Why do goal-based incentives cause cheating? Unpacking the confounding effects of goals, social comparisons and pay

Matthew Chao, Williams College
Ian Larkin, UCLA Anderson School of Management

This version: August 20, 2017

Abstract

Recent studies suggest that goal-based incentive systems cause cheating. However, goal-based incentives comprise a number of distinct elements, including statement of a goal, the goal's justification and framing, and rewards for meeting the goal. Previous experimental research has simultaneously varied several of these elements within a single condition, leading to interpretational confounds in assessing the impact of goal assignment on cheating. We use a carefully-designed 2x2x2 experiment that isolated cheating caused by assigning a goal, by providing peer comparison justifications, and by paying for performance. This design measures the individual impact of each element on cheating, and any possible interaction effects. We find that only pay-for-performance and peer comparison framing increased cheating, while the mere assignment of goals (without these other elements) did not. Our results suggest that care is necessary when examining the impact of incentive system elements on cheating, and that mere goals themselves do not cause cheating.

Introduction

Goal-based incentives are used by many firms to motivate their employees (e.g., Larkin 2014; Latham and Baldes 1975; Latham and Kinne 1974). A wide range of studies have shown that setting difficult but attainable goals can increase worker effort and persistence on a large variety of tasks (for summaries, see Locke et al. 1981; Locke and Latham 1990, 2002). As a result, goal-setting advocates claim that "goal-setting can be used effectively in any domain where an individual or group has some control over the outcomes" (Locke and Latham 2006).

However, some scholars warn that goal-setting should not be viewed as a widespread solution for improving performance, as it can also induce unwanted consequences (Staw and Boettger 1990, Ordóñez et al. 2009a, 2009b; Ordóñez and Welsh 2015). Recent laboratory studies showed that goal-based incentive systems can increase worker cheating on a task (Schweitzer et al. 2004; Cadsby et al. 2010). These experiments demonstrated that subjects were more likely to cheat on a task when they were incentivized using a goal-based structure versus a do-your-best structure. As a result, using goal-based incentives could lead to significant organizational losses, such as employee "gaming" of incentive systems (Larkin 2014) or outright deception (Edelman and Larkin 2015).

Advocates of goal-setting counter that these criticisms falsely ascribe some of the downsides of financial incentives to goals (Locke and Latham 2009), since the aforementioned experiments did not equate financial returns to cheating across conditions. In particular, in one study the marginal financial return to cheating in the goal condition could be as much as nine times higher than in other conditions using pay-for-performance incentives (Cadsby et al. 2010). In another study, subjects in some goal-based conditions were in total paid two-thirds less than subjects in non-goal conditions (unless they decided to cheat), leading to substantial differences in absolute pay across conditions (Schweitzer et al. 2004). Thus, the observed effects of goals in these studies are potentially confounded with effects of these differences in financial incentives.

Some studies also confounded their results by simultaneously varying another common element of goal-based systems: the framing of the goal, which is used to increase goal commitment. Goal commitment is a key modulator for any effect of goals on performance, especially when tasks are difficult (Seijts and Latham 2000; Klein et al. 1999); to increase subject commitment to the goal, experimental studies on goals (e.g., Schweitzer et al. 2004) often instructed subjects that many of their peers were able to attain the goal, or other similar peer comparison language. However, subjects in non-goal conditions were never given any information on peer performance, since there was no goal to justify. This therefore makes peer comparison framing a potential confound; it could be that the introduction of an unfavorable peer comparison, rather than the introduction of a goal, caused the increased incidence of cheating,

becasuse unfavorable peer comparisons have been shown to increase deception and cheating (e.g., Moran and Schweitzer 2008; John et al. 2014; Edelman and Larkin 2015).

In short, existing experimental research has identified the role of goal-based incentive *systems* on cheating, but without separately examining the effects of individual *elements* of the incentive system. As a result, scholars have conclusively shown that goal-based systems cause cheating, but not which specific element (such as goal assignment) caused cheating. An incentive system involves many independent elements, including the *structure* of the incentives (e.g., goal-based or not), the *framing* of the incentives (e.g., peer comparisons or not), and the actual *pay scheme* itself (e.g., pay-for-performance vs. fixed pay). These potential confounds underscore the fact that incentive systems are by nature complex. Comparing two inherently different incentive systems, such as goal-based versus piece rate incentives, requires understanding how differences in each of these independent elements can influence behavior.

In this study, we use a carefully-designed experiment to isolate the effects of pay, goals, and peer comparisons on cheating. Each of these three elements is commonly used in many goal-based incentive systems (Ordóñez and Welsh 2015). The experiment was designed so that the base pay, the use of peer comparison framing, the marginal financial returns for task performance, and the marginal financial returns for cheating were all the exact same across treatments unless that specific element was the one being varied in a given treatment. Unlike previous research, we can exactly compare how the addition of one element, such as peer comparison framing to a goal-based system, affects cheating, without simultaneously altering any other element. We also test for interaction effects between these elements by running treatments that vary two or all three of these elements. This allows us to test whether these main effects are additive for incentive systems that utilize multiple elements simultaneously.

Our study finds that cheating in goal-based systems occurs due to financial rewards to cheating and to social comparison framing. Assigning a goal on its own without increased pay or social comparison framing did *not* lead to an increase in cheating relative to a do-your-best counterfactual. However, the use of either financial rewards or social comparison framing led to cheating. In addition, adding a goal to financial rewards or social comparisons did not increase cheating beyond the main effects of the financial rewards or social comparisons. Interestingly, we find no evidence of interaction effects across the three elements, suggesting a hard "cheat or not" trigger seems to exist for most subjects who chose to cheat.

Our study has three main contributions. First, it provides a methodological contribution for research on incentive systems by introducing an experimental design that allows direct comparison of system elements in a way not seen in previous work. Second, it finds that although goal-based *systems* lead to cheating, it is not the use of goals itself that generates cheating but rather the use of other incentive elements – pay for performance and framing or justification of the goal – that are commonly found in

goal-based systems. Finally, much of the current literature on social comparisons and unethical behavior has focused on unequal pay, but our study demonstrates that a simple social comparison reference point can also cause cheating, even absent a link between performance and pay. Altogether, we hope this study can clarify some of the issues discussed in the ongoing debate among scholars over the consequences of goal-based incentives (Locke and Latham 2009; Ordóñez et al. 2009a and 2009b), which we believe has been fueled in part by the inability to separately identify the precise causes of cheating in previous laboratory studies of goal-based incentive systems.

Literature Review & Theoretical Framework

In the last 40 years, research on goals has been one of the most prominent and influential literatures on compensation systems in organizational behavior (Ordóñez and Wu 2013). The most widely cited summary of this literature states that "goals have been shown to increase performance on well over 100 different tasks involving more than 40,000 participants in at least eight countries" (Locke and Latham 2002, p. 714). In the 15 years since this summary was published, hundreds of other studies on goals have demonstrated how and why goal-based systems can so often lead to higher levels of personal and organizational performance (Ordóñez and Wu 2013). Goals have been shown to increase performance through many mechanisms: among others, they energize and motivate, increase task persistence, redirect effort towards valuable tasks, and promote task-relevant knowledge building (Locke and Latham 2002). There are also multiple documented moderating factors: among others, goals are more effective when actors are more committed to them and believe they are more important, when goals are more difficult to achieve while still being attainable, and when tasks are more complex (Locke and Latham 2002).

A more recent, smaller, but quickly growing literature shows that goal-based systems also can also carry significant downsides, particularly in that they can motivate cheating and unethical behavior. The first such study was Schweitzer et al. (2004), which showed that the use of a goal-based incentive system led to higher levels of inaccurate self-reporting of performance compared to a "do-your-best" alternative. Interestingly, cheating in goal-based treatments was similar regardless of whether financial rewards for the goal were present. At least 30 other subsequent studies in the laboratory and field have shown a link between goal-based incentive systems and unethical behavior such as cheating (Ordóñez and Welsh 2015).

However, most of these studies, as well as the literature reviews summarizing the research, use the term "goal" in two distinct ways: first as an abbreviation of an "incentive system that gives an explicit target for participants to reach," and second as the target itself. This problem appears particularly acute in the more recent literature on goals and deception; for example, a prominent recent review of this literature explicitly states that it included "terms such as expectations (e.g., analyst expectations), targets,

aspiration, quotas, status quo and reference points that have an impact on individual and organizational behavior" (Ordóñez and Welsh 2015, p. 63). Critics of the literature on goals and deception, on the other hand, point out that by using such a broad definition of "goals," the deception literature may ascribe problems stemming from non-goal elements of a given system to goals (Locke and Latham 2009).

As such, while the existing literature clearly shows a causal link between goal-based systems and cheating, it does not definitively show whether goals themselves, or some other systemic element, lead to the increased cheating. For example, the initial Schweitzer et al. (2004) study contains three experimental conditions but varies multiple elements of the incentive system across conditions. In its "mere goal" and "reward goal" treatments, subjects were explicitly given a goal (of unscrambling 9 words in one minute) and told that "many students were able to create 9 or more words per round" in previous studies. In the "do-your-best" condition, subjects were not given a goal, nor were they given the social comparison information. The study therefore cannot differentiate whether the goal communication or the social comparison framing caused the observed cheating, or whether both were required.

Another study on goal-based systems examined cheating across four types of compensation systems – do your best, goals, piece rates, and tournaments – and found that goals had the highest incidence of cheating by a wide margin (Cadsby and Song 2010). However, in this study subjects in the goal condition were given not only a stated goal but also highly different financial returns to cheating. In the "goal" condition, subjects received $3.60 for cheating on a question that caused them to meet the goal (making the marginal return to cheating as high as $3.60 if they are one question short of the goal), while in the piece rate scheme they only received $0.40 per question; their expected financial return from cheating in the tournament scheme was approximately $0.30.[1] It is highly plausible that the much larger financial return to cheating led to increased prevalence of cheating, and not the communication of the goal itself.

There are many other laboratory and field studies linking various elements of goal-based systems to cheating (Ordóñez and Welsh 2015), but the current research on system-level antecedents of cheating has not revealed the exact reasons goals cause unethical behavior. It is important to be precise about how elements of goal-based systems cause cheating since the antecedents of cheating in organizations are often so subtle and indirect (Gino 2015).

Of course, it is impossible to fully test the separate effect of every commonly-used element in goal-based systems. Incentive systems are by nature complex; organizational scholars (Jensen 1998) define incentive systems as comprising the delegation of authority, the measurement of performance on

[1] Schweitzer et al. (2004) also has a potential confound from financial rewards since subjects faced unequal monetary returns for baseline performance across conditions. Subjects in the "do-your-best" and "mere goal" conditions were paid a fixed $10, while subjects in the "reward goal" treatment would be paid less than $4 on average if they chose to accurately report their performance. If subjects had ex-ante expectations about fair pay in an experiment, they might have cheated in the "reward goal" condition simply to meet these expectations. Alternatively, they may have foreseen that the study was about cheating given the very low amount of earned pay.

delegated tasks, and the subsequent allocation of rewards and punishments. Variations in any of these "legs" can potentially alter both positive and negative behavior (see Gerhart et al. 2009 for a summary). In this study, we focus on resolving the confounds from earlier studies by explicitly examining the differential impact of assigning a goal, using social comparison frames, and offering financial rewards.

Merely assigning a goal to an employee could lead to unethical behavior in a variety of ways (Ordóñez and Welsh 2015). First, just as goals direct employees towards desired behavior at the expense of non-productive behavior (Locke et al. 1981), goals may direct employees away from desirable behavior (and towards undesirable behavior) that is not part of the goal (Barsky 2008). Second, although challenging goals are better at increasing performance than goals that are easy to achieve (Erez and Zidon 1984; Crossley et al. 2013), they may also lead to higher rates of ego depletion and thereby cheating (Welsh and Ordóñez 2014; Mead et al. 2009). Third, establishing a goal sets a reference point for employees, and employees who fall short of the reference point may end up in the loss portion of their utility curve (Heath et al. 1999); they may therefore be more risk-seeking in preferences (Larrick et al. 2009), which can lead to unethical behavior (Gino and Margolis 2011). We therefore hypothesize:

> H1: Communicating a goal to an employee will lead to greater incidence of cheating, even absent social comparison framing or compensation differentials.

There is a large literature linking financial incentives to cheating and other deceptive behavior in lab experiments (Camerer et al. 1999) and field studies (Zitzewitz 2012). For instance, bonuses, such as those often associated with goals, have been shown to lead to deception (Larkin 2014). These types of effects extend to a diverse range of occupations including CEOs (Zhang et al. 2008; Singh and Larkin 2015), athletes (Wolfers 2006) and teachers (Jacob and Levitt 2003). Research has also investigated multiple psychological mechanisms behind the link between financial incentives and cheating. For instance, receiving relatively lower absolute pay than deemed fair can lead to deception (Greenberg 1993). In addition, financial gains can skew the balance between a desire for personal gain and a desire to maintain a positive self-concept (Mazar et al. 2008). When financial rewards are introduced or increased, it increases the personal gain that can result from deception, leading self-concept maintenance to matter less relative to personal gain. Moreover, motivated reasoning and "hypermotivation" can help individuals justify their deception, thus allowing them to obtain the personal gain from the financial incentives without sacrificing their self-concept or self-image (Rick and Loewenstein 2008). This leads to our second hypothesis:

> H2: Paying for performance will lead to greater incidence of cheating, even absent an explicit goal or social comparison framing.

While the initial literature on social comparisons is well established (Festinger 1954), the empirical literature on the consequences of unfavorable social comparisons is relatively new but is growing rapidly. Field experiments demonstrate that peer comparisons alone can influence task performance (Mas and Moretti 2009; Barankay 2010; Blanes i Vidal and Nossol 2011). Research suggests that this occurs because peer comparison information can lead to increased self-efficacy (Stajkovic and Luthans 1998, Larkin 2011) and higher individual effort on a task (Cohn et al. 2014), leading to improved performance. However, unfavorable social comparisons have been shown to lead to deception in the lab (Moran and Schweitzer 2008) and the field (Edelman and Larkin 2015), in part because they invoke negative emotions (Festinger 1954), such as resentment (Weiner 1986), low self-esteem (Tesser 1988), and envy (Feather 1989). However, this research is not specific to compensation incentives; existing research on unfavorable social comparisons in compensation has largely focused on comparisons about pay rather than performance. This literature shows that subjects or employees who know they are being paid less than referent others are more likely to feel dissatisfied with their job and engage in negative behavior such as deception (John et al. 2014; Card et al. 2012). Although pay is one important metric by which employees compare themselves (O'Reilly et al. 1988), research in psychology demonstrates that social comparisons about performance can also be key to employees; specifically, they can negatively affect self-esteem and job satisfaction (Bandura and Jourden 1991). We therefore hypothesize that:

> H3: The use of social comparison framing will increase the likelihood of cheating by subjects who perform lower than the communicated level, even absent an explicit goal or financial incentives.

Of course, incentive systems commonly weave several of these elements together. Since we hypothesize that each of these individual elements causes a psychological strain on employees, their use in combination should lead to even higher cheating than any one element alone. In particular, adding monetary and psychological incentives together will further skew the balance between the rewards for cheating and the desire to maintain a positive self-concept (Mazar et al. 2008). Our final hypothesis is therefore:

> H4: The use of any combination of goal setting, social comparison framing, and paying for performance will increase cheating compared to the use of any one element by itself; additionally, the use of all three elements at the same time leads to the highest amount of cheating.

Experimental design

Overview

To test these hypotheses, we designed an experiment that can conclusively isolate the individual effects on cheating of goal setting, social comparison framing, and pay for performance, as well as their

interactions. To accomplish this, our design equalized marginal pay, the marginal financial return to cheating, and expected total pay across conditions. In each condition of the experiment, the design varied only one of the three elements relative to the relevant comparison condition.

Experimental tasks

Our study asked subjects to solve a series of "word jumble puzzles," which are commonly found in Sunday newspapers. Each word jumble requires unscrambling a set of four words, which then reveals scrambled letters to a fifth word that must also be unscrambled. This last unscrambled word is the answer to the puzzle. Regardless of how many of the initial four words a subject successfully unscrambled, the subject was only credited with solving the puzzle if the fifth word was unscrambled successfully. An example puzzle, and the instructions that we gave subjects for this task, is shown in Figure 1.[2] Subjects were given eight minutes in a round to solve up to five jumbles puzzles. As is common in studies of cheating, we later asked subjects to grade their own work. However, subjects did not know they would be doing so when solving the puzzles.

We chose this task, which we term the "Jumble Task," for several reasons. First, it is similar to the anagram task used in the previous literature on goals and unethical behavior (Locke and Latham 1990; Schweitzer et al. 2004; Cadsby et al. 2010), since both tasks involve unscrambling words. Second, since the Jumble Task required using several "sub-solutions" to arrive at the puzzle solution, it was more time-consuming than the anagram task. This therefore presented subjects with a similar amount of unscrambling as in the anagrams, but made performance more lumpy (or static); as explained in detail later, this was necessary to equalize marginal financial returns across conditions.

Subjects were also tasked with completing the risk elicitation mechanism from Dohmen and Falk (2011). In this task, subjects made choices between fixed payments and a lottery; the fixed payments ranged from $0.25 to $3.75 in $0.25 increments, while the lottery yielded $0 with 50% probability and $4 with 50% probability. This provided a subject-specific certainty-equivalence measure so we could control for risk preferences when testing the relationship between incentive systems and cheating.

Experimental Conditions

Our experimental conditions varied the incentive system utilized for the Jumbles Task. We varied three incentive elements in particular: *structure*, *pay*, and *social frames*. *Structure* indicates how performance is measured; we used a baseline structure that measured performance via a simple count of jumbles puzzles correctly solved (the control condition), and we compared this to a structure that measured performance according to whether a threshold goal was reached (denoted in this paper as condition G). *Pay* indicates how subject compensation was determined; we used a baseline that paid subjects a flat fee regardless of their performance (control), and we compared this to pay-for-performance

[2] Full instructions for the experiment are attached as an appendix.

schemes (P). *Social frames* indicates whether any information was given about the performance of others; the baseline condition made no reference to any social frames (control) while the treatment condition used peer comparison language to frame task performance (S). In order to study interaction effects, we also studied the use of two or all three of the non-control factors. This therefore represents a 2x2x2 experimental design. The study's eight conditions are:

Condition	Description*
DYB	The do-your-best control condition with no goals, social frames, or pay-for-performance.
G	The goal condition, where performance was measured relative to a goal threshold.
S	The social frames condition, where explicit peer comparison language was used to frame task performance.
P	The pay-for-performance condition, where subjects were paid based on their performance level as measured by a simple count.
GP	The incentivized goal condition, where performance and pay were both determined relative to a goal.
PS	The pay-for-performance condition with social comparison language.
GS	The goal condition with social comparison language.
GPS	The incentivized goal condition with social comparison language.

*In all conditions, subjects were asked to self-report the number of puzzles correctly solved.

This design contrasts in important ways with previous literature. In the parlance of this study, Schweitzer et al. (2004) only used the GS, GPS, and DYB conditions. The Cadsby et al. (2010) study only used P, GP, and DYB (and a tournament-based scheme). Our study therefore fills in the missing conditions. It also equates expected total pay, marginal pay, and marginal returns to cheating across all conditions, unlike the two previous studies.

We used the same language as in Schweitzer et al. (2004) to communicate goals and social frames. In any condition containing G, the goal was communicated as:

> "In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**."

In conditions involving S, we added the statement:

> "In previous studies, many subjects were able to solve **3** or more of these same puzzles within the allotted time."

Finally, in all conditions, participants were asked to "please solve as many puzzles as you can." Figure 2 gives the full instructions that were read to subjects and given in paper form before the experiment.

It was necessary to equalize marginal financial returns to cheating when comparing the rates of cheating across all conditions. To do this, we used pilot studies to select word jumbles that took the average subject three to four minutes to complete. This meant the median subject could complete two jumbles, but not three, in the eight minutes allotted for each round. We then set pay so that the average

subject, who solves two jumbles in a round, would earn the same total pay and face the same marginal return to cheating regardless of condition. This was done according to the table below.

Condition	Fixed Pay Awarded	Pay-for-Performance	Two Puzzles Solved Yields:	Marginal Return to Cheating (if two puzzles solved)
DYB	$6 per round.	N/A	$6 per round.	$0
G	$6 per round.	N/A	$6 per round.	$0
P	N/A	$3 per puzzle^	$6 per round.	$3
GP	$6 per round.	$3 when three puzzles are solved in a round*	$6 per round.	$3

Note: Conditions containing S (social comparisons) also followed the table of incentives above, since S did not alter financial incentives.
^This is identical to traditional piece-rate schemes.
***This is a standard pay-for-performance structure that utilizes goal thresholds.**

This design is not without shortcomings. While the median subject faces equivalent returns to cheating across conditions, subjects above or below the median may not. That is, subjects in the P and PS conditions could earn less than $6 in a round, while those in the GP and GPS conditions were guaranteed at least $6 in each round;[3] similarly, subjects in the P and PS conditions had the opportunity to earn even more pay by solving (or misreporting) performance higher than three jumbles in a round. However, our results are completely robust to only considering cheating that led to reported performance of three or less; moreover, all of our results are also robust to only examining choices of participant-rounds with actual performance of exactly two.

Participants, Materials & Setup

The experimental sessions were carried out using Computer Laboratory for Experimental Research (CLER) subjects at Harvard Business School. Each session recruited a maximum of 35 subjects per session, with an average of 20 participants showing up for each. All subjects were students who spoke English as their primary language.

Each session was conducted in a standard Harvard Business School MBA classroom with 90 seats. The same 35 seats were pre-numbered in each session so that they were spread across the room and had only one other nearby subject. Each such seat was labeled with a number and provided with pens and a stack of experimental paperwork containing the Jumble Task. These pairs of seats each contained an even-numbered seat and an odd-numbered seat, with even numbered-seats receiving blue-colored pens and odd-numbered seats receiving black-colored pens.

An empty black trash bin was placed between each pair of seats. This was provided so subjects could discard used test materials when instructed to do so during the experiment. At the end of the

[3]While guaranteed pay may reduce effort, we show later in the paper that performance was not affected by having guaranteed pay.

experiment, the contents of each trash bin contained only one copy of materials in blue ink (corresponding to the closest even seat number), and one copy of materials in black ink (corresponding to the closest odd seat number). By recovering the contents of each trash bin, we connected all task worksheets to a specific seat number and session, but not to a subject's name.

Procedure

Upon entering the room, subjects were instructed to choose a numbered seat and to put away all personal materials. They were instructed not to look through the papers in front of them until they were told to do so. They were truthfully instructed that the experiment was about decision making, and then were told of the Jumble Task and how they would be compensated (as shown in Figure 2).

The stack of papers at each seat included all materials necessary for the Jumble Task. First, it contained a one-page "answer sheet" specifically formatted for subjects to record *only* the final one-word solution to each jumble, and not the intermediate answers. There were also three packets of jumble puzzles corresponding to three eight-minute rounds; each round contained five different jumbles. At the end of every round, subjects were instructed to discard their test booklets, which contained the actual jumbles questions, into the trash bin next to them; they were told this was to "minimize the clutter" on their desks, since the experiment consisted of many packets of paper. They were told to keep their "answer sheet" for further rounds. Thus, subjects retained only the one-word solutions they derived during the round, and they discarded all other evidence of any progress they had made on each jumble.

After the third round, subjects were told that they would grade their own work. Crucially, subjects were not told that they would be self-grading their work until after they had completed the task, which prevented subjects from anticipating the ability to cheat and shirking as a result. Underneath the last jumble booklet was a pay slip indicating the subject's seat number and a simple format for reporting task performance (See Figures 3a and 3b for pay slips from the G and P conditions).[4] Subjects were instructed to use a solution key found underneath the pay slip to compare the one-word solution on their answer sheet to the one-word solution on the solution key. In all conditions, subjects were instructed to count the number of correctly-solved puzzles per round and to record this on the pay-slip. In goal conditions, subjects also had to report whether they met the goal in each round by circling a "Y" or "N" on the pay-slip. After they completed grading, subjects were told to discard their answer sheets and solution key in the trash bin, and to keep only their pay slip.

The subjects then completed the risk-elicitation mechanism, denoted as simply the "Situation Task," which was contained in the next packet of papers. In a table, subjects were presented with a set of fifteen choices labeled as "situations." Each situation was a choice between a fixed payoff (ranging from \$0.25 to \$3.75 in \$0.25 increments) and a lottery with a 50% probability of paying \$0 and a 50%

[4] Subjects could not see the pay-slip until after they completed the last jumbles round.

probability of paying $4. Subjects were told that one situation would be chosen for payment. After subjects made their choice for each situation, we paid subjects according to their choice on the "risk neutral" situation, which asked subjects to choose between a fixed payoff of $2 and the 50-50 lottery. For subjects who chose the lottery for this situation, we played the lottery using a Microsoft PowerPoint tool that displayed "win" or "lose" with a 50% probability of each. Subjects were then told to fill out their pay slip with the appropriate amount ($0, $2, or $4), and to discard the Situation Task worksheet in the bin.

The experimenter then collected all pay-slips into a single pile and arranged payment into envelopes. Each envelope was labeled according to seat numbers that corresponded to each pay-slip (thus maintaining anonymity). During this time a questionnaire was administered. This questionnaire contained personality measures for the Five Factor Personality Inventory, overconfidence, ethics, and demographics indicators,[5] as well as measures of subject perceptions of the fairness and difficulty of the Jumble Task. Subject payments were delivered in the envelopes as they finished this task. Once subjects finished the questionnaire, they placed the questionnaire in a box at the front of the classroom and were dismissed.

Post-experiment identification of cheating

Unbeknownst to subjects, we were able to tie all test booklets and answer sheets to their corresponding pay slip by salvaging the contents of each trash bin. Materials in each trash bin corresponded to only the two seats within its vicinity, and pen color distinguished between the two seats. Each questionnaire was also tied to a seat number via a unique question that was seat-specific; in all cases, this question closely resembled other questions on the questionnaire. Through this design, we avoided the need to make any false statements to subjects during the course of the experiment.

Cheating was identified using an objective, double-blinded procedure. Materials from all sessions were coded and mixed together to allow for double-blinded grading by research assistants (RAs). This included transferring the self-reported performance information on each pay-slip to another form that was identical across conditions; this way, RAs could not identify the corresponding condition according to the pay-slip. RAs identified cheating by comparing this new form to subjects' salvaged test booklets and answer sheets, which contained all work and solutions derived by the subject. All materials were graded twice by separate RAs, who did not have any details about the study's conditions or what the study was about.[6] The RAs were instructed to count the following evidence as cheating:

1. A round with more claimed correct answers than actually written on their answer sheet.
2. A question with an incorrect answer crossed off and the correct answer written in; this was counted only if all of the evidence in the workbook supported the crossed-off answer.

[5]The Big 5 questions were based on scales developed by Rammstedt and John (2007), Saucier (1994), and Gosling et al. (2003), and the overconfidence questions were based on Larkin and Leider (2012). The ethics scales were based on Robin et al. (1996) and Forsyth (1980).

[6]In rare cases where the graders disagreed, the materials were re-graded twice more. In those situations, answers were only marked as cheating if at least 3 out of the 4 gradings identified cheating.

3. A question with the correct answer on the answer sheet, but little or no evidence on the discarded test booklet to indicate that a subject had made significant progress on the question (e.g., the subject had revealed zero to two of the letters of the solution[7]).

This definition of cheating is in fact quite conservative, and is much more likely to fail to detect instances of cheating than it is to falsely identify cheating. For example, if a subject's work indicated she had correctly unscrambled three of the five letters, we gave her credit for actually solving the puzzle. In pilots (that did not involve cheating), virtually no participants had correct answers with fewer than four of the letters correctly unscrambled. Furthermore, in instances where evidence of cheating was more ambiguous, we gave the benefit of the doubt to the subject in order to reduce false positives. Our results are robust and in many ways stronger when broader definitions of cheating are used.

Results

There were 412 subjects across the eight conditions. The PS condition has fewer subjects than the other conditions. This is because the last session we ran was PS, and due to the large number of subjects in the study (including the pilot sessions) we were unable to recruit enough subjects. Each of the other conditions has at least 45 subjects.

Calibration of Performance and Pay

We attempted to calibrate the Jumbles Task so that the average subject would solve two puzzles per round. Going by actual performance and not self-reported performance, exactly two jumbles were solved successfully 32% of the time, while exactly three were solved 20% of the time. Four or five were solved successfully only 8% of the time, and one jumble was solved 26% of the time. The remaining 14% of rounds yielded zero correct solutions. Therefore, the median and mean subject did indeed solve two puzzles per round.

For each condition, Table 1 shows average pay and average "notional pay," which represents what average pay would have been if every subject had accurately reported performance. Both pay and notional pay are statistically indistinguishable across conditions, indicating that the study design successfully equalized total pay for all conditions. There was a greater variance in pay in the P and PS conditions because some subjects were very good or very bad at the Jumbles Task.

Average Cheating on the Jumbles Task across Conditions

Table 2 shows the percentage of individual jumble puzzles on which subjects cheated across the experiment's eight conditions. Subjects cheated on 3% of all jumbles puzzles. As expected, cheating was

[7]A few subjects did try to "force solve" puzzles without first unscrambling all of the initial four words, but their work in the worksheet made it clear when they were successful in doing so. These subjects wrote down all possible letter combinations of the jumble solution and all possible letter sequences of each possible combination, until a plausible solution was derived. This was rare because there were too many possible letter combinations.

least in the DYB control; nonetheless, a few subjects in the DYB condition still cheated. This may indicate measurement error, but based on comments in the questionnaires, some DYB subjects likely chose to misreport their performance because they did not agree with the solution provided or because they felt they deserved credit for the progress they made on puzzles they did not complete.[8]

Figure 4 shows cheating levels and standard errors for the four non-interacted conditions: DYB, G, S, and P. It also shows the results of a series of tests of proportions comparing the mean levels of cheating in each condition. Cheating is statistically the same in the DYB and G conditions ($p=0.21$). The S condition has significantly more cheating than G ($p=0.02$), and the P condition has significantly more cheating than S ($p=0.01$). These basic results indicate that, at the pay levels of the study, the use of financial incentives causes the most cheating, followed by the use of social comparisons. The use of goals does not increase cheating compared to simply asking subjects to do their best and counting their performance at the end.

Figure 5 shows the same analysis for only the conditions containing goals: G, GS, GP and GPS. In this analysis, the baseline goal condition with no social frames or pay-for-performance serves as the condition for comparison. All the interacted conditions contain significantly more cheating than the baseline G condition, further indicating that the use of social comparisons and financial incentives will increase cheating. However, there is no difference across interacted conditions; thus, introducing financial incentives or social comparisons to goals increases cheating by approximately the same amount, but including both is not additive.

Figures 6 and 7 present similar analyses for the S conditions (S, GS, PS, GPS) and the P conditions (P, GP, PS, GPS), respectively. In both cases, there is never significantly more cheating than in the baseline P or S conditions. In other words, there are similar levels of cheating across all S conditions, and likewise there are similar levels of cheating across all P conditions. This is consistent with the previous finding that the effects of P and S on cheating are not additive.

We also constructed a measure of "adjusted cheating," which corrects for the fact that subjects in the P and PS conditions had financial incentives to report performance above three, while subjects in the GP and GPS conditions did not. For this measure, we ignored cheating in all conditions for questions that increased reported performance above three. As shown in Table 2, the basic patterns in cheating across conditions are very similar under the "adjusted cheating" measure, although cheating in the P condition falls the most with the adjustment, as expected. All of our results hold when using "adjusted cheating" rather than the baseline "cheating" measure. We therefore only report results using the "baseline" measure in the paper.

[8] This occurred more frequently when the subject listed a plural of a word as a solution (the instructions explicitly stated that plurals were not acceptable solutions, but subjects may have felt this rule was unfair). They were clearly told before grading that they were not supposed to mark any solution as correct if it was not listed on the official answer key.

Cheating: Regression Analysis

Regression analysis confirms the basic results reported above. We use a random effects logistic regression with subject as the panel variable and question-round as the time variable. An indicator variable representing "cheated on the question" is used as the dependent variable, and the conditions are the explanatory variables (DYB is the omitted condition). The results of this regression are reported in Model 1 of Table 3. Since the condition variables are all indicator variables, the coefficients for each condition indicate the difference in probability that a subject cheats on a question when in that condition. All the conditions cause significantly more cheating than DYB except for the baseline G condition.

Regression analysis allows us to examine the separate estimated effect of goals, social comparisons, and financial incentives across all conditions containing these elements. To do this, we introduce an indicator variable that takes a value of one for all conditions containing the element. For example, the "has goals" indicator is set equal to one for G, GS, GP, and GPS; the "has social comparisons" and "has financial pay" indicators are set up in similar fashion. These results are reported in Model 2 of Table 3. Conditions containing incentive pay and social comparisons are associated with significantly greater cheating, while conditions with goals are not.

We also run Model 1 on only those subject-rounds with three or fewer actual correct answers in order to eliminate instances where a subject's true performance in P or PS would have earned more than $9 in a round. The results, reported in Model 3 of Table 3, are nearly the same. The PS treatment becomes insignificant because there were fewer subjects in that condition. The results also tell a similar story when we run Model 2 on only rounds where a subject's true performance is exactly two questions (see Model 4 of Table 3), although the "has social comparisons" and "has financial pay" indicators are only marginally significant due to the smaller sample sizes.

Risk Preferences

We elicited risk preferences using the Situation Task, a series of choices between fixed payments and a 50-50 lottery. Model 5 in Table 3 uses an indicator to represent whether subjects were risk-averse or not (i.e., if they switched from lottery to fixed payments when the fixed payment was still under $2.00). Risk-aversion was not predictive of cheating on the Jumbles Task for any of the conditions. In addition, including risk-aversion as a control does not significantly alter any of the previously reported results on cheating in the Jumbles Task, although results for "has social comparisons" loses statistical significance due to the inclusion of so many interaction terms. Additionally, interactions between risk-aversion and condition were not significant.

Jumbles Performance across Conditions

In order to better equilibrate total pay and marginal pay across conditions, we chose a task that measured performance more discretely than the prior literature, which used simple anagrams.

Consequently, our study is less conducive than previous studies for measuring performance differences across different incentive systems. However, unlike the previous literature, our analysis of task performance is not subject to possible confounds from differences in marginal or total pay across conditions.

To determine if subjects performed differently across conditions, we ran a regression that measured the effects of indicator variables for "has G," "has P," and "has S" on an indicator for whether the question was correctly answered (without cheating). The results are reported in Table 4. None of the condition indicators are predictive of a change in performance. This is consistent with previous studies that used similar word unscrambling tasks and also found no treatment effect on performance (Schweitzer et al. 2004; Cadsby et al. 2010). If any of these incentives influenced subject effort or persistence, then these results would suggests that one's performance on this task may not be purely a function of effort. Of course, these results may also be due in part to the fact that our task was not conducive to measuring changes in performance.

To examine the effects of coming close on performance, we coded an indicator variable equal to one if a subject successfully identified four or more of the letters of the final fifth word, but did not come up with the correct answer; in these instances, subjects successfully unscrambled at least three of the four initial words to a jumble. As it turns out, subjects across all conditions were equally likely to come close (but fail to complete) an individual jumble puzzle.

Table 5 reports the effect on cheating of "almost" solving a puzzle and also interacts this variable with "has G," "has S," and "has P." Just falling short of completing a puzzle is highly predictive of cheating; the interaction effects demonstrate that the effect is almost twice as strong in the G conditions. Prior research also suggests that subjects are more likely to cheat if they come close to reaching a goal but fall just short (Schweitzer et al. 2004).

Task Perceptions and Emotion

In the questionnaire, we asked several questions that solicited subject feelings about the Jumble Task as it relates to fairness, frustration, and pride (see experimental instructions for the entire list of questions). These were also balanced in terms of positive and negative frames. Examples include:

> *I am proud of the number of Word Jumbles I was able to solve in today's experiment.*
>
> *The Word Jumble puzzles were frustrating to me.*
>
> *The rules to the Word Jumble task were unfair.*

Subject feelings of fairness, pride, and frustration regarding the Jumble Task were not significantly different across conditions. In other words, none of the conditions seemed to influence subject feelings of fairness, pride, or frustration with the task.

Model (1) in Table 6 reports regression results of these task-specific emotional measures on a subject's propensity to cheat. Reporting pride in one's performance on the task was positively associated with cheating, while a general tendency toward frustration was negatively associated with cheating on the task. These findings suggest that cheating involves an emotional component, and that emotions induced by the task (but perhaps not by the incentive system elements) may play a role in promoting cheating. For instance, subjects who took pride in their verbal, vocabulary, or problem-solving skills may have felt that their self-concept was threatened by performing poorly on the task, and they may therefore be more likely to cheat. However, since these results are only correlational, it remains for future studies to examine the causal relationship between these emotions and cheating.

The questionnaire also asked subjects to guess the purpose of our experiment. We asked this to determine whether subjects suspected the study was about cheating; research has suggested that subjects who believe an experimenter wants them to behave a certain way will tend to comply with those expectations (Zizzo 2008). This measure may overestimate the number of subjects who suspected we were measuring cheating, since some subjects may not have given it a second thought until the question was asked. Nevertheless, only 15% of subjects across all conditions guessed that we were measuring cheating, honesty, or trust (as coded by double-blinded research assistants). Subjects in pay-for-performance and social comparison framing conditions were slightly more likely to guess that the study was related to honesty. Model (2) in Table 6 demonstrates that the point estimates for "has P" and "has S" are robust to controlling for this guess, but statistical significance is lower due in part to the inclusion of so many interaction terms.

Discussion

This study examined the specific elements of goal-based incentive systems that lead to cheating, using a careful 2x2x2 design that allowed for direct and precise comparisons of the effects of goal setting, social comparison framing, and pay for performance. The study found that goal-based systems cause cheating through their framing of the goal in social comparison terms, and when tying the goal directly to monetary compensation. Having a goal by itself did not lead to cheating that is different from a "do your best" condition.

This study provides important context for the existing literature that has typically varied more than one incentive system element at once. It also provides important evidence for the ongoing debate about the link between "goals" and cheating. Do "goals" cause cheating? Our study says the answer is

"no" and "yes" – "no" if one is referring to the use of a performance target as an element in an incentive plan, but "yes" if one is referring to the collective set of elements typically found in goal-based systems. We believe much of the existing debate on goals has unfortunately been caused by researchers talking at different levels of analysis, and we hope that this study will focus the debate on the differences between incentive systems overall and incentive system elements.

This study also contributes to the burgeoning literature on the downside of social comparisons. It is fascinating and important that negative social comparisons can trigger cheating even when the social comparison has nothing to do with pay, but only with performance standards. This is highly suggested by existing field research but has not yet been precisely and causally demonstrated.

Theoretical Implications

The study has a number of interesting theoretical implications. The goals literature has clearly shown that goals are more effective when employees are committed to them and when goals are more difficult to achieve. However, the positive results that stem from commitment and difficulty are clearly a double-edged sword. Social comparison framing is one classic method used by managers to increase goal commitment, yet this commitment may be exactly what causes greater cheating. The problem with social comparisons is only exacerbated when they are given at a level of high performance. Again, this is a positive for performance because it increases goal difficulty, but it also makes the number of employees for whom negative social comparisons are invoked all the greater. Our study suggests that some of the elements that are needed to make goal-based systems effective are also those that lead to higher cheating. This is a classic ethical dilemma when designing incentives; it is often impossible to separate the productive and counterproductive actions induced by incentive systems, since given elements of a system usually influence the likelihood of both in the same direction (Larkin and Pierce 2016).

Our results also suggest that emotion may play a role in choosing to cheat, and in particular that cheating may be a response to maintain pride in one's performance on a task. However, it is unclear whether these emotions are the cause of the decision to cheat, or a consequence of that decision. It is clear that better models of the emotional mechanisms behind cheating are needed. This would potentially be consistent with literature suggesting that cheating is highly task-dependent (Bazerman and Gino 2012).

Limitations

Our study has a number of limitations. It only modeled a small number of incentive system elements; real-world incentive systems are a complex system of formal rules, informal expectations and psychology (Larkin et al. 2012). Although the three elements studied in the paper were chosen to resolve confounds in the existing literature, more complex experiments, including field experiments, are needed to improve the generalizability of the study. Like all laboratory studies, generalizability to the workplace is not guaranteed. Indeed, our sample of primarily Harvard undergraduates may be particularly

unrepresentative of the broader workplace population in that these subjects worked extremely hard even in the do-your-best conditions. They also might be particularly prone to social comparison framing if they (correctly) believed that the communicated social comparison involved the performance of their classmates. Finally, the lumpy nature of task performance, necessitated by the requirement to equalize total and marginal pay across treatments, meant that the study could not delve into many interesting aspects of incentive system elements on performance and other key outcome variables.

Future Directions

We view our largest likely contribution in this paper to be a call to more and better research at the systems level of incentive design, and to specifically consider how individual elements of a system contribute to the positive and negative effects of a given system. The debate on goals clearly shows that academics are interested in system-level effects, and this is surely the level of interest of most managers. However, most of the research in organizational behavior is focused at the level of the system element, rather than the design of the overall system; however, much of the discussion in these papers is focused on systems-level analysis.

Indeed, we hope this study increases interest in systems-level research that carefully designs many treatment arms within a given system, in addition to the standard experimental methodology of repeating a given element-level experiment in many domains. Both approaches have great merit, but the latter approach inhibits the ability to perform systems-level research. This type of research can help bridge the significant differences that exist in the literature on incentive design, both within and across disciplinary boundaries, and can also help resolve the many unresolved questions and strong debates about optimal incentive design.

References

Bandura, A., F. Jourden. 1991. Self-regulatory mechanisms governing the impact of social comparison on complex decision-making. *Journal of Personality and Social Psychology*. 60(6), 941-951.

Barankay, I. 2010. Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment. Working paper.

Bazerman, M.H., Gino, F. 2012. Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and the Social Sciences*. 8, 85-104.

Barsky, A. 2008. Understanding the ethical cost of organizational goal-setting: a review and theory development. *Journal of Business Ethics*. 81, 63-81

Blanes i Vidal, J., M. Nossol. 2011 Tournaments without prizes: Evidence from personnel records. *Management Science*. 57, 1721-1736.

Cadsby, C. B., F. Song, F. Tapon. 2010. Are you paying your employees to cheat? An experimental investigation. *The B.E. Journal of Economic Analysis & Policy*. 10(1).

Camerer, C.F., Hogarth, R.M., Budescu, D.V., Eckel, C. 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*. 19(1-3), 7-48.

Card, D., A. Mas, E. Saez. 2012. Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*. 102(6), 2981-3003.

Cohn, A., E. Fehr, B. Herrmann, F. Schneider. 2014. Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association.* 12(4), 877-898.

Crossley, C. D., Cooper, C. D., Wernsing, T. S. 2013. Making things happen through challenging goals: Leader proactivity, trust, and business-unit performance. *Journal of Applied Psychology*. 98(3), 540-549.

Dohmen, T., A. Falk. 2011. Performance pay and multi-dimensional sorting: Productivity, preferences and gender. *American Economic Review*. 101(2), 556-90.

Edelman, B, Larkin. I. 2015. Social comparison and deception across workplace hierarchies: Field and experimental evidence. *Organization Science*. 26(1), 78-98.

Eres, M., I. Zidon. 1984. Effects of goal acceptance on the relationship of goal setting and task performance. *Journal of Applied Psychology*. 69, 69-78.

Feather, N. T. 1989. Attitudes towards the high achiever: The fall of the tall poppy. *Australian Journal of Psychology*. 41(3), 239–267.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations*. 7(2), 117-140.

Forsyth, D. R. 1980. A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology*. 39(1), 175-184.

Gerhart B, S. Rynes, I. S. Fulmer. 2009. Pay and performance: individuals, groups, and executives. *Academy of Management Annals*. 3, 251–315.

Gino, F. 2015. Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences*. 3, 107–111.

Gino F, Margolis J.D. 2011. Bringing ethics into focus: how regulatory focus and risk preferences influence (un)ethical behavior. *Organizational Behavior and Human Decision Process*. 115, 145-156.

Gosling, S. D., P. J. Rentfrow, W. B. Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*. 37, 504-528.

Greenberg J. 1993. Stealing in the name of justice: Informational and interpersonal moderators of theft reactions to underpayment inequity. *Organizational Behavior and Human Decision Processes*. 54, 81-103.

Gubler, T., Larkin, I., Pierce, L. 2016. Motivational spillovers from awards: Crowding out in a multitasking environment. *Organization Science*. 27(2), 286-303.

Heath C., R. P. Larrick, G. Wu. 1999. Goals as reference points. *Cognitive Psychology*. 38, 79-109.

Jacob, B. A., S. D. Levitt. 2003. Rotten apples: An investigation of the prevalence and predictors of teacher cheating**.** *Quarterly Journal of Economics. 118*(3), 843-77.

Jensen, M. C. 1998. *Foundations of Organizational Strategy*. Harvard University Press: Cambridge.

John, L.K., Loewenstein, G., Rick, S.I. 2014. Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat. *Organizational Behavior and Human Decision Processes, 123*, 101-109.

Klein, H., Wesson, M., Hollenbeck, J., B. Alge. 1999. Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology*. *84*, 885-896.

Larkin, I. 2011. Paying 30,000 for a gold star: An empirical investigation into the value of peer recognition to software salespeople. Working paper.

Larkin, I. 2014. The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, 32(2), 1-29.

Larkin, I., Leider, S. 2012. Incentive schemes, sorting and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics*. 4(2), 184-214.

Larkin, I., Pierce, L. 2016. Compensation and employee misconduct: The inseparability of productive and counterproductive behavior in firms. In Palmer, D., Smith-Crowe, K., Greenwood, R. (Eds.): *Organizational Wrongdoing: Key Perspectives and New Direction*, Cambridge University Press, 270-295.

Larkin, I. Pierce, L., Gino, F. 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*. 33(10), 1194-1214.

Larrick, R.P., Heath, C., Wu, G. 2009. Goal-induced risk taking in negotiation and decision making. *Social Cognition*. 27, 342-364.

Latham, G.P., J. Baldes. 1975. The "practical significance" of Locke's theory of goal setting. *Journal of Applied Psychology*. 60, 122-124.

Latham, G.P., S. B. Kinne. 1974. Improving job performance through training in goal setting. *Journal of Applied Psychology*. 74, 411-416.

Locke, E. A., G. P. Latham. 1990. *A theory of goal setting and task performance*. Prentice Hall, Englewood Cliffs, NJ.

Locke, E. A., G. P. Latham. 2002. Building a practically useful theory of goal setting and task motivation. *American Psychologist*. 57(9), 705–717.

Locke, E.A., G. P. Latham. 2006. New directions in goal-setting theory. *Current Directions in Psychological Science*. 15(5), 265-268.

Locke, E. A., G. P. Latham. 2009. Has goal setting gone wild, or have its attackers abandoned good scholarship? *Academy of Management Perspectives. 23*(1), 17-23.
Locke, E.A., Shaw, K.N., Saari, L.M. & Latham, G.P. 1981. Goal setting and task performance: 1969-1980. *Psychological Bulletin*. 90, 125-152.
Mas, A., Moretti, E. 2009. Peers at work. *American Economic Review*. 99, 112–145.
Mazar, N., O. Amir, D. Ariely. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*. 45, 633-644.
Mead N.L., Baumeister, R.F., Gino, F., Schweitzer, M.E., Ariely D. 2009. Too tired to tell the truth: self-control resource depletion and dishonesty. *Journal of Experimental and Social Psychology*. 45, 594-597.
Moran, S., M. Schweitzer. 2008. When better is worse: Social comparison and the use of deception. *Negotiation and Conflict Management Research*. 1(1), 3-29.
Ordóñez, L., M. E. Schweitzer, A. D. Galinsky, M. H. Bazerman. 2009a. On good scholarship, goal setting, and scholars gone wild. *Academy of Management Perspectives.* 23(1), 6-16.
Ordóñez, L., M. E. Schweitzer, A. D. Galinsky, M. H. Bazerman. 2009b. Goals gone wild: How goals systematically harm individuals and organizations. *Academy of Management Perspectives.* 23(3), 82-87.
Ordóñez, L., D. Welsh. 2015. Immoral goals: How goal setting may lead to unethical behavior. *Current Opinion in Psychology*. 6, 93-96.
Ordóñez, L. D., Wu, G. 2013. Goals and decision making. In S. Highhouse, E. Salas and R. S. Dalal (Eds): *Judgment and Decision Making*, SIOP Organizational Frontiers Series, 123-139.
O'Reilly III, C.A., Main, B.G., Crystal, G.S. 1988. CEO compensation as tournament and social comparison: A tale of two theories. *Administrative Science Quarterly*. 33(2), 257-274.
Rammstedt, B., O. P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*. 41, 203-212.
Rick, S., G. Loewenstein. 2008. Hypermotivation. *Journal of Marketing Research*. 45, 645-653.
Robin, D. P., R. E. Reidenbach, P. J. Forrest. 1996. The perceived importance of an ethical issue as an influence on the ethical decision-making of ad managers. *Journal of Business Research*. 13(1), 17-28.
Saucier, G. 1994. Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment.* 63, 506-516.
Schweitzer, M. E., L. D. Ordóñez, B. Douma. 2004. Goal setting as a motivator of unethical behavior. *Academy of Management Journal*. 47(3), 422–432.
Seijts, G. H., G. P. Latham. 2000. The construct of goal commitment: Measurement and relationships with task performance. In R. Goffin and E. Helmes (Eds.), *Problems and solutions in human assessment.* Dordrecht, the Netherlands: Kluwer Academic, 315-332.
Singh, R., Larkin, I. 2015. Auditor conservatism, incentive compensation, and the quality of financial reporting. *Journal of Law Economics and Organization*. 31(4), 721-751.
Stajkovic, A. D., Luthans, F. 1988. Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*. 124(2), 240-261.
Staw, B. M., R. D. Boettger. 1990. Task revision: A neglected form of work performance. *Academy of Management Journal*. 33(3), 534-559.
Tesser, A. 1988. Toward a self-evaluation maintenance model of social behavior. L. Berkowitz, ed. *Advances in Experimental Social Psychology* (Vol. 21). Academic Press, New York, 193-355.
Weiner, B. 1986. *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
Welsh, D., Ordóñez, L. D. 2014. The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior. *Organizational Behavior and Human Decision Process,* Special Issue on Behavioral Ethics. 123, 79-89.
Wolfers, J. 2006. Point shaving: Corruption in NCAA basketball. *American Economic Review*. 96(2), 279-283.
Zhang, X., Bartol, K.M., Smith, K.G., Pfarrer, M.D., Khanin, D.M. 2008. CEOs on the edge: earnings manipulation and stock-based incentive misalignment. *Academy of Management Journal*. 51, 241-258.
Zitzewitz, E. 2012. Forensic economics. *Journal of Economic Literature*. 50(3), 731-69.
Zizzo, D. 2010. Experimenter demand effects in economic experiments. *Experimental Economics*. 13(1): 75-98.

Figure 1: Word Jumble Instructions and an Example Word Jumble and Solution

Instructions given to participants (in all treatments)
Your first task is to solve a series of "Word Jumble" puzzles. Each puzzle consists of four scrambled words that you must unscramble. Once correctly unscrambled, the letters that are circled form a fifth scrambled word that you must also unscramble. This fifth word is the Puzzle Solution. A puzzle is considered correct AS LONG AS this FIFTH scrambled word (the Puzzle Solution) is correct.

Please note that all words are English words that can be found in any typical English dictionary. The words in these exercises do not have any relation to one another (unless by coincidence). Words can be of any kind (nouns, verbs, adjectives, adverbs, etc.). Most are words commonly used in speech or writing. Proper nouns, plurals, slang or jargon are NOT considered words in this exercise. Verbs can be in any tense.

Example puzzle:

S E R E A

E N T G A

S C I B A

P R O W E

Scrambled 5th word: ___ ___ ___ ___ ___

Solution:

Solution to Example:
Erase, Agent, Basic, Power

Scrambled letters:
R T A W E

Puzzle Solution: WATER

Figure 2: Instructions by Treatment Type

In all treatments, subjects were instructed to "Please complete as many puzzles as you can." In addition, subjects were given the following instructions based on treatment:

In non-incentivized and all goal conditions (DYB, G, S, GS, GP, GPS)
"In each round, you will be paid **$6**, regardless of how many puzzles you correctly solve. Please complete as many puzzles as you can."

In all goal conditions (G, GS, GP, GPS):
"In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**."

In all incentivized goal conditions (GP, GPS):
"You will be paid an additional **$3** in each round where the goal is met."

In all social comparison conditions (S, GS, PS, GPS):
"In previous studies, many subjects were able to solve **3** or more of these same puzzles within the allotted time."

In incentivized treatments without goals (P, PS):
"In each round, you will be paid **$3** for each puzzle you solve. Please complete as many puzzles as you can."

Instructions for treatments with multiple elements would contain all relevant instructions. For example, instructions for the GPS treatment read:

> In each round, you will be paid **$6**, regardless of how many puzzles you correctly solve. Please complete as many puzzles as you can. In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**. You will be paid an additional **$3** in each round where this goal is met. In previous studies, many subjects were able to solve **3** or more of these same puzzles within the allotted time.

Instructions for the PS treatment read:

> In each round, you will be paid **$3** for each puzzle you solve. Please complete as many puzzles as you can. In previous studies, many subjects were able to solve **3** or more of these same puzzles within the allotted time.

Figure 3a – Pay Slip for Non-Incentivized Goal Conditions (G, GS)

Pay Slip
Seat #1

TASK PERFORMANCE:

Task	Results	Goal
Jumble Task Rd 1	____ correct	Goal Met? Y/N
Jumble Task Rd 2	____ correct	Goal Met? Y/N
Jumble Task Rd 3	____ correct	Goal Met? Y/N

Total Number of Times You Met the Goal:
(0, 1, 2, or 3)

TASK PAY:

Task	Pay
Jumble Task Rd 1-3	$18
Situation Task	$_____

Total Pay:
($18 + pay for Situation Task)

Figure 3b – Pay Slip for Incentivized Non-Goal Treatments (P, PS)

Pay Slip
Seat #1

TASK PERFORMANCE:

Task	**Results**
Jumble Task Rd 1	____ correct
Jumble Task Rd 2	____ correct
Jumble Task Rd 3	____ correct

Total Performance: [ ]

TASK PAY:

Task	**Pay**
Jumble Task Rd 1-3	$_____ ($3 x Total # Correct)
Situation Task	$ _____

Total Pay: [ ]
(sum of earnings in jumble task rds 1-3 and earnings from the situation task)

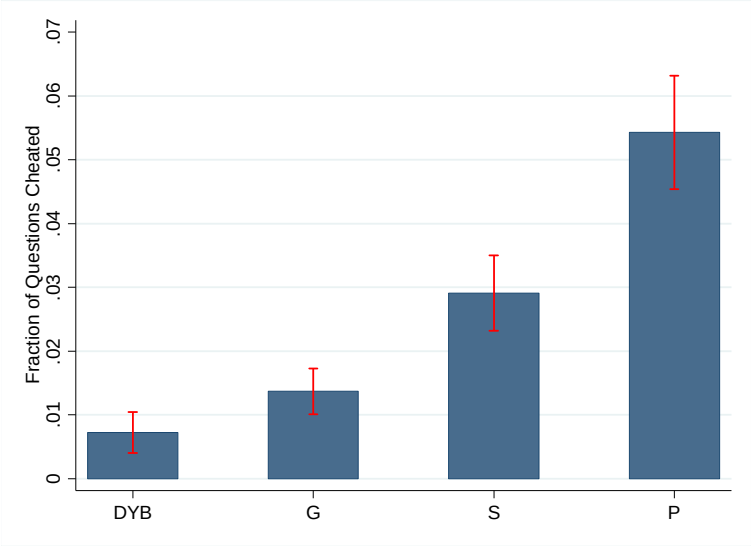
Table 1 – Subjects and average pay by condition (includes Situation Task pay)

Condition	# Subjects	Average Payment	SD	Average "notional" pay	SD
DYB	46	$20.91	1.01	$20.91	1.01
G	68	$20.15	1.00	$20.15	1.00
S	55	$20.15	1.21	$20.15	1.21
P	43	$21.68	8.35	$19.23	7.89
GS	59	$19.90	1.20	$19.90	1.20
GP	67	$24.00	3.40	$21.54	1.62
PS	23	$21.39	8.85	$19.96	7.00
GPS	51	$23.20	3.26	$20.61	1.66
Total	412	$21.43	4.18	$20.37	3.30

Table 2 – Cheating and "adjusted cheating" by condition

Condition	Cheating	SE	Adjusted cheating	SE
DYB	0.0072	0.0032	0.0072	0.0032
G	0.0137	0.0036	0.0108	0.0032
S	0.0291	0.0059	0.0255	0.0055
P	0.0543	0.0089	0.0496	0.0086
GS	0.0339	0.0061	0.0316	0.0059
GP	0.0428	0.0064	0.0418	0.0063
PS	0.0319	0.0095	0.0174	0.0071
GPS	0.0431	0.0074	0.0405	0.0071
Average	*0.0316*	*0.*	*0.0254*	*0.*

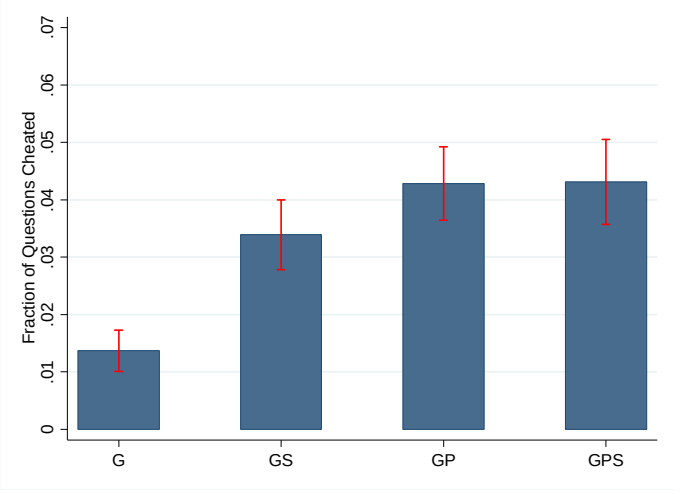
Figure 4 – Cheating comparisons across non-interacted conditions



Note: bars represent +/- one standard error

Condition 1	Condition 2	T-statistic (two-tailed)	P-value
DYB	G	1.254	0.210
G	S	2.313	0.021
S	P	2.443	0.015

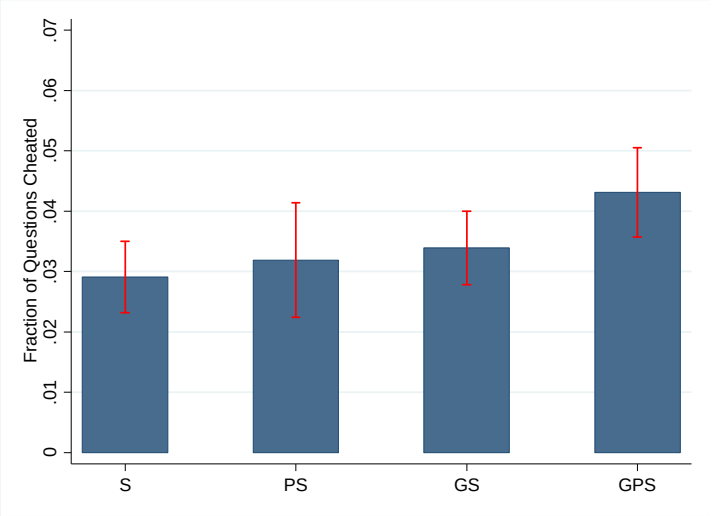
Figure 5 – Cheating comparisons across goal conditions



Note: bars represent +/- one standard error

Condition 1	Condition 2	T-statistic (two-tailed)	P-value
G	GS	2.928	<0.003
GS	GP	1.000	0.317
GP	GPS	0.036	0.971

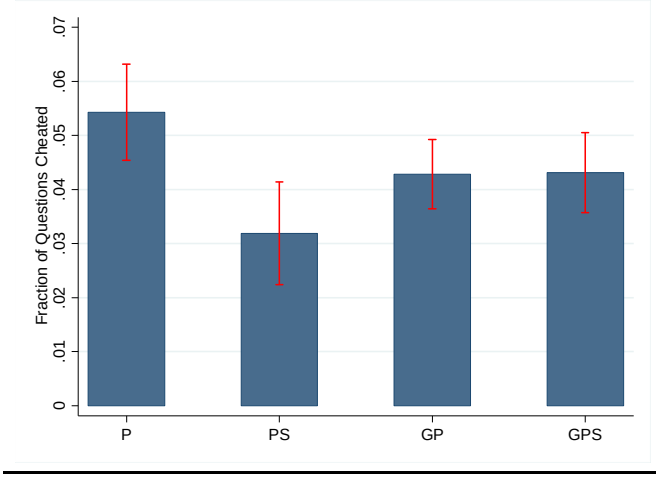
Figure 6 – Cheating comparisons across social comparison conditions



Note: bars represent +/- one standard error

Condition 1	Condition 2	T-statistic (two-tailed)	P-value
S	PS	0.256	0.798
PS	GS	0.177	0.860
GS	GPS	0.976	0.329

Figure 7 – Cheating comparisons across incentive pay conditions



Note: bars represent +/- one standard error

Condition 1	Condition 2	T-statistic (two-tailed)	P-value
P	PS	1.595	0.111
SP	GP	0.891	0.373
GP	GPS	0.036	0.971

Table 3 – Cheating regressions

Random effects logit model; time variable is question-round and panel variable is subject
Dependent variable is indicator=1 if a subject cheated when grading the answer
Standard errors in parentheses.

Variables	(1) All questions	(2) All questions	(3) Drop rounds where actual perf.>=3	(4) Rounds where actual perf. ==2	(5) All questions
G	0.538		0.123		
	(0.701)		(0.778)		
S	1.440**		1.266*		
	(0.679)		(0.736)		
P	1.884***		1.584**		
	(0.689)		(0.741)		
GS	1.664**		1.538**		
	(0.664)		(0.708)		
GP	1.863***		1.823***		
	(0.650)		(0.693)		
PS	1.572**		1.083		
	(0.793)		(0.921)		
GPS	1.912***		1.814**		
	(0.670)		(0.719)		
Has G		0.160		0.363	0.210
		(0.281)		(0.369)	(0.329)
Has S		0.538*		0.597*	0.393
		(0.278)		(0.356)	(0.327)
Has P		0.856***		0.678*	0.736**
		(0.279)		(0.358)	(0.329)
Risk Averse					-0.070
					(0.724)
Risk Averse * Has G					-0.201
					(0.640)
Risk Averse * Has S					0.433
					(0.630)
Risk Averse * Has P					0.309
					(0.632)
Constant	-6.063***	-5.439***	-5.786***	-5.077***	-5.398***
	(0.603)	(0.360)	(0.639)	(0.509)	(0.395)
Observations	6180	6180	4420	1975	6180
Number of subjects	412	412	372	266	412

*** p<0.01; ** p< 0.05; * p<0.10

Table 4 – Task performance

Variables	(1)
Has G	-0.144
	(0.097)
Has S	0.059
	(0.091)
Has P	-0.047
	(0.091)
Constant	-0.508***
	(0.091)
Observations	6180
Number of subjects	412

***p<.01

Random effects logit model; time variable is question-round and panel variable is subject
Dependent variable is whether a question was correctly answered (mean=0.37; SD = 0.48)
Standard errors in parentheses.

Table 5 – Effect on cheating of "coming close" on a puzzle

Variables	(1)
Almost	1.122**
	(.497)
Has G	-0.140
	(0.311)
Has S	0.533*
	(0.311)
Has P	1.070***
	(0.315)
Almost*Has G	1.183***
	(0.451)
Almost*Has S	0.147
	(0.414)
Almost*Has P	-0.646
	(0.417)
Constant	-5.743***
	(0.400)
Observations	6180
Number of subjects	412

*** p<0.01; ** p< 0.05; * p<0.10

Random effects logit model; time variable is question-round and panel variable is subject
Dependent variable is indicator=1 if a subject cheated when grading the answer
Standard errors in parentheses.

Table 6 – Subject attributes predicting cheating

Random effects logit model; time variable is question-round and panel variable is subject
Dependent variable is indicator=1 if a subject cheated when grading the answer
Standard errors in parentheses.

Variables	(1)	(2)
Idealist ethics	0.0466	
	(0.237)	
Relativist ethics	-0.198	
	(0.227)	
Task pride	0.571***	
	(0.172)	
General pride	-0.300	
	(0.244)	
Task frustration	-0.001	
	(0.159)	
General frustration	-0.408**	
	(0.0.176)	
Task fairness	0.062	
	(0.297)	
General fairness	0.138	
	(0.190)	
Honesty		-1.291
		(0.965)
Has G	0.294	0.127
	(0.272)	(0.304)
Has S	0.503*	0.428
	(0.264)	(0.302)
Has P	0.727***	0.572*
	(0.268)	(0.302)
Honesty * Has G		-0.252
		(0.800)
Honesty * Has S		0.388
		(0.759)
Honesty * Has P		1.982**
		(0.954)
Constant	-4.934**	-5.233***
	(2.063)	(0.374)
Observations	6165	6180
Number of subjects	411	412

Note: one subject did not fill out the "personality test" questions
*** p<0.01; ** p< 0.05; * p<0.10

Why do goal-based incentives cause cheating?
Unpacking the confounding effects of goals, social comparisons and pay

FULL SET OF EXPERIMENTAL MATERIALS

This document represents the full set of materials that each subject received during the experiment. Subjects were instructed not to leaf through the stack of papers, and no subjects were observed doing so.

Where the materials differed by experimental condition, it is clearly noted in the attached. The word jumble puzzles, answer key, Situation Task/risk elicitation, and questionnaire did not vary across treatments. The basic instructions also did not differ by treatment type, except as noted in Figure 2A of the attached. The pay slips differed slightly by treatment, as show in Figure 6A. Note subjects did not see the Pay Slip, or know they would be grading their own work, until *after* they had completed the word jumble task.

Figure 1A – Sample Instructions (Goal, Incentivized and Social Comparison Treatment)

Instructions

Welcome and thank you for participating. At this point, please don't talk to any other participant or look at their materials. Please remove any personal materials from your desk (you may place them underneath your seat or off on a side table). You should not use any materials during the course of this experiment that are not provided to you by the experiment. If you have any questions, raise your hand.

On your desk are two consent forms. If you haven't yet, please read through these and sign both copies. One will be handed in to me, and the other is for your records. After you have signed them, please put them in the empty manila folder in front of you. At this time, please also place your receipt into the manila folder. This receipt is for HBS CLER record-keeping purposes, and will not be used by the researchers in any way for any research purposes.

Today's session is a study on individual judgment. Over the course of the experiment, you will solve puzzles and make choices that may affect your total earnings. I will explain the details of these tasks as we proceed.

Please note that everything about today's experiment is completely anonymous. As such, please DO NOT write your name on any materials (except the consent form). Your name will NOT be used in any way by the researchers conducting this study. In fact, the researchers will not have access to your name or anything that identifies you.

You have a set of papers on the desk in front of you. Please only take a paper or set of papers when I instruct you to do so. To avoid clutter, please pay careful attention to what I tell you to do with each set of papers. By the end of the experiment, I will have told you to put nearly all the papers in front of you into the bin that's next to you. I will be using the same tasks for future CLER experiments. It is therefore very important that you not leave the room with any of the papers that have the tasks or answers on them. Please make sure to put papers in the bin as I ask you to do so.

Figure 1A (Continued) – Sample Instructions

Task 1: Word Jumbles

Your first task is to solve a series of "Word Jumble" puzzles. Each puzzle consists of four scrambled words that you must unscramble. Once correctly unscrambled, the letters that are circled form a fifth scrambled word that you must also unscramble. This fifth word is the Puzzle Solution. A puzzle is considered correct AS LONG AS this FIFTH scrambled word (the Puzzle Solution) is correct.

Please note that all words are English words that can be found in any typical English dictionary. The words in these exercises do not have any relation to one another (unless by coincidence). Words can be of any kind (nouns, verbs, adjectives, adverbs, etc.). Most are words commonly used in speech or writing. Proper nouns, plurals, slang or jargon are NOT considered words in this exercise. Verbs can be in any tense.

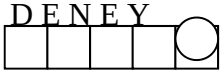
When the instructor gives you permission, please take 3 minutes to attempt the practice puzzle below. This is a puzzle of average difficulty. If you finish early, please sit quietly at your desk and wait for further instructions.

Example 1:

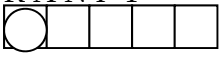
O P Y P P

H O T U M

D E N E Y



R A N I T



Scrambled 5th word: ___ ___ ___ ___ ___

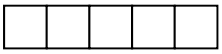
Puzzle Solution: 

Figure 1A (Continued) – Sample Instructions

Solution to Example 1:
Poppy, Mouth, Needy, Train

Scrambled letters:
O U H Y T

Puzzle Solution: Youth

We will be doing the Word Jumble task over three rounds. In each round, you will try to solve as many jumbles as you can. In each round, there are 5 total jumbles, but it is unlikely you will be able to solve all 5 in the time allotted. Each round lasts 8 minutes.

You MAY skip puzzles if you like, but keep the following factors in mind:

1. As mentioned, you MUST have the fifth scrambled word correct in order to receive credit. There is no partial credit for incomplete puzzles or incorrect solutions.

2. The questions are ordered in increasing difficulty. Thus, while skipping puzzles may be necessary if you are stuck, the next puzzle is likely to be more difficult.

It is important that you complete as many "Word Jumble" puzzles as you can. In return for completing as many puzzles as you can, we will compensate you at the end of the experiment. Your compensation for the "Word Jumbles" task is as follows:

> In each round, you will be paid **$6**, regardless of how many puzzles you correctly solve. Please complete as many puzzles as you can. In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**. You will be paid an additional **$3** in each round where this goal is met. In previous studies, many students were able to solve **3** or more of these same puzzles within the allotted time.

The test booklet for Round 1 is on top of the stack of papers in front of you. Please take this booklet now, but do not open it. The next sheet of paper is the Answer Sheet. Please take this sheet now. Once you have these materials in front of you, please discard this set of instructions into the bin next to you.

Figure 2A: Instructions by Treatment Type

All Treatments except Piece Rate:
"In each round, you will be paid **$6**, regardless of how many puzzles you correctly solve. Please complete as many puzzles as you can."

Goal Treatments:
"In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**."

Incentivized Treatments:
"You will be paid an additional **$3** in each round where the goal is met."

Social Comparison Treatments:
"In previous studies, many students were able to solve **3** or more of these same puzzles within the allotted time."

Piece Rate Treatment:
"In each round, you will be paid **$3**for each puzzle you solve. Please complete as many puzzles as you can."

Instructions for treatments with multiple elements would contain all relevant instructions. For example, instructions for treatment with Goal, Incentivized and Social comparison would read:

> In each round, you will be paid **$6**, regardless of how many puzzles you correctly solve. Please complete as many puzzles as you can. In addition, it is important that you commit to a specific yet attainable goal. Your goal for each round is to solve **3 puzzles**. You will be paid an additional **$3** in each round where this goal is met. In previous studies, many students were able to solve **3** or more of these same puzzles within the allotted time.

Figure 3A – Word Jumbles

JUMBLE ROUND 1
Number of Puzzles: 5
Minutes to Complete: 8
Start time: ___ : ___ End time: ___ : ___

On the Word Jumble Tasks, words can be of any kind (nouns, verbs, adjectives, adverbs, etc.). Verbs include all present, past, or future tenses. Words do NOT include: proper nouns, slang/jargon, or plural tenses.

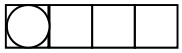
All solutions can be found in typical English dictionaries. The words in these exercises do not have any relation to one another (unless by coincidence).

In almost all instances, there is only one possible solution for each word. In rare cases, the fifth word of a puzzle (the Puzzle Solution) may have two possible solutions. In these cases, both solutions are acceptable.

Do NOT open the test booklet until the instructor gives you permission to.

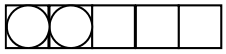
PUZZLE 1:

M I S L



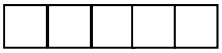
T A P H

C A K R T



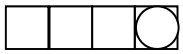
D A R N B

Scrambled 5th word: ___ ___ ___ ___ ___

Solution: 

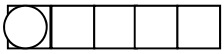
PUZZLE 2:

E A F S



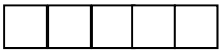
M A R F

O R Y T F



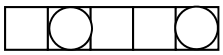
S O T U C

Scrambled 5th word: ___ ___ ___ ___ ___

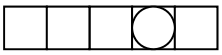
Solution: 

PUZZLE 3:

T I T A R



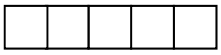
B R O H A



O R N O H

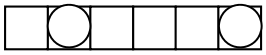
H I G E W

Scrambled 5th word: ___ ___ ___ ___ ___

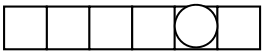
Solution: 

PUZZLE 4:

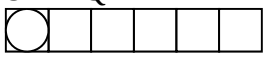
T U B E C K



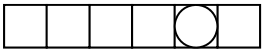
C I T O X E



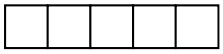
U N I Q T A



F A N E E D

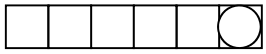


Scrambled 5th word: ___ ___ ___ ___ ___

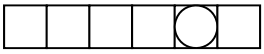
Solution: 

PUZZLE 5:

T O N B E N



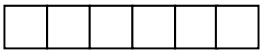
T O M S E D



K R A N E T

F L U E E Y

Scrambled 5th word: ___ ___ ___ ___ ___ ___

Solution: 

STOP! THIS IS THE END OF WORD JUMBLE ROUND 1.

Figure 3A (Continued) – Word Jumbles

JUMBLE ROUND 2
Number of Puzzles: 5
Minutes to Complete: 8
Start time: ___ : ___ End time: ___ : ___

On the Word Jumble Tasks, words can be of any kind (nouns, verbs, adjectives, adverbs, etc.). Verbs include all present, past, or future tenses. Words do NOT include: proper nouns, slang/jargon, or plural tenses.

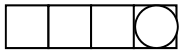
All solutions can be found in typical English dictionaries. The words in these exercises do not have any relation to one another (unless by coincidence).

In almost all instances, there is only one possible solution for each word. In rare cases, the fifth word of a puzzle may have two possible solutions. In these cases, both solutions are acceptable.

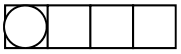
Do NOT open the test booklet until the instructor gives you permission to.

PUZZLE 1:

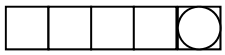
A C L P



D A I C

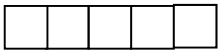


S L A R N



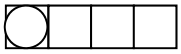
T O M C E

Scrambled 5th word: ___ ___ ___ ___ ___

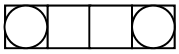
Solution: 

PUZZLE 2:

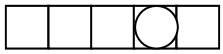
V O N E



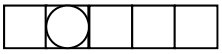
A M F E



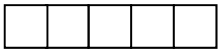
K I H C T



W O N C R

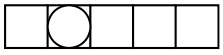


Scrambled 5th word: ___ ___ ___ ___ ___

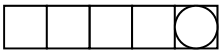
Solution: 

PUZZLE 3:

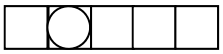
S E R E A



E N T G A

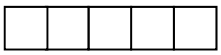


S C I B A



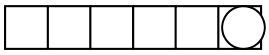
P R O W E

Scrambled 5th word: ___ ___ ___ ___ ___

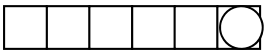
Solution: 

PUZZLE 4:

B R U Z Z E



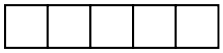
T A J E C K



M I N M U E

E N C O S H

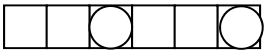
Scrambled 5th word: ___ ___ ___ ___ ___

Solution: 

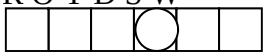
PUZZLE 5:

Y U C I J

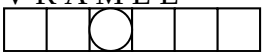
M I N O O T



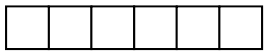
R O Y D S W



V R A M L E



Scrambled 5th word: ___ ___ ___ ___ ___ ___

Solution: 

STOP! THIS IS THE END OF WORD JUMBLE ROUND 2.

Figure 3A (Continued) – Word Jumbles

JUMBLE ROUND 3
Number of Puzzles: 5
Minutes to Complete: 8
Start time: ___ : ___ End time: ___ : ___

On the Word Jumble Tasks, words can be of any kind (nouns, verbs, adjectives, adverbs, etc.). Verbs include all present, past, or future tenses. Words do NOT include: proper nouns, slang/jargon, or plural tenses.

All solutions can be found in typical English dictionaries. The words in these exercises do not have any relation to one another (unless by coincidence).

In almost all instances, there is only one possible solution for each word. In rare cases, the fifth word of a puzzle may have two possible solutions. In these cases, both solutions are acceptable.

Do NOT open the test booklet until the instructor gives you permission to.

PUZZLE 1:

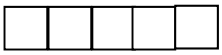
N E T U

I W V E

T H W A R

S S E C H

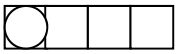
Scrambled 5th word: ___ ___ ___ ___ ___

Solution: 

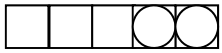
PUZZLE 2:

W L N A

K R A L

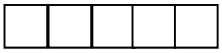


P R O N A



T H L O C

Scrambled 5th word: ___ ___ ___ ___ ___

Solution: 

PUZZLE 3:

U P T L I

N I P C A

A L T H O

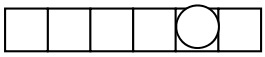
C H E E N

Scrambled 5th word: ___ ___ ___ ___ ___

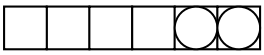
Solution: 

PUZZLE 4:

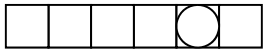
L U F T I E



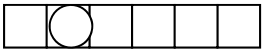
T I L P U F



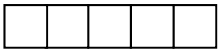
T E L L M A



R U B B U S

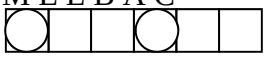


Scrambled 5th word: ___ ___ ___ ___ ___

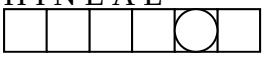
Solution: 

PUZZLE 5:

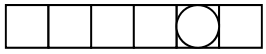
M E E B A C



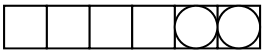
H I N E A L



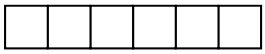
L I T G U Y



P L I T O E



Scrambled 5th word: ___ ___ ___ ___ ___ ___

Solution: 

STOP! THIS IS THE END OF ROUND 3.

Figure 4A – Answer Sheet

ANSWER SHEET.

Please record your answers for Word Jumble Rounds 1, 2, and 3 on this Answer Sheet. Answers written on the Jumble Worksheets will NOT receive credit. You must record your answers on this sheet DURING each 8-minute Round. You are not allowed to write on this Answer Sheet at any other time.

JUMBLE Round 1:

1.) ☐☐☐☐☐

2.) ☐☐☐☐☐

3.) ☐☐☐☐☐

4.) ☐☐☐☐☐

5.) ☐☐☐☐☐☐

JUMBLE Round 2:

1.) ☐☐☐☐☐

2.) ☐☐☐☐☐

3.) ☐☐☐☐☐

4.) ☐☐☐☐☐

5.) ☐☐☐☐☐☐

JUMBLE Round 3:

1.) ☐☐☐☐☐

2.) ☐☐☐☐☐

3.) ☐☐☐☐☐

4.) ☐☐☐☐☐

5.) ☐☐☐☐☐☐

Figure 5A – Answer Key

Instructions for grading:

* Compare the answers below to your Answer Sheet. If it exactly matches, give yourself credit for one completed puzzle. There is no partial credit for uncompleted puzzles or incorrect solutions.
* For each round, enter the TOTAL number of puzzles you successfully completed on your pay slip. Do not enter it on this sheet, which you will put into the bin after you are finished grading.
* Use the pay slip to calculate your pay for each round.

PUZZLE SOLUTIONS for each Word Jumble Puzzle.

Round 1:
PUZZLE 1: START
PUZZLE 2: FEAST
PUZZLE 3: WORTH *or* THROW
PUZZLE 4: QUIET *or* QUITE
PUZZLE 5: FASTEN

Round 2:
PUZZLE 1: CLAMP
PUZZLE 2: FORCE
PUZZLE 3: WATER
PUZZLE 4: COURT
PUZZLE 5: WINTRY

Round 3:
PUZZLE 1: TWINE
PUZZLE 2: CLOWN
PUZZLE 3: LOCAL
PUZZLE 4: FLUTE
PUZZLE 5: BATTLE *or* TABLET

Figure 6A – Pay Slip (Do Your Best Treatment)

Pay Slip
Seat #

TASK PERFORMANCE:

Task	**Results**
Jumble Task Rd 1	____ correct
Jumble Task Rd 2	____ correct
Jumble Task Rd 3	____ correct

Total Number Correct: [ ]

TASK PAY:

Task	**Pay**
Jumble Task Rd 1-3	$18
Situation Task	$ _____

TOTAL Pay:
($18 plus pay for the Situation Task) [ ]

Figure 6A (continued) – Pay Slip (Goal Treatment)

Pay Slip
Seat #

TASK PERFORMANCE:

Task	Results	Goal
Jumble Task Rd 1	____ correct	Goal Met? Y/N
Jumble Task Rd 2	____ correct	Goal Met? Y/N
Jumble Task Rd 3	____ correct	Goal Met? Y/N

Total Number of Times You Met the Goal:
(0, 1, 2, or 3)

TASK PAY:

Task	Pay
Jumble Task Rd 1-3	$18
Situation Task	$_____

Total Pay:
($18 + pay for Situation Task)

Figure 6A (continued) – Pay Slip (Piece Rate Treatment)

Pay Slip
Seat #

TASK PERFORMANCE:

Task	**Results**
Jumble Task Rd 1	____ correct
Jumble Task Rd 2	____ correct
Jumble Task Rd 3	____ correct

Total Performance: [ ]

TASK PAY:

Task	**Pay**
Jumble Task Rd 1-3	$_____ ($3 x Total # Correct)
Situation Task	$ _____

Total Pay: [ ]
(sum of earnings in jumble task rds 1-3 and earnings from the situation task)

Figure 7A – Risk Task

Situation Task

For each Situation in this table, please select either the fixed payment or the lottery (each Situation corresponds to a single row in the table). For example, in Situation 1 you are choosing between a fixed payoff of $0.25, and a lottery with a 50% chance of a $4 payoff, and a 50% chance of a $0 payoff. Once you have made a selection for all 15 Situations, we will select one of the 15 Situations for payment. If you chose the Fixed Payment for that Situation, you will receive the amount listed in the table. If you chose the Lottery for that Situation, I will use a random-number generator in PowerPoint to determine whether the result of the Lottery is $4.00 or $0.00.

"Situation" Task

Situation	Lottery	Fixed Payoff	I choose:
1	50% Chance $4.00; 50% Chance $0.00.	$0.25	☐ Lottery ☐ Fixed Payment
2	50% Chance $4.00; 50% Chance $0.00.	$0.50	☐ Lottery ☐ Fixed Payment
3	50% Chance $4.00; 50% Chance $0.00.	$0.75	☐ Lottery ☐ Fixed Payment
4	50% Chance $4.00; 50% Chance $0.00.	$1.00	☐ Lottery ☐ Fixed Payment
5	50% Chance $4.00; 50% Chance $0.00.	$1.25	☐ Lottery ☐ Fixed Payment
6	50% Chance $4.00; 50% Chance $0.00.	$1.50	☐ Lottery ☐ Fixed Payment
7	50% Chance $4.00; 50% Chance $0.00.	$1.75	☐ Lottery ☐ Fixed Payment
8	50% Chance $4.00; 50% Chance $0.00.	$2.00	☐ Lottery ☐ Fixed Payment
9	50% Chance $4.00; 50% Chance $0.00.	$2.25	☐ Lottery ☐ Fixed Payment
10	50% Chance $4.00; 50% Chance $0.00.	$2.50	☐ Lottery ☐ Fixed Payment
11	50% Chance $4.00; 50% Chance $0.00.	$2.75	☐ Lottery ☐ Fixed Payment
12	50% Chance $4.00; 50% Chance $0.00.	$3.00	☐ Lottery ☐ Fixed Payment
13	50% Chance $4.00; 50% Chance $0.00.	$3.25	☐ Lottery ☐ Fixed Payment
14	50% Chance $4.00; 50% Chance $0.00.	$3.50	☐ Lottery ☐ Fixed Payment
15	50% Chance $4.00; 50% Chance $0.00.	$3.75	☐ Lottery ☐ Fixed Payment

Figure 8A – Questionnaire Task

Instructions: Please answer the following questions about yourself to the best of your ability.

1. Age: ___

2. Gender (circle one)

   •Male
   •Female

3. Race (circle one)

   •Asian
   •Black
   •Caucasian
   •Hispanic
   •Other (specify): _____

4. Which university do you attend? (circle one)

   •Boston College
   •Boston University
   •Harvard
   •MIT
   •Northeastern
   •Tufts
   •Umass
   •Other (specify): _____

5. What year in school are you? (circle one)

   •Freshman
   •Sophomore
   •Junior
   •Senior
   •None of the Above / Other (Specify): _____

6. What do you expect your salary to be in 5 years?
   $ _____ (nearest $5,000)

7. What is your declared undergraduate major or intended major?
   (e.g. Mathematics, English, Physics, History, etc.).

   A: _____

8. Do you intend to pursue a graduate degree? (Y / N) If so, in which field?
(e.g. Medicine, Law, Business, Engineering, Chemistry, Literature, History, etc.).

   A: _____

9. What kind of company or organization do you expect to work for after graduation? (circle one)

   •Traditional Manufacturing (automotive, chemical, aerospace, etc.)
   •Technology (hardware, software, technology services, etc.)
   •Financial Services (banking, insurance, investment products, etc.)
   •Other Professional Services (legal, consulting, accounting, real estate, etc.)
   •Arts / Entertainment / Media / Journalism
   •Health Care
   •Education
   •Government Services (including military)
   •None of the Above
   •Don't Know

10. What kind of job do you expect to have after graduation? (circle one)

    •Engineer or Computing / Technical Professional
    •Management
    •Sales
    •Marketing
    •White-Collar (banking, consultant, lawyer)
    •Doctor or other Health Care Professional
    •Professor or Teacher
    •Researcher
    •Administrative / Clerical
    •Artist / Entertainer / Writer
    •None of the Above
    •Don't Know

11. What do you think today's experiment was measuring?

    A: _____
    _____
    _____

12. In one or two words, please describe your feelings about the word Jumble Task:

    A: _____

13. Approximately how many CLER studies have you participated in prior to this one?

    A: _____

Below are statements describing people's behaviors and attitudes. Please use the rating scale below, which ranges from Strongly Disagree to Strongly Agree, to describe the degree to which you agree or disagree with each statement. There are no right or wrong answers. Please read each statement carefully, and then fill in the corresponding bubble for your choice.

Note: Some statements refer specifically to you; in these instances, describe yourself as you generally are now (not as you wish to be in the future) and as you honestly see yourself (in relation to other people you know of the same gender and general age as you). Other statements do not refer directly to you, but instead represent a commonly held opinion, some of which you will agree with and others you will disagree with. For all types of statements, simply indicate the degree to which you agree or disagree.

	Statement:	Disagree 1	2	Neutral 3	4	Agree 5
1.	I see myself as someone who is reserved and quiet.	o	o	o	o	o
2.	It is never necessary to sacrifice the welfare of others.	o	o	o	o	o
3.	It is very easy for me to admit when I make a mistake.	o	o	o	o	o
4.	I am poor at making financial decisions compared to the average student.	o	o	o	o	o
5.	The word jumble task was not set up in a fair way.	o	o	o	o	o
6.	Moral standards are not personal rules, and thus can be used to judge others.	o	o	o	o	o
7.	If I have done well on a job or assignment, I insist on a reasonable reward.	o	o	o	o	o
8.	The dignity and welfare of people should not be the most important concern in any society.	o	o	o	o	o
9.	The word jumble consisted of reasonable expectations.	o	o	o	o	o
10.	I am worse at standardized tests than most of my friends.	o	o	o	o	o
11.	I see myself as someone who is bold and energetic.	o	o	o	o	o
12.	People should ensure their actions never intentionally harm another even minutely.	o	o	o	o	o
13.	The word jumbles task was frustrating.	o	o	o	o	o
14.	I see myself as someone who is moody and temperamental.	o	o	o	o	o
15.	No matter what I do, I always take great pride in the outcome.	o	o	o	o	o
16.	There are ethical principles so important they should be in all codes of ethics.	o	o	o	o	o
17.	I see myself as someone who is generally trusting.	o	o	o	o	o
18.	Ethics in interpersonal relations are so complex that individuals should formulate their own codes.	o	o	o	o	o

19.	I am worse at spelling than my peers.	o	o	o	o	o
20.	The word jumbles task was fun.	o	o	o	o	o
21.	I see myself as someone who is sympathetic and warm.	o	o	o	o	o
22.	One should never psychologically or physically harm another person.	o	o	o	o	o
23.	I do not like it when friends or colleagues see I've done a poor job.	o	o	o	o	o
24.	Different sets or types of moralities cannot be compared as to their "rightness."	o	o	o	o	o
25.	I see myself as someone who has few artistic interests.	o	o	o	o	o
26.	What one person considers moral may be immoral to another.	o	o	o	o	o
27.	I am a better driver than most people.	o	o	o	o	o
28.	I get frustrated easily.	o	o	o	o	o
29.	I see myself as someone who is outgoing and sociable.	o	o	o	o	o
30.	I feel I did a great job in today's experiment.	o	o	o	o	o
31.	I see myself as someone who is dependable and systematic.	o	o	o	o	o
32.	In general, it is important to me to be reasonably rewarded based on the amount of effort I put forth.	o	o	o	o	o
33.	I see myself as someone who is disorganized and careless.	o	o	o	o	o
34.	I tend to get frustrated when I can’t finish a task.	o	o	o	o	o
35.	Risks to another, no matter how small, should never be tolerated.	o	o	o	o	o
36.	I see myself as someone who is rude and harsh.	o	o	o	o	o
37.	Whether a lie is moral or immoral does not depend upon the circumstances.	o	o	o	o	o
38.	I see myself as someone who is bashful and shy.	o	o	o	o	o
39.	I enjoyed the challenge of the word jumbles.	o	o	o	o	o
40.	Rigid ethical positions can stand in the way of bettering human relations.	o	o	o	o	o
41.	Even when facing a challenging situation, I tend to stay calm and unemotional.	o	o	o	o	o
42.	I am better at very difficult tasks than most of my friends.	o	o	o	o	o
43.	In general, it is of little consequence to me if I am not fairly rewarded for a personal accomplishment.	o	o	o	o	o

44.	If an action could harm an innocent other, it should not be done.	o	o	o	o	o
45.	I am proud of the number of Word Jumbles I was able to solve in today's experiment.	o	o	o	o	o
46.	I see myself as someone who is unenvious of others.	o	o	o	o	o
47.	I know I could have done better in today's experiment than I actually did.	o	o	o	o	o
48.	Deciding an action by balancing the positive and negative consequences is not immoral.	o	o	o	o	o
49.	If I have performed poorly on a task, it is only fair that I am rewarded poorly.	o	o	o	o	o
50.	I have better decision making skills than most people	o	o	o	o	o
51.	Eight minutes was not a fair amount of time for each round of the word jumbles.	o	o	o	o	o
52.	Moral actions are not those that closely match ideals of the "perfect" action.	o	o	o	o	o
53.	I do not take it personally when I've done a bad job or made a mistake	o	o	o	o	o
54.	I see myself as someone who is relaxed and handles stress well.	o	o	o	o	o
55.	I do not feel hassled when performing a task under time pressure.	o	o	o	o	o
56.	General rules concerning lying can be formulated, as it does not vary between situations.	o	o	o	o	o
57.	What is ethical does not vary from one situation or society to another.	o	o	o	o	o
58.	I see myself as someone who is conventional and uncreative.	o	o	o	o	o
59.	I see myself as someone who tends to find fault with others.	o	o	o	o	o
60.	I was fairly paid for my work on the word jumble task.	o	o	o	o	o
61.	Determining what is ethical for everyone is impossible since it is up to the individual.	o	o	o	o	o
62.	I see myself as someone who is complex and open to new experiences.	o	o	o	o	o
63.	I see myself as someone who tends to be lazy.	o	o	o	o	o
64.	Potential harm to others is not always wrong; it depends on the benefits gained.	o	o	o	o	o
65.	I see myself as someone who gets nervous easily.	o	o	o	o	o
66.	The time limit on the word jumbles task was annoying to me.	o	o	o	o	o
67.	One may at times perform an action that might threaten the dignity or welfare of another.	o	o	o	o	o

68.	I am embarrassed about my performance in today's Word Jumble task.	o	o	o	o	o
69.	I see myself as someone who has an active imagination.	o	o	o	o	o
70.	I see myself as someone who does a thorough job.	o	o	o	o	o
71.	I am afraid of looking foolish in social situations.	o	o	o	o	o

---

14. Do you have any further comments about today’s session?

    A: _____
    _____
    _____