our experiment is closest in design to KRW. Kachelmeier and Williamson (2010) extend KRW to the *self-selection* effects of creativity incentives, finding that more creative people prefer creativity as an element of their compensation package and indeed are more creative initially, although quantity-based pay eventually predominates in terms of creativity-weighted productivity. Chen et al. (2012) examine incentives for *group* creativity, finding that creativity incentives work best when implemented across groups, not within groups.

enhance creativity, particularly at the highest levels of creativity. More specifically, while KRW find that a quantity-only pay scheme outperforms all other pay schemes they test in terms of "creativity-weighted productivity," it is important to note that KRW's measure of creativity-weighted productivity is linked to total quantity by construction. That is, even if one weights workers' production quantity by creativity scores, the more one produces of *any* creativity, the higher is the creativity-weighted productivity score. As the authors explain, the creativity-weighted productivity advantage enjoyed by KRW's quantity-only participants is mostly a reflection of the "volume strategy" achieved from these participants' greater overall quantity. In contrast, for *high-creativity* production, which KRW define as ideas in the top quartile of the overall distribution of creativity ratings, quantity-only incentives result in a "dead heat" with the other incentive conditions they test (including fixed pay), with no discernable treatment differences. Essentially, KRW find that quantity-based pay increases overall productivity without increasing or lowering the number of high-creativity ideas.

Given that business success often depends on highly creative ideas (Fairbank and Williams 2001; Fallon and Senn 2006), we direct this study to the open question of whether performance-based incentives can improve not just productivity in general, but *high-creativity* productivity in particular. Outside accounting, this question remains highly controversial, with strongly held positions on both sides (Byron and Khazanchi 2012). The position opposing incentives generally rests on the premise that explicit incentives undermine any intrinsic motivation to do well at the task in the absence of incentives. Although not involving creative design, a foundational example for this view is Deci's (1971) experiment investigating the effect of performance-based incentives for solving thought-intensive puzzles. Relative to participants in a no-incentive control condition, participants in Deci's incentive condition solved more puzzles during the thirteen-minute incentive period, but were less willing to work on the task for an additional eight minutes after the incentives were

removed. This and other studies reviewed by Deci, Koestner, and Ryan (1999) support the position that performance incentives reframe an otherwise interesting task as payment for aversive work, undermining performance once the incentives are removed. Amabile (1996) extends the argument to creativity, taking the position that extrinsic rewards are not just ineffective for stimulating creativity, but are actually harmful.

Taking the opposite position, Prendergast (1999) revisits the debate from the more economics-based perspective that incentives are effective and functional when used in appropriate ways. Prendergast acknowledges the accumulated evidence that incentive-based pay can decrease effort once the incentives are removed, but he offers a different interpretation. Namely, Prendergast observes that precisely because incentives are effective, they lead to greater (relative) exhaustion among those working under incentives, with the resulting fatigue leaving those participants less willing to work on the task after the incentives are removed. Returning to Deci's (1971) experiment, for example, it is possible that after working harder than the control-condition participants for thirteen minutes, the incentivized participants do not work as hard for the eight subsequent minutes without incentives simply because the additional effort they have already exerted leads them to desire some rest.

Prendergast's (1999) argument suggests that a fair test of performance-based incentives requires more than simply removing the incentives and observing the immediate consequences. Especially when extended to *creative design* tasks that go beyond Deci's (1971) and others' examinations of incentive effects in problem-solving tasks, the longer-term effects of incentives likely take days or weeks to arise, not minutes. From a neuroscientific perspective, Dietrich (2004) provides evidence from brain research that creativity requires a sustained, "defocused attention" that transitions between the conscious and subconscious mind over time. This perspective helps to explain why incentive-based

differences in creativity experiments are unlikely to arise instantly, which we address by designing a longer-term test.

A longer-term perspective on creativity incentives also builds on the "progress principle" advanced by Amabile and Kramer (2010; 2011), which holds that success in thought-intensive tasks arises in part from making meaningful short-term progress. In short, "priming the pump" with some progress leads to increased motivation, which in turn leads to more progress. Although Amabile and Kramer do not advocate the use of extrinsic rewards for creative tasks, some elements of their "progress principle" suggest a role for incentives in generating the progress that can lead to long-term creative success. For example, in a multi-year study in which workers kept diaries of one significant event each day, Amabile and Kramer (2011) find that external forces such as deadlines led to significant progress, even if stressful at the time, which then improved the workers' subsequent success. Consistent with this premise, Grabner (2014) finds from survey-based evidence that creativity-intensive firms employ *more* performance-contingent rewards than do firms that engage in more routine tasks. Relative to typical laboratory experiments, Grabner's (2014) evidence has the advantage of capturing actual firm behavior, which is inherently more long-term in nature.

Extended to a creative-design setting, the progress principle suggests that quantity-based incentives could "prime the pump" with additional initial ideas. While the creativity of these ideas is likely to vary, the increased volume provides more practice with the task. Such practice might not lead to greater creativity *initially* because it is difficult to force creative ideas on the spot (KRW 2008), in addition to the fatigue factor noted by Prendergast (1999) and the absence of the sustained, defocused attention that nurtures creativity in the brain (Dietrich 2004). Over time, however, if initial exposure to quantity-based pay simulates a greater immersion in the task, that immersion is likely to pay off later in terms of sustaining greater long-term creativity than would be exhibited by individuals receiving fixed pay.

Because the literature is characterized by competing arguments, we test two null hypotheses, one for initial creativity and the other for longer-term creativity. That is, while we are not predicting the null, we test the null against two-sided alternative theory-based arguments that performance-based incentives could either improve or worsen high-creativity production.

> **H1:** Relative to fixed pay, quantity-based compensation will have no effect on high-creativity production at the time of the initial experiment.
>
> **H2:** Relative to fixed pay, quantity-based compensation will have no effect on high-creativity production several days after the incentives are removed.

Incentivizing Creativity Directly

The arguments discussed thus far apply to quantity-based performance incentives, which reward productivity but do not explicitly reward creativity. Thus, it is natural to also ask what effect *creativity*-based incentives might have on creative production. Accordingly, as a supplemental test to shed additional insights on our primary hypotheses H1 and H2, we also test a condition with *high-creativity* incentives instead of quantity-only incentives. KRW and Kachelmeier and Williamson (2010) find that adding creativity weightings to the compensation mix *lowers* creativity-weighted productivity relative to quantity-only pay, without increasing the number of high-creativity ideas. In explaining this finding, KRW offer the reasoning that incentives to be creative are unlikely to be effective immediately, given that creativity is more likely to arise from a trial-and-error process than from fixating on the need to be creative. From the perspective of the progress principle, it would seem that the most initial progress would come simply from trying as many ideas as possible, as quantity-only incentives would reward, such that we have no *ex ante* reason to believe that high-creativity incentives would outperform quantity-only incentives in the long-run. The issue is ultimately an empirical question that we address later in the paper.

III. Method, Task, and Design

Task

We recruited 79 student volunteers from undergraduate business classes for participation in a compensated laboratory experiment. To facilitate comparability with prior research, we patterned the experimental task after the "rebus puzzle" experiments conducted by KRW (2008) and by Kachelmeier and Williamson (2010). As illustrated in Figure 1 with examples from the current study, a rebus puzzle "is a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase" (quoted from the experimental instructions). To make the task creative, participants *design* rebus puzzles rather than solve them, although we asked participants to write the solution at the bottom of each puzzle in order to facilitate creativity ratings, discussed later. For a creative production task, designing rebus puzzles has the advantage of meaningful variation in both quantity and creativity, as evidenced in prior research. We provided participants with the same instructional examples as those reproduced in the appendix to KRW (2008, 368-372), informing them (truthfully) that an independent panel of creativity raters would evaluate their submitted puzzles afterwards for creativity, "where creativity refers to puzzles that are original, innovative, and clever" (quoted from the instructions).

In the first-stage experiment, participants used a stack of blank index cards to design and submit rebus puzzles for 20 minutes, placing each puzzle in an "output box" when finished. One important departure from prior research is that KRW informed their participants in general terms that the researchers value quantity and creativity, whereas our instructions focus on the high-creativity end of the scale, informing participants that "we value the *number of high-creativity puzzles* you can construct" (emphasis in original). We then informed participants (truthfully) that, "in previous experiments using this task, approximately 15 percent of puzzles received a creativity rating at or above 6," such that "a

rating at or above 6 would be considered a high-creativity puzzle." This statement provides all participants with the common understanding that we wanted them to maximize the production of high-creativity puzzles. Using a different task, Kachelmeier, Thornock, and Williamson (2014) find that value statements of this nature can influence behavior, which for our study is consistent with focusing participants' attention on our primary dependent variable, the number of high-creativity puzzles submitted.

Design

Our experimental design involves two manipulated factors: compensation scheme and experimental stage. Our first treatment factor, manipulated between participants, is the compensation scheme participants face in the first stage, which either provides fixed pay of $25 as a control condition or bases payment on the quantity produced. In the *fixed-pay* condition, participants (n = 26) read the following:

> You will receive a fixed payment of $25.00 for constructing rebus puzzles for 20 minutes. In about two weeks, all participants with this version of the research will receive $25.00 in cash. You will not need to do anything else, and you will get $25.00 no matter what you do today. The primary reason for waiting two weeks is that different versions of the research require waiting, and we want to pay all participants at the same time.

In contrast, participants in the *quantity-pay* condition (n = 27) read the following:

> Your compensation will be based on **how many** rebus puzzles you can construct in 20 minutes, irrespective of the creativity ratings of those puzzles. That is, to determine your compensation, we will simply ***count the number of rebus puzzles you submit, no matter what creativity ratings those puzzles receive***. For the participants using this version of the research, we will determine a cash payment rate that results in **$45.00** total compensation for the participant (or participants, if tied) with the ***highest*** number of puzzles submitted, and **$5.00** for the participant (or participants, if tied) with the ***lowest*** number of puzzles submitted. Everyone else in this version of the research will receive something in between $5.00 and $45.00, depending on the number of puzzles submitted, thus resulting in an expected average compensation around **$25.00**. In other words, the more puzzles you submit, irrespective of their creativity ratings, the more money you will make. We will pay you in two weeks, after we have analyzed the results to determine the payment rate that achieves this compensation.

By determining the pay rate afterwards, we are able to set the *average* compensation in the quantity-pay condition at $25, the same payment provided to all participants in the fixed-pay condition. Thus, our compensation scheme factor manipulates the *nature* of compensation without manipulating the average *amount* of compensation. This control is important because we are interested in the incentive effects of performance-based pay as a compensation design scheme, not merely the behavioral effects of paying more or less money to experimental participants.

As mentioned above, we also implement a third condition involving a *high-creativity* pay scheme (n = 26), operationalized in a manner similar to our quantity-pay scheme, but rewarding only high-creativity puzzles (i.e., those rated six or better). We comment on this this condition as part of our supplemental analyses.

Our second experimental factor, manipulated within participants, is the *experimental stage*. The first stage is the 20-minute experimental task described above, which we use as a within-participants control for comparison to high-creativity production in the second-stage of the experiment, discussed next. Importantly, participants do not know at the time of the first-stage experiment that there will be a second stage. We simply ask participants to come to a specified location ten days later within a specified block of time to collect their compensation for the experiment. Thus, participants have no reason to continue to think about the task between the two stages other than any motivation they might have developed from the initial experiment.

An important clarification is that our experiment does not implement *long-term* incentives, such as informing incentivized participants at the time of the first-stage experiment that we would continue to apply the same incentive structure for any additional ideas provided ten days later. Although explicit long-term incentives might have provided a stronger long-term effect, the risk of such an approach is that, outside the control of the

laboratory environment, participants might have been tempted to "cheat" by accessing rebus puzzles from online sources and/or from other participants. We desired an environment in which creative ideas emerge spontaneously from the participants' ideas, not from their research or extraneous efforts. Also, from a theoretical perspective, we wanted to test the extent to which the *initial* incentive prompts the sustained, defocused attention over time that Dietrich (2004) asserts is pivotal to creativity. Accordingly, we took the conservative approach and designed the experiment to eliminate any incremental effect of explicit long-term incentives, while recognizing that long-term incentives could potentially generate a stronger effect.

Upon arrival at the specified room ten days after the initial experiment, we provided *all* participants with a document containing the following wording:

> Thank you again for participating in our experiment last week. You will receive your cash compensation for last week's session in just a few minutes. However, if you are willing to provide responses to just a few more items, we will pay you an additional $10 in cash today, on top of what you have already earned. We expect that these additional responses will require no more than 15 minutes.
>
> Presuming that you are willing to give us these few additional minutes before collecting your payment, please provide responses to the items on the next page. After you are finished, please hand these responses to the administrator, who will give you your cash compensation, including the additional $10. If you have to leave now and cannot give us a few more minutes, please see the administrator to get paid what you earned from last week's participation, but we certainly hope that you can spare just a few more minutes for an additional $10 of compensation.

As this wording indicates, we asked for the additional participation *before* providing participants with their first-stage compensation, thereby minimizing the risk of influencing quantity-pay participants in the second stage by revealing their earnings from the first stage. We also gave *all* participants the same promise of $10 additional, fixed compensation for a few more minutes of time, such that there is no incremental treatment manipulation of participant incentives in the second stage. As previously discussed, the entirety of our

between-participants incentive manipulation occurred ten days before we collected the second-stage responses. We provided $10 *fixed* pay to all participants in the second stage simply as a goodwill gesture in exchange for our request for a few extra minutes, which all participants volunteered to provide.

Immediately after the second-stage instructional excerpts quoted above, participants read the following:

> We are interested in your approach to the task now that several days have elapsed since the experiment. In your envelope are a few more blank index cards. Please use ***one*** of these cards to construct the most creative puzzle ***that now comes to mind,*** writing the solution at the bottom of the card as you did before***. Please also write "MOST CREATIVE" on the back of the card.*** We are asking everyone to submit this one "most creative" puzzle. Then, if you wish to do so, at your option you can construct ***up to*** ten additional creative rebus puzzles, printing the solution to each puzzle at the bottom of the card. You will receive $10 additional compensation no matter how many additional puzzles you submit. When completed, please place your puzzles in the envelope that contained these instructions.

Note from these instructions that our second-stage request was modest, only asking for one "most creative" puzzle from each participant, with the option to provide up to ten more creative puzzles strictly at the participants' discretion. Participants received $10 additional compensation irrespective of how many additional puzzle(s) they submitted. In short, our second-stage instructions attempt to capture the voluntary willingness that participants might have to continue with the task from ten days earlier, without providing any additional performance-based incentives. We asked participants to put any new puzzles in an envelope along with any remaining unused index cards in order to minimize any sense of pressure from the experimenters.

The second-stage materials closed with a post-experimental questionnaire, which participants submitted along with the envelope containing any new puzzles. Participants then collected their cash compensation from the first-stage experiment and the $10 additional fixed pay for the second stage.

Creativity Ratings

To construct our dependent variable, we obtained creativity ratings from four panels of eight independent raters each – three panels for the first-stage experiment and a fourth panel for the second-stage puzzles. We recruited creativity raters from undergraduate business honors classes that did not provide experimental participants, thus avoiding any overlap between the participant pool and the rater pool. Raters received fixed compensation of $50 each for their time (about 2½ hours). Creativity raters first read the same background instructions and examples as did the experimental participants, but without any information on treatment manipulations. They then used radio-frequency response devices to rate each puzzle on a 1 (lowest) to 10 (highest) scale, with puzzles projected one at a time. We needed four panels to mitigate rater fatigue, as each panel had to evaluate approximately 500 rebus puzzles. The order of puzzles was randomized for rating purposes, and raters were blind to treatment conditions. To maximize consistency across panels, we began each rating session with the same set of 40 puzzles in order to provide the same initial calibration. For the fourth session of 230 second-stage puzzles, we also added 170 first-stage puzzles (randomly selected and interspersed), in order to make the distribution of puzzles as similar as possible across rating sessions.

Because creativity ratings are subjective and noisy, we dropped the highest and lowest individual ratings for each puzzle to reduce the effect of inter-rater variability. Thus, each puzzle's rating reflects the average rating awarded to that puzzle by six of the eight panelists.

IV. Results

Descriptive Statistics and Overview of Findings

Table 1 shows descriptive statistics for the *overall* quantity and *high-creativity* quantity of rebus puzzles submitted in both experimental stages, with "high-creativity quantity" defined as puzzles receiving a creativity rating of six or better, consistent with the

definition provided in the instructions. For the first-stage experiment, participants compensated for quantity produce significantly more puzzles (average of 24.15) than do those with fixed pay (average of 13.50), with the difference significant at $p < 0.001$. This difference is not surprising, given the incentive scheme, and is consistent with the similar result reported by KRW. The production advantage gained by quantity-compensated participants does not extend to high-creativity productivity in the first stage, however, in which the number of first-stage high-creativity puzzles submitted is 2.44 and 2.73 in the quantity-based pay and fixed-pay conditions, respectively. Overall, we observe an average of 18.92 total puzzles and 2.58 high-creativity puzzles across conditions in the first stage, such that high-creativity production is 13.6 percent of total production, similar to the 15 percent estimate stated in the instructions.

Thus, for the first-stage experiment, participants operating under quantity-based pay submit significantly more puzzles in total, but about the same number of high-creativity puzzles as their fixed-pay counterparts. These first-stage findings are consistent with results reported previously by KRW. In the second-stage ten days later, participants submit only 2.34 puzzles in total, on average, which is far fewer than the average of 18.92 in the first stage. This large difference reflects the structural difference in how we implemented each stage. Specifically, from the participants' perspective, the formal "experiment" occurs only in the first stage, in which we asked participants to design rebus puzzles for 20 minutes. Participants come to the second-stage location only to collect their first-stage compensation, at which time we ask for only *one* additional puzzle from each participant in exchange for $10 additional fixed compensation. Participants then have the option to provide up to ten more second-stage puzzles, but only if they wish to do so. Predictably, many participants provided the minimum of one second-stage puzzle, collected their compensation, and left. As we report next, however, we do observe a difference in creativity in the second stage.

Although participants compensated for quantity in the first stage continue to submit more puzzles than their fixed-pay counterparts even in the second stage (2.74 vs. 1.92), the more striking comparison is that almost all of this difference is due to puzzles that meet the high-creativity threshold. Specifically, 1.44 of the 2.74 puzzles submitted by those assigned previously to the quantity-based pay condition reach the high-creativity threshold, on average, in comparison to only 0.69 of the 1.92 puzzles submitted by those assigned previously to fixed pay.

Tests of Hypotheses

To test our hypotheses, we first standardize the dependent variable by computing Z-score transformations of high-creativity production within each stage of the experiment, subtracting the overall mean high-creativity production for each stage from each corresponding observation and dividing by the overall standard deviation for each stage. Z-score transformation controls for the structural difference between the two experimental stages by removing this overall difference from our data.[2] Nevertheless, we can still evaluate the compensation scheme × stage *interaction* to test whether our primary manipulation of quantity-based vs. fixed pay has different effects across the two stages of our study.

We also add a covariate to the analysis for *openness*, extracted from one of the personality dimensions in the "Big-Five Inventory Questionnaire" (John, Donahue, and Kentle 1991). McCrae (1987) finds a significant correlation between individual creativity and tolerance for openness, which captures "an interest in varied experience for its own sake" (1987, 1259). In short, more "open" people are more receptive to new and different things. Finally, of the 240 second-stage puzzles we received, ten are duplicates of the same participants' first-stage puzzles. Submitting a duplicate is inconsistent with our research

[2] That is, after the Z-score transformation, the overall mean within each stage is constrained to be zero.

objective of assessing differences in second-stage creativity, so we delete these ten observations from all second-stage analyses.[3]

An analysis of covariance (ANCOVA) on Z-score-transformed high-creativity production with the openness covariate generates a significant compensation scheme × stage interaction (F = 5.05; two-tailed p = 0.029), as depicted in Figure 2. As this figure shows, after standardizing via the Z-score transformation to remove the overall decline in the quantity of puzzles submitted from the first stage to the second stage, quantity-based pay appears to benefit high-creativity production only in the second stage of the experiment, even though the compensation scheme itself applies only to the first stage. Below, we report statistical tests for each of our hypotheses.

For H1, Panel A of Table 2 reports no statistically significant differences in high-creativity production at the time of the initial experiment attributable to quantity-based pay or the openness covariate. Thus, we fail to reject null hypothesis H1. This finding is consistent with prior research by KRW, and likely reflects Dietrich's (2004) neuroscientific perspective that creativity does not happen instantly. Thus, although participants paid for quantity respond to this incentive by submitting significantly more first-stage puzzles do than their fixed-pay counterparts, they are unable to transform this initial advantage into a greater number of first-stage *high-creativity* puzzles.

For H2, we obtain a different result. As Figure 2 depicts and Panel B of Table 2 tabulates, participants paid for quantity in the first stage of the experiment generate significantly more high-creativity puzzles *in the second stage* than do their fixed-pay counterparts (F = 6.53; two-tailed p = 0.014), even though the quantity-based incentive no longer applies in the second stage. Thus, we reject null hypothesis H2 in favor of the alternative hypothesis that, relative to fixed pay, quantity-based pay improves high-creativity

[3] Our research conclusions do not depend on deleting the ten second-stage duplicates.

production long after the incentive is removed. We also see a significant second-stage effect for openness ($F = 5.79$; $p = 0.014$).

Our findings for H2 pose a challenge to the view often expressed in psychology and management that incentives undermine creativity. Consistent with Prendergast's (1999) conjecture that evidence of the dysfunctional effects of incentives in experiments may simply reflect the fact that incentivized participants get fatigued in a short-term task, it may well be that performance-based incentives are of little creative benefit in short-term settings measured at the time that the incentives are introduced. What we find ten days later, however, is that participants initially paid for quantity are more successful than their fixed-pay counterparts in sustaining high-creativity production over time. This finding is consistent with the "progress principle" (Amabile and Kramer 2010; 2011) that meaningful experience towards a goal, even if that goal is not attained initially, helps to further long-term success. In our study, quantity-compensated participants generated significantly more puzzles than did fixed-pay participants in the first stage of the experiment. These additional puzzles *did not* translate to additional high-creativity puzzles in the first-stage, but they likely served to "prime the pump" with additional meaningful experience that influenced participants after the experiment ended. Ten days later, these quantity-incentivized participants voluntarily submitted not just a higher number of second-stage puzzles, but also puzzles that were more creative than those submitted by fixed-pay participants.

Supplemental Analyses

Alternative Measure of Second-Stage High-Creativity Production

Recall that our second-stage instructional materials request only one "most creative" puzzle, with the option to submit up to ten more puzzles "if you wish to do so." Predictably, given these instructions, the modal participant in stage two submitted only one puzzle, which sometimes met the high-creativity threshold and sometimes not. This structure suggests a

natural division of our second-stage responses into three categories: (1) those who submitted no high-creativity puzzles, (2) those who submitted one high-creativity puzzle, and (3) those who submitted two or more high-creativity puzzles. Coding these three categories as 0, 1, and 2, respectively, in an ordered logistic regression supports the conclusion of greater second-stage creativity among those compensated for quantity in the first stage (two-tailed $p = 0.046$, untabulated). Eight of the 27 participants (30 percent) in the quantity-based pay condition fall into the highest category of two or more second-stage high-creativity puzzles, in contrast to only two of the 26 participants (8 percent) in the fixed-pay condition. Importantly, if we apply the same categorical test in the first-stage experiment, we *do not* observe a significant difference ($p = 0.205$, in the opposite direction), supporting our general conclusion that performance-based incentives can lead to *sustained* creativity even if they have no effect on creativity initially.

Change in Creativity

As noted previously, our second-stage setting is structured less formally than our first-stage experiment. Indeed, it is likely that participants do not even view the second stage as a continuation of the experiment, given that they come to the second-stage location simply to collect their first-stage compensation. Participants submit far fewer puzzles overall in the second stage than in the first, such that from the simple aspect of total volume, the second stage offers fewer opportunities to reach the high-creativity threshold. Given this environment, it is interesting to examine the best puzzle submitted by each participant in each stage. We find that 52 percent (28 percent) of the participants in the quantity-compensated condition (fixed-pay condition) submit a best puzzle in the second stage that is of equal or greater creativity than their best puzzle from the first stage. These percentages significantly differ in a logistic analysis that controls for the openness measure discussed earlier as a covariate (two-tailed $p = 0.05$). We conclude that quantity-compensated participants were

more successful than their fixed pay counterparts in sustaining high creativity from the first to the second stage.

Mediation Analysis

The "progress principle" that we draw upon as the theoretical basis for our findings rests on the notion that people have to experience meaningful progress towards a goal in order to achieve more success meeting that goal. For our task, we observe no difference in high-creativity production between quantity-compensated and fixed-pay participants in the first-stage experiment. Accordingly, the differential progress between the two incentive conditions likely arises from the increased first-stage production in the quantity-compensated condition of "reasonable" puzzles that were close to but did not quite reach the high-creativity threshold. To capture this progress, we extract the number of "moderate-creativity" puzzles generated by each participant in the first-stage experiment, defining this category as puzzles attaining a rating greater than four but less than six. Reflecting their overall greater volume, quantity-compensated participants enjoy a significant first-stage advantage over their fixed-pay counterparts in moderate-creativity production ($p = 0.003$).

We then employ the first-stage moderate-creativity count in a bootstrapping-based mediation analysis, as recommended by MacKinnon et al. (2002) and explained procedurally by Preacher and Hayes (2004; 2008). Based on the results of 1,000 bootstrapped samples, this procedure supports the conclusion that first-stage quantity-based pay exerts an indirect effect on second-stage high-creativity production that is mediated by first-stage moderate-creativity production ($p < 0.05$). Importantly, we do not obtain the same conclusion from an alternative mediation variable based on the number of first-stage *low-creativity* puzzles (i.e., those rated lower than four), for which the bootstrapping analysis does not support a mediating effect. Thus, it would appear that "progress" for our task hinges on experiencing some success in the first stage that at least gets close to the high-creativity threshold. With this progress,

participants are more likely to sustain high-creativity production after the incentives have been removed. Because quantity-compensated participants achieve more first-stage progress at the moderate-creativity level, they sustain more high-creativity success in the second stage.

Time Spent Thinking About the Task Between the First and Second Stages

Deci (1971) concludes that extrinsic incentives undermine intrinsic task motivation based on the finding that participants with incentives spent significantly less time on the task immediately after the incentives were removed. Prendergast (1999), in turn, conjectures that one reason for this finding might simply be that incentives led to a short-term "burn out" effect, suggesting that a fairer assessment of the willingness to spend time on the task would require a longer-term perspective. To shed some insights towards this end, we asked participants in a post-experimental questionnaire administered after the second-stage task to self-report the number of minutes spent thinking about rebus puzzles over the days after the first-stage experiment. These self-reports exhibit high variance, and are at best a noisy proxy for the actual time spent. Nevertheless, we find at least marginally significant evidence that participants with quantity-based pay in the first stage self-report spending more time thinking about rebus puzzles in the days following the first-stage experiment than do participants with fixed pay in the first stage (two-tailed $p = 0.085$, using a rank-transformed measure of time spent to reduce the variance of responses).

We do not wish to overinterpret this finding, as we are unable to find significant evidence that self-reported time spent mediates our primary findings, nor do we even find systematic evidence that self-reported time spent thinking about rebus puzzles after the first-stage experiment is correlated with second-stage high-creativity production.[4] As noted, a self-reported measure of time spent from a post-experimental questionnaire is a noisy proxy at

[4] One exception is that, if we split the sample on self-reported task familiarity, we find a marginally significant positive correlation between self-reported time spent thinking about rebus puzzles and second-stage high-creativity puzzles among the subsample that falls below the median on familiarity with rebus puzzles (two-tailed $p = 0.069$).

best, and accordingly it does not lend itself well to rigorous follow-up analyses. At a minimum, however, we can say that Deci's (1971) finding of *less* time spent on the task by incentivized participants after the removal of the incentives does not appear to extend to our longer-term setting, in which case the direction reverses if anything.

Supplemental Treatment Condition: High-Creativity Incentives

Insofar as our primary dependent variable is high-creativity production, it is natural to question whether explicit incentives for high-creativity ideas would exhibit different effects from those we observe for quantity-only incentives. Intuition might suggest that the best way to motivate high creativity is to incentivize high creativity, but there is reason to expect otherwise, as KRW find that adding creativity to the incentive mix fixates individuals too much on being creative rather than being productive. Because it is difficult to force creativity, such fixation leads KRW's creativity-compensated workers to be less productive overall but no more successful in high-creativity production than those paid for quantity only. Extended to our interest in the progress principle, if participants compensated for high-creativity puzzles are less productive in the first stage than those compensated for overall quantity, they would experience less initial progress, and hence would have less to draw on for sustained creativity ten days later.

Consistent with this reasoning, we test a supplemental "high-creativity" treatment condition (n = 26) in which we compensate participants in a manner similar to the quantity-based condition in the first-stage experiment but reward only those puzzles rated six or higher. Participants in this condition produce a first-stage average total quantity of 14 that is nearly identical to that observed in the fixed-pay condition, while significantly lower than the first-stage quantity of 24 generated by participants in the quantity-only incentive condition. The number of first-stage high-creativity puzzles is nearly identical across all three conditions (average of 2.5). In the second stage, participants compensated for high-creativity puzzles in

the first stage achieve second-stage high-creativity production that lies in between the quantity-based and fixed-pay conditions. Specifically, first-stage high-creativity-compensated participants produce an average of 0.92 high-creativity puzzles in the second stage. This average is more than the 0.69 produced in the second stage by participants with fixed pay, but it is less than the 1.44 produced by participants compensated for quantity only. Neither of these comparisons with the supplemental "high-creativity" incentive condition is statistically significant at conventional levels, although, as reported earlier, the fixed-pay and quantity-based incentive conditions significantly differ from each other.

On balance, it seems safe to conclude that, if the goal is sustained creativity over time, initial incentives for high-creativity production fare no better than incentives for total quantity. From the perspective of the progress principle, quantity-based compensation stimulates the greatest initial experience (in the sense of volume) from which workers can draw on later. It is interesting to contemplate this reasoning in the context of Grabner's (2014) recent survey-based evidence on the use of performance-based incentives among creativity-intensive firms in practice. Grabner finds that creative intensity leads firms to be *more* inclined to use performance-based compensation, although her survey does not specify the exact kind of performance that performance-based compensation is rewarding. We suggest as an avenue for future field-based research that it would be interesting to extend Grabner's (2014) findings to explore whether creativity-intensive firms are explicitly rewarding *creativity*, as opposed to rewarding other dimensions of productivity that generate more progress towards high-creativity production even if they do not reward creativity directly.

V. Conclusions

Management accountants have long been interested in the effects of performance-based incentives. Only recently, however, has this interest extended beyond routine

production exercises to the softer dimensions of productive quality, such as creativity. Creativity is of interest not only because of its importance to contemporary business, but also because creativity does not necessarily arise from immediate effort. This characteristic poses a significant challenge to laboratory experimentation, as experiments have the advantage of controlled, *ceteris paribus* manipulation of treatment factors, but also carry the disadvantage of a short-term focus. We mitigate this disadvantage by collecting additional data on creativity ten days after a laboratory experiment. Our primary finding is that performance-based compensation that rewards quantity in a first-stage experiment has no effect on high-creativity production at that time, relative to fixed pay. Ten days later, however, those paid for quantity in the initial experiment voluntarily submit significantly more ideas that meet a high-creativity rating threshold than do those with fixed pay. Indeed, our fixed-pay participants generally "gravitate to the bottom" in terms of providing minimal second-stage production and creativity. These findings highlight the importance of capturing creativity in longer-term settings.

Our findings pose a challenge to the view often expressed in the psychology and management literatures that explicit incentives are harmful for creative and other thought-intensive tasks. As Prendergast (1999) observes in defending the economic perspective that appropriate incentives are functional and effective, the dysfunctional effects of incentives observed in past laboratory studies may simply reflect the fact that incentivized participants get tired and fatigued. Thus, while those operating under performance-based incentives might do worse than those with fixed pay in the minutes immediately after a laboratory task, such as in the seminal study by Deci (1971), we direct our study to what happens days after the first-stage experiment is over. Our longer-term perspective aligns more closely with a neurobehavioral perspective on creativity, which holds that creativity requires a sustained, defocused attention over time (Dietrich 2004). We test the potential for performance-based

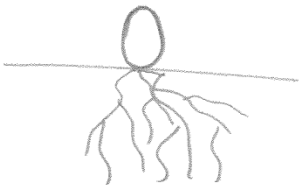
incentives to serve as the catalyst that seeds the initial progress necessary to motivate this longer-term sustained attention.

We encourage further research on the effects of incentive-compensation schemes on creativity and other performance dimensions that characterize an idea-based economy. Ideally, this research can draw from multiple methods, as is the case with recent experimental and field-based accounting research on creativity. What our study adds to the mix is a longer perspective than that captured in most laboratory experiments. That being said, our study is limited by the fact that our second-stage collection of data occurs in a setting in which participants have limited time and no advance knowledge of our request for additional participation. These features enable us to isolate the long-term effects of initial incentives that arise without any additional long-term incentives or anticipation of long-term requests. Future research can expand on our baseline by incorporating explicit long-term incentive and feedback structures that facilitate learning and practice along with the basic incentive effects we identify.

Figure 1
Examples of Second-Stage Rebus Puzzles Submitted by Participants

Panel A: High-Creativity Puzzles

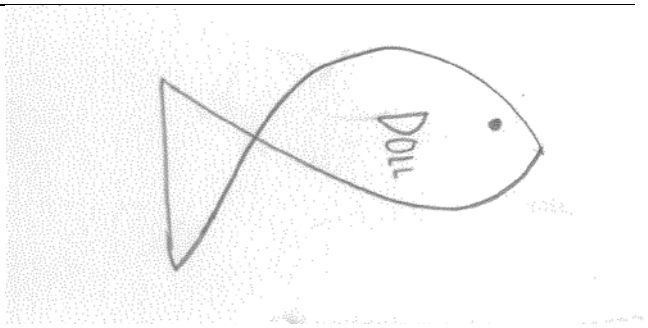
Solution: Obama
Creativity rating: 8.67
Condition: Fixed pay



Solution: Eggplant
Creativity rating: 7.83
Condition: Quantity-based pay

Panel B: Borderline High-Creativity Puzzles

Solution: Widespread panic
Creativity rating: 6.00
Condition: Fixed pay



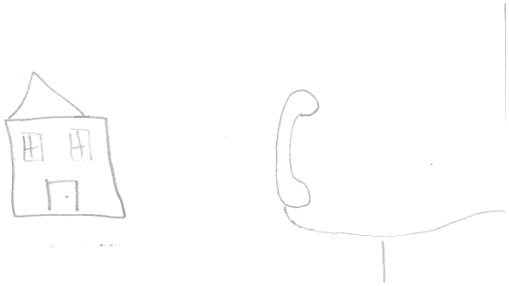
Solution: Dolphin
Creativity rating: 6.00
Condition: Quantity-based pay

Panel C: Moderate-Creativity Puzzles

Panel E

Solution: Cover-up
Creativity rating: 4.17
Condition: Fixed pay

Panel F



Solution: House Call
Creativity rating: 4.00
Condition: Quantity-based pay

Figure 2
Standardized High-Creativity Production by Compensation Scheme and Stage

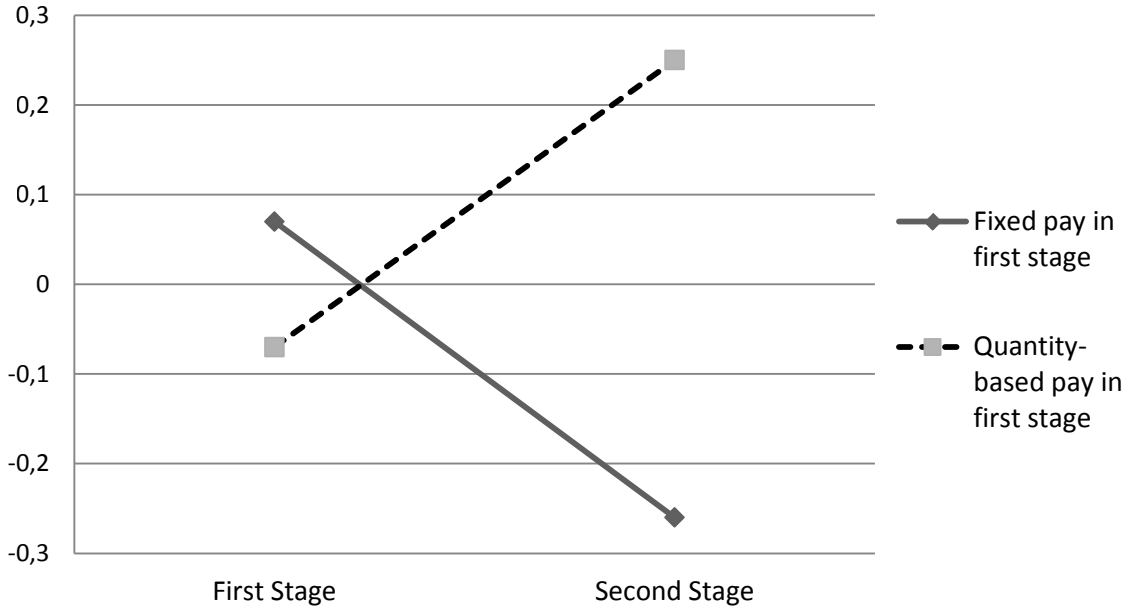


Table 1
Descriptive Statistics for Rebus-Puzzle Production: Means and (Standard Deviations)

	Fixed Pay in First Stage	Quantity-Based Pay in First Stage	Overall
Number of participants	26	27	53
Panel A: First-stage experiment			
Total quantity	13.50 (4.93)	24.15 (12.21)	18.92 (10.73)
High-creativity puzzles	2.73 (1.73)	2.44 (2.38)	2.58 (2.07)
Z-Score transformation of high-creativity puzzles[a]	0.07 (0.84)	-0.07 (1.15)	0.00 (1.00)
Panel B: Second-stage follow-up			
Total quantity	1.92 (2.30)	2.74 (2.90)	2.34 (2.63)
High-creativity puzzles	0.69 (0.88)	1.44 (1.85)	1.08 (1.49)
Z-Score transformation of high-creativity puzzles[a]	-0.26 (0.59)	0.25 (1.24)	0.00 (1.00)

[a] The Z-score transformation of high-creativity production adjusts for the structural differences between the two experimental stages by subtracting the overall mean and dividing by the overall standard deviation within each stage. Removing the main effect of stage in this manner and standardizing for the different variances across stages allows direct comparison of the compensation conditions across stages.

Table 2
Tests of Hypotheses

	df	MSE	*F*	*p*-value
Test of H1: First-stage high-creativity production (after Z-score transformation):				
Compensation scheme	1	0.18	0.18	0.675
Openness covariate	1	0.06	0.06	0.810
Error term	50	1.05		
Test of H2: Second-stage high-creativity production (after Z-score transformation):				
Compensation scheme	1	5.12	6.53	0.014
Openness covariate	1	5.79	6.67	0.012
Error term	50	0.87		

References

Amabile, T. M. 1996. *Creativity in Context: Update to the Social Psychology of Creativity.* Boulder, CO: Westview Press.

Amabile, T. M., and S. J. Kramer. 2010. What really motivates workers. *Harvard Business Review* 88 (1): 44-45.

Amabile, T. M., and S. J. Kramer. 2011. *The Progress Principle: Using Small Wins to Ignite Joy, Engagement, and Creativity at Work.* Boston, MA: Harvard Business Review Press.

Byron, K., and S. Khazanchi. 2012. Rewards and creative performance: A meta-analytic test of theoretically derived hypotheses. *Psychological Bulletin* 138 (4): 809-830.

Chen, C. X., M. G. Williamson, and F. H. Zhou. 2012. Reward system design and group creativity: An experimental investigation. *The Accounting Review* 87 (November): 1885-1911.

Deci, E. L. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18 (1): 105-115.

Deci, E. L., R. Koestner, and R. M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6): 627-668.

Dietrich, A. 2004. The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review* 11 (6): 1011-1026.

Fairbank, J. F., and S. D. Williams. 2001. Motivating creativity and enhancing innovation through employee suggetion system technology. *Creativity and Innovation Management* 10 (2): 68-74.

Fallon, P., and F. Senn. 2006. *Juicing the Orange: How to Turn Creativity into a Powerful Business Advantage.* Boston, MA: Harvard Business School Press.

Grabner, I. 2014. Incentive system design in creativity-dependent firms. *The Accounting Review* 89 (September): in press.

Hannan, R. L. 2005. The combined effect of wages and firm profit on employee effort. *The Accounting Review* 80 (January): 167-188.

John, O. P., E. M. Donahue, and R. L. Kentle. 1991. *The Big Five Inventory – Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kachelmeier, S. J., B. E. Reichert, and M. G. Williamson. 2008. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46 (May): 341-373.

Kachelmeier, S. J., T. A. Thornock, and M. G. Williamson. 2014. Communicated values as informal controls: Gaining accuracy while undermining productivity? *Contemporary Accounting Research*, in press.

Kachelmeier, S. J., and M. G. Williamson. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review* 85 (September): 1669-1691.

MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 7: 83-104.

McCrae, R. R. 1987. Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology* 52(6): 1258-1265.

McGraw, K. 1978. The detrimental effects of reward on performance: a literature review and a prediction model. In *The Hidden Costs of Reward*, edited by M. Lepper and D. Greene. Hillsdale, NJ: Lawrence Erlbaum Associates.

Preacher, K. J., and A. F. Hayes. 2004. SPSS and SAS procedures for estimating indirect effect in simple mediation models. *Behavior Research Methods, Instruments & Computers* 36: 717-731.

Preacher, K. J., and A. F. Hayes. 2008. Contemporary approaches to assessing mediation in communication research. In *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research*, edited by A. F. Hayes, M. D. Slater, and L. B. Snyder, 13-54. Thousand Oaks, CA: Sage Publications.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37(1): 7-63.