Author/s:
Ranasinghe, Ajanee Bhagya

Title:
Designing Incentives to Elicit Creativity

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/241718

Designing Incentives to Elicit Creativity

Ajanee Bhagya Ranasinghe

ORCID identifier: 0000-0002-4952-6498

Doctor of Philosophy

August 2020

Department of Accounting, Faculty of Business and Economics

The University of Melbourne

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

In this thesis, I investigate the effectiveness of incentives for creative tasks. While organizations value and pursue creativity, they must also promote the efficient use of limited resources. Prior research finds that incentives, despite being a key control mechanism, are ineffective for simultaneously encouraging both creativity and efficiency. However, this relatively recent literature has not yet examined how critical attributes relating to individuals, tasks, and incentives may influence incentive efficacy. I conduct two studies to examine the differential effects of combined creativity and quantity incentives for (1) individuals with different creative capacities, (2) distinct types of creative tasks, and (3) varying strengths of incentives.

In the first study, I focus on individual creative capacity, which is a crucial, yet unexplored characteristic of individuals engaged in creative tasks. Using an experiment, I hypothesize and find that combined creativity and quantity incentives lead to sustained high creative performance for individuals with high creative capacity. Thus, contrary to prior beliefs that combined creativity and quantity incentives are ineffective, I show that they can be effective for the type of individuals typically employed for creative work.

In the second study, I examine creative tasks with pre-specified problems, which is a prevalent task type that is under-studied in the prior literature. Contrary to expectations, my experiment results show that combined creativity and quantity incentives can be detrimental to creative tasks with pre-specified problems. Comparing my study with prior studies on pre-specified problems, I note that incentive efficacy appears to be sensitive to incentive strength. Therefore, I design a follow-up experiment to examine the effect of different incentive strengths on the performance in a pre-specified task. I find that, for pre-specified tasks, combined creativity and quantity incentives are detrimental to performance, regardless of

incentive strength, although the effect is predictably lower with weak incentives. This shows that combined creativity and quantity incentives are not just ineffective, as found in the prior literature, but can impede creative performance in tasks with pre-specified problems. Furthermore, weaker incentives may alleviate some of the negative effects of incentives on performance.

Overall, my thesis highlights the importance of having a more granular understanding of incentive design in the creativity sphere. It extends prior research by demonstrating that incentives can lead to higher or lower creative performance depending on characteristics of individuals, tasks, and incentives. This highlights the importance of future studies on the use of performance management and rewards to elicit creative performance.

Declarations

This is to certify that

(i) this thesis comprises only my original work towards the Doctor of Philosophy;

(ii) due acknowledgement has been made in the text to all other material used; and

(iii) this thesis is fewer than the maximum word limit (100,000 words) in length, exclusive of tables, maps, bibliographies and appendices.

Signature ……

Ajanee Bhagya Ranasinghe

Acknowledgements

I was fortunate to receive supervision from a remarkable group of researchers. I am grateful to Professor Margaret Abernethy, who was both my supervisor and Doctoral Program Director, for her support in all the many challenges in my PhD journey; Professor Anne Lillis for her meticulous comments packaged lovingly; Professor Jennifer Grafton for valuable feedback and help with presentations; Professor Donald Moser for working so closely with me across time zones that I often forgot the 16,000 kilometers between us; and Professor Naomi Soderstrom, Committee Chair, for being a loving and supportive presence. Their mentorship has helped me develop both as a researcher and as a person, for which, I will be eternally grateful.

My research was made possible by generous funding from the Department of Accounting at The University of Melbourne. I thank the funding support by Professor Matthew Pinnuck, Head of Department, for my experiments, conferences, workshops, doctoral consortia, and a four-month visit to the University of Pittsburgh, USA. I have grown much as a researcher from all these experiences. My gratitude also goes to the Faculty of Business and Economics, and the Australian Government Research Training Program for funding my doctoral studies and research travel. My research also benefitted from the resources of the Nectar Research Cloud, backed by the National Collaborative Research Infrastructure Strategy.

I sincerely thank the thoughtful comments from: Michael Davern; Martin Weisner; other (visiting) faculty at the Department; Steven Kachelmeier; faculty and doctoral students at the University of Pittsburgh; participants in the doctoral symposiums of the American Accounting Association 2019, the Accounting Workshop for Experiments 2018, 2019, and the Accounting and Finance Association of Australia and New Zealand 2017; and seminar participants at Tilburg University, University of Amsterdam, IESEG School of Management, Warwick Business School and the London School of Economics.

It was a pleasure working alongside my PhD colleagues at the Department. I am grateful for the friends in Kerui, Sabutay and Hrishikesh. We started together and made it through together! I am so proud of us, and excited for the next steps. I also wish to thank: past students Shaffique, Ra-Pee, Ruidi, Sujay, Kevin, Mark, Wei (Li), Wei (Zeng), Sharon, Michelle and Yan, for showing me the ropes when I first joined the doctoral program; and the current group Chris, James, Jin, Yunhe, Xing, Fei, Jingru, Jenny, Anson and Junyoung, for the insightful discussions, fun chats, good food and generally good times.

Much support for my PhD journey came in the familiar form of friends. I wish to thank Achini, for being my friend, doctor (the medical kind), test subject, proof-reader, and anything else I asked of her; Pete, for advice from his own PhD journey, help with my work, quality humor and companionship; Nimna, for appearing suddenly in my life to become a close and supportive friend; Damayanthi, for listening to my complaints despite her own PhD worries; Kedar, for inspiring me to keep improving myself; Azmi, Pasan, Dimithri and Damith for coding support; Amy, for keen help with my experiments; Sampath, for good naturedly offering help whenever I asked; and Ajahn Hassapagno from the Vimokkharam Forest Monastery, for his wise guidance when I needed it the most.

The unfailing support that saw me through the PhD came from my family. My mother, Malanee Ranasinghe, was my closest confidante and got me back on my feet each time I fell. My father, Sujanaratna Ranasinghe, was my constant reminder to be excited about life. Living in Melbourne would not have been the same without my sister, Wasanee, and brother-in-law, Suranga. In many ways, they made up for what I missed about Sri Lanka.

Lastly, as I write this, creative minds around the world are working tirelessly to help humanity cope with one of the worst global pandemics in living memory. I am deeply grateful to all those unsung heroes of the crisis who remind me of the true value and power of creativity.

උගතෙකු යන්නෙහි සැබෑ අරුත
මට මුලින්ම පෙන්වාදුන්
අම්මාට

To my mother,
who showed me the ABCs
of being a true scholar

Contents

List of Tables and Figures

Chapter 2

Chapter 3

List of Abbreviations

FD	Full Duration
FH	First Half
HCC	High Creative Capacity
KRW	Kachelmeier, Reichert, and Williamson (2008)
KW	Kachelmeier and Williamson (2010)
LCC	Low Creative Capacity
SH	Second Half

Chapter 1
Introduction

This thesis examines the effectiveness of including creativity as a performance dimension in an incentive contract. In two separate studies, I examine how our prior understanding on the use of creativity measures in incentives can be advanced by considering individual traits, task characteristics, and incentive characteristics.

1.1. Background

Creativity – defined as the production of novel, potentially useful ideas (Amabile, Conti, Coon, Lazenby, & Herron, 1996; Malik & Butt, 2017; Shalley, Zhou, & Oldham, 2004) – is now a necessity for modern organizations to remain competitive (Amabile & Khaire, 2008; Unsworth, 2018). Both technology organizations (e.g. Google) and non-technology organizations (e.g. Walmart) explicitly highlight the importance of creativity as a crucial component of their strategy and workforce culture (Alphabet Inc., 2019; Walmart Inc., 2018), suggesting that creativity is ubiquitous.

Although organizations are keen to encourage creativity, they cannot afford it at all costs (Chang & Birkett, 2004; Grabner, 2014). When performing tasks, they must use limited resources to achieve multiple, and often contradictory goals (Holmstrom & Milgrom, 1991; Kachelmeier, Reichert, & Williamson, 2008; Kachelmeier & Williamson, 2010; Prendergast, 1999). Efficiency is one such conflicting goal for creativity, because attempts to be creative

include trial and error that require time and effort, which, in turn, reduces short-run efficiency.[1] Therefore, firms need to effectively communicate the desired trade-offs between creativity and efficiency to their employees.

While management control systems, particularly incentive contracts, are common tools used to communicate organizational strategy to employees, incorporating creativity into these tools is challenging due to some unique characteristics of creativity.[2] Creativity generally involves complex tasks that are difficult to specify and measure (Grabner, 2014; Kachelmeier et al., 2008). In addition, the relationship between input effort and outcome creativity is often unclear (Amabile et al., 1996; Kachelmeier et al., 2008). Creativity, in many cases, is also driven by intrinsic motivation, and explicit incentives may crowd out intrinsic motivation (Deci, 1971; Shalley et al., 2004).

There is little consensus among firms on how to encourage the desired trade-offs between creativity and efficiency. Furthermore, the use of innovation-specific performance metrics in performance evaluation and reward systems remains low (Ittner, Larcker, & Rajan, 1997; McKinsey & Company, 2012). The long held belief in the management and psychology literatures is that such metrics will crowd out intrinsic motivation (Deci, Koestner, & Ryan, 1999).[3] However, this literature does not take a commercial perspective and does not, therefore, consider incorporating other dimensions such as efficiency alongside creativity in the incentive structure (Kachelmeier et al., 2008; Unsworth & Luksyte, 2015). Nevertheless, the presence of other performance measures could render the crowding-out effect irrelevant. Recent studies in accounting do examine the use of incentives for creativity alongside other dimensions (e.g.

[1] Efficiency, often referred to as "doing things right", is defined as getting the most output from the least amount of inputs (e.g. Robbins & Coulter, 2012)

[2] The general usefulness of performance management systems to encourage performance on various dimensions is well documented in the literature. Ferreira and Otley (2009) provide an extensive review.

[3] The psychology literature has hotly debated whether explicit incentives on creativity lead to differences in creative performance. See Malik and Butt (2017) for a comprehensive review of this literature.

Kachelmeier et al., 2008; Kachelmeier, Wang, & Williamson, 2019; Kachelmeier & Williamson, 2010) and demonstrate that embedding creativity measures in incentives will not necessarily generate gains in highly creative output. But this finding is attributed to the notion that 'trying harder' does not necessarily lead to more creative performance, and not to crowding out of intrinsic motivation.

However, this relatively recent stream of literature on multidimensional incentives for creativity does not consider differential attributes relating to individuals, tasks, or incentives that could influence the relationship between incentives and performance.[4] In this thesis, I explore three such important attributes that can alter the implications of prior findings to theory and practice. In Study 1, I focus on creative capacity, which is a crucial, yet unexplored characteristic of individuals engaged in creative tasks. In Study 2, I examine creative tasks with pre-specified problems, which is an important task type that has been largely ignored in the prior literature. In the same study, I also look at the effects of different incentive strengths for tasks with pre-specified problems.

1.2. Summary of Study 1

Study 1 examines the effectiveness of incentivizing "efficient creativity", by way of creativity-weighted-quantity incentives, for individuals with different creative capacities. While early studies find that creativity-weighted-quantity incentives lead to no improvement in creative performance, and are detrimental to the overall joint performance on creativity and quantity (hereafter referred to as "overall performance"), they do not consider the effect of an individual's capacity to be creative (e.g. Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010; Liu, Ditillo, & Van den Abbeele, 2019). This is despite the

[4] Consistent with the prior literature, I refer to 'multidimensional incentives' as incentives that are based on multiple dimensions of performance (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010). This is not to be confused with various attributes of incentives such as incentive form, frequency, strength, and horizon.

high likelihood that, in practice, it is individuals with high creative capacity who are being selected for such tasks (Hunter, Cushenbery, & Friedrich, 2012).

To test the differences in incentive efficacy across individuals with different creative capacities, I conduct a laboratory experiment with student participants. I measure *individual creative capacity* prior to the experiment (using a self-rated measure on creative self-efficacy) and manipulate *incentive type* (quantity-only incentives or creativity-weighted-quantity incentives). The creative task is adapted from Kachelmeier et al. (2008). Participants construct rebus puzzles – riddles in which words and/or diagrams are used to represent a familiar word or phrase (Kachelmeier et al., 2008) – during a 20-minute period. I capture the quantity of puzzles created by participants as well as the creativity of their puzzles, as evaluated by an independent panel of raters.

Drawing on psychology and economics literatures, I predict and find that creativity-weighted-quantity incentives lead to an improvement in creativity for individuals with high creative capacity (HCC), but not for individuals with low creative capacity (LCC). Specifically, the HCC group has stronger creative performance under creativity-weighted-quantity incentives (compared to quantity-only incentives), both initially and later in the 20-minute task. In contrast, the LCC group under creativity-weighted-quantity incentives (compared to quantity-only incentives) has a burst of creative performance initially but is unable to sustain this performance as time passes. For the pooled sample of HCC and LCC groups, I replicate the prior finding that creativity-weighted-quantity incentives are no better than quantity-only incentives for the number of highly creative puzzles (Kachelmeier et al., 2008). The findings are robust to different thresholds for highly creative puzzles and creativity rating quality. The

key findings are not affected by excluding those who fail manipulation checks or including a range of covariates.[5]

The findings provide an important qualification to previously held beliefs about the efficacy of creativity incentives. Contrary to prior beliefs, I show that creativity-weighted-quantity incentives can elicit higher creative performance from individuals with HCC. This finding is especially important because firms are more likely to employ individuals with HCC for creative roles (Hunter et al., 2012).

I am also the first to capture the time taken for each creative idea, providing less noisy data on "initial" versus "later" creativity. Time is a scarce resource to be managed, and employees are required to be mindful of time when producing creative work (Kachelmeier & Williamson, 2010). In this study I measure the time taken to produce each unit of output, and this allows me to observe, with more precision, whether creative performance sustains over time. This approach sets a benchmark for future studies on creativity, by identifying the importance of capturing performance over time. The findings also contribute to the creativity literature in management and psychology by shedding light on interactive effects of individual and contextual factors on creative performance (Shalley et al., 2004).

Finally, the study has some important implications for practice. Rewards for creative performance, alongside efficiency performance, can be effective if employees have HCC, but not if they have LCC. Furthermore, I identify an established scale that can effectively capture an individual's creative capacity, and this can be used to sort individuals in the employee selection and incentive design processes. This kind of artificial screening based on personality

[5] Covariates include demographic characteristics (e.g. gender, age, level of education, language background), individual characteristics and preferences (e.g. drawing skills, prior experiences with rebus puzzles, ambiguity aversion, risk aversion, and preferences for workplace structure) and self-evaluations about performance in the task (e.g. intrinsic motivation, overconfidence, and efforts made to increase creativity and quantity).

characteristics, such as the use of assessment centers, are already common in employee recruitment (Diekmann & König, 2015).

1.3. Summary of Study 2

Study 2 examines the effectiveness of creativity-weighted-quantity incentives for (1) creative tasks with pre-specified problems, and (2) different levels of incentive strength. While prior research finds that creativity-weighted-quantity incentives do not improve performance, different types of creative tasks have not been considered. This is despite the repeated calls to recognize the diversity of tasks labelled 'creative' (e.g. Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015).

In this study, I draw on prior creativity research that differentiates closed tasks (i.e. those that require creative solutions to already-specified problems) from open tasks (i.e. those that require generating creative ideas without pre-specified problems) (Charness & Grieco, 2018; Unsworth, 2001; Unsworth & Luksyte, 2015). Even though closed tasks are widely prevalent in practice (Unsworth, 2001), there is little research on the topic.

To test the differences in incentive efficacy across different types of creative tasks, I conduct an experiment manipulating the *creative task type* (closed task or open task) and *incentive type* (quantity-only incentives or creativity-weighted-quantity incentives). The open task is identical to the task in Study 1. The closed task also requires participants to construct rebus puzzles, but for specified word-phrases. This reflects the "pre-specified problem" feature of a closed task.

I predict that, for closed tasks, where uncertainty is reduced (Charness & Grieco, 2018; Unsworth & Luksyte, 2015) and there is greater analyzability of search procedures (Perrow, 1970), creativity-weighted-quantity incentives will be more effective compared to open tasks. However, I find the opposite effect. Specifically, creativity-weighted-quantity incentives, when

compared to quantity-only incentives, lead to significantly lower creative and overall performance for the closed task. Participants' self-ratings of task complexity or difficulty do not explain this surprising result. For the open task, I replicate prior studies that creativity-weighted incentives neither improve nor impede creative performance, on average (Kachelmeier et al., 2008), but increase creative performance for the subset of individuals with high creative capacity (Study 1). The findings are robust to alternative measures of highly creative puzzles and quality of creativity ratings. The key findings are not affected when I exclude manipulation check failures or include covariates.[6]

To investigate the unexpected result for closed tasks, I conduct a follow-up experiment with the same incentive types, but with two levels of *incentive strength* (high or low performance-based incentives, compared to base pay). This is motivated by the observation of the different strengths of incentives used in prior studies of closed tasks, with their associated mixed findings. Due to subject pool constraints, I conduct this experiment with limited student participants, and combine the data with comparative data from the initial experiment.

Based on behavioral theories of the negative effects of high-powered incentives, I expect weak incentives to reduce the negative effect of incentives on performance. I find this to be true for overall performance, but not for creative performance. In addition, quantity-only incentives outperform creativity-weighted-quantity incentives for closed tasks, regardless of incentive strength. This suggests that incentive strength is only a partial explanation for the results of the initial experiment, and the mixed findings on creativity incentives for closed tasks in the prior literature.

This study contributes to the literature on incentivizing creativity in several ways. It is among the few to explicitly compare incentive efficacy for open and closed creative tasks, and

[6] All covariates for Study 1, mentioned in footnote 5, were used in Study 2, except preferences for work-place structure. This was not collected in the post-experiment questionnaire for Study 2.

the first to study multidimensional incentives, such as creativity-weighted-quantity incentives, for closed tasks. Previously, published studies suggest that creativity-weighted-quantity incentives do not lead to more creative ideas, but do not impede creativity either (e.g. Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). However, I show that creativity-weighted-quantity incentives can negatively influence creative performance for closed tasks. This is vital because closed tasks are commonplace in many areas such as design, technology, manufacturing and marketing (Unsworth & Luksyte, 2015).

I also reconcile some mixed findings in the literature on the effectiveness of creative incentives for closed tasks. Some studies find a positive relationship between creative incentives and performance on a creative task (e.g. Charness & Grieco, 2018; Liu et al., 2019), while others, including the initial experiment in this study, show a negative relationship (e.g. Ariely, Gneezy, Loewenstein, & Mazar, 2009). I contribute to the literature by showing that incentive strength, another important management control choice available within firms, can be a partial reason for these mixed findings. This opens avenues for future studies to explore other contributory factors for the mixed findings.

1.4. Thesis outline

Chapters 2 and 3 are designed to stand alone and can be read independently. The footnotes, appendices, figures and tables are numbered specific to each chapter. Chapter 2 presents Study 1 on the effectiveness of creativity-weighted-quantity incentives for individuals with different creative capacities. Chapter 3 presents Study 2, on the effectiveness of creativity-weighted-quantity incentives for closed tasks at different levels of incentive strength. Chapter 4 synthesizes the results and conclusions of Study 1 and Study 2, and provides a discussion of contributions and suggestions for future research.

Chapter 2
Creative capacity, multidimensional incentives and creative performance

Abstract

I conduct an experiment to examine whether an individual's creative capacity – the potential to produce creative outcomes – influences the efficacy of creativity-weighted-quantity incentives. To remain commercially viable, organizations desire creativity while also maintaining efficiency. The challenge is to design a management control system to achieve a trade-off between these desired dimensions. Prior research finds that incentivizing both creativity and quantity of production is detrimental to quantity and has no effect on creativity. However, these studies do not consider the creative capacity of individuals. Drawing on psychology and economics literatures, I predict and find that creativity-weighted-quantity incentives lead to an improvement in creativity for individuals with high creative capacity, but not for individuals with low creative capacity. The creative performance is sustained over time for individuals with high creative capacity but declines for those with low creative capacity. These results suggest that creativity-weighted-quantity incentives are effective for the types of individuals who are typically employed for creative work.

Key words: Creative capacity, creative self-efficacy, creativity, quantity, incentives

2.1. Introduction

Creativity and innovation are a crucial source of competitive advantage in rapidly changing business environments, and organizations are increasingly attempting to incorporate them into business strategies and cultures (e.g. Alphabet Inc., 2019; Toyota Motor Corporation, 2019).[1] However, "art for art's sake" is rarely the end-goal for most organizations that desire to promote creativity while maintaining efficiency (Chang & Birkett, 2004; Grabner, 2014).[2] Designing a management control system to achieve this goal can be challenging. Prior research on multidimensional incentives, such as creativity-weighted-quantity incentives (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010), concludes that incentives lead to no improvement in creative performance and are detrimental to the overall joint performance on creativity and quantity (hereafter referred to as "overall performance").[3]

It is likely that the effectiveness of incentives depends on the individuals' capacity to be creative (e.g. Tierney & Farmer, 2002). Creative personality characteristics have received considerable attention in the psychology literature, using a variety of measures to capture personality types (e.g. Costa Jr & McCrae, 2008; Gough, 1979; Kirton, 1976; Tierney & Farmer, 2002). Despite this, there is a dearth of studies on the interactive effects between creative personality characteristics and rewards (Shalley et al., 2004). While there is some empirical evidence that incentive effectiveness differs for individuals with different levels of

[1] Creativity, is defined as the production of novel, potentially useful ideas, while innovation is defined as the development and successful implementation of novel and useful ideas (Amabile et al., 1996; Malik & Butt, 2017; Shalley et al., 2004). While I view creativity and innovation as distinct constructs, I will refer to relevant literature from both areas throughout the chapter because creativity is an integral part of innovation.

[2] Efficiency, often referred to as "doing things right" is defined as getting the most amount of outputs from the least amount of inputs (e.g. Robbins & Coulter, 2012). I acknowledge that creativity and efficiency are not always competing factors. Indeed, they can be complements in the long run if creative ideas lead to efficiency improvement. The resource trade-off discussed in this study is based on the trial and error nature of creative idea generation. With limited resources such as time, effort, and money, individuals frequently need to make trade-offs between creativity and efficiency in the idea generation process. The focus of the study is on this trade-off.

[3] One exception is the suggested usefulness of creativity-weighted-quantity incentives as a tool for attracting individuals with high creative ability under certain conditions (Kachelmeier & Williamson, 2010). However, this is not a performance-inducing characteristic of the incentive contract.

creative self-efficacy (e.g. Malik, Butt, & Choi, 2015), prior studies do not consider creative capacity when examining the effectiveness of multidimensional incentives such as creativity-weighted-quantity incentives. This is despite employees hired for creative roles (e.g. architects, researchers, employees of innovation labs) typically having a high capacity to be creative (e.g. Hunter et al., 2012; Tierney & Farmer, 2002). The primary goal of my study is to understand the effectiveness of creativity-weighted-quantity incentives for individuals with different creative capacities.

Incentives are only effective if an individual possesses the capacity, ability and knowledge to perform the activity (Locke, 1968). Individuals with high creative capacity (HCC) have the capacity to produce creative outcomes if induced to do so. Therefore, I argue that individuals with HCC will be able to effectively respond to creativity-weighted-quantity incentives by producing creative outcomes, which will compensate for any efficiency loss in the process of attempting to be creative. The creativity component of the incentives will encourage them to utilize their creative potential. In contrast, if only paid for quantity, HCC individuals will primarily focus on output quantity, even though they have the potential to be more creative.

For individuals with low creative capacity (LCC), I argue that creativity-weighted-quantity incentives can be detrimental. For them, trying to be creative takes up valuable time, but does not necessarily lead to better creativity. Thus, there will be an efficiency loss, with no corresponding creativity gain. If individuals with LCC are only paid for quantity, this efficiency loss can be avoided. Therefore, for individuals with LCC, creativity-weighted-quantity incentives, when compared to quantity incentives, are expected to have no difference on creative performance and a reduction in overall performance.

Furthermore, when performing a task repeatedly over a period, I expect the creative and overall performance of individuals with HCC to be relatively stable (Feist, 1998). However, for individuals with LCC, I expect that, even though creativity-weighted-quantity incentives could encourage some early creative performance (and consequently, overall performance), they will be unable to sustainably produce creative ideas as time passes. The unsuccessful effort invested in creativity is expected to divert effort from quantity as well. Overall, for individuals with LCC, creative performance (and consequently, overall performance) will likely suffer over time.

To test these predictions, I conduct a 2x2 between-subject experiment. I adopt the creative task from Kachelmeier et al. (2008) (hereafter referred to as KRW). Participants construct rebus puzzles – a riddle in which words and/or diagrams are used to represent a familiar word or phrase (Kachelmeier et al., 2008) – in a 20-minute period. This task allows the measurement of both the creativity and quantity of the puzzles produced. The task is performed on touch-enabled computer screens, which allows the recording of the exact time each puzzle is submitted, facilitating a more precise analysis of puzzles produced at initial and later stages of the experiment. Independent raters, recruited through the online crowdsourcing platform Prolific, rate the creativity of the puzzles.

The key variables in the experiment are *incentive type* (manipulated) and *individual creative capacity* (measured). *Incentive type* is manipulated as quantity-only incentives, and creativity-weighted-quantity incentives.[4] Quantity-only incentives are based on the number of puzzles produced. Creativity-weighted-quantity incentives are based on a creativity-weighted total score, which is the sum of creativity ratings of all puzzles produced per participant. For both conditions, an approximate average pay of $25 per person is maintained. *Individual*

[4] See Appendix A for details on experiment conditions and variables, and Appendix B for the complete experiment instrument.

creative capacity is captured by a widely-used instrument for creative self-efficacy developed by Tierney and Farmer (2002). Data for this instrument was collected from the participants prior to the experiment task.

Based on the experiment data, I find that if individuals have HCC, then creative performance can be improved through creativity-weighted-quantity incentives. The HCC group under creativity-weighted-quantity incentives has higher creative performance, when compared to the HCC group with quantity-only incentives, and the difference is more evident as time passes. Individuals in the HCC group successfully manage the trade-offs between creativity and quantity, regardless of the incentives. In the absence of creativity-weighted-quantity incentives, individuals with HCC have significantly lower creative performance over time. For individuals with LCC, the effectiveness of creativity-weighted-quantity incentives is less clear. Creativity-weighted-quantity incentives encourage some early creativity, although this is offset by a decline in creativity over time.

This study provides important insights on incentive system design for creative work. Prior literature suggests that creativity-weighted-quantity incentives are not successful in inducing more creative performance, but can serve as a mechanism for attracting high ability individuals (Kachelmeier & Williamson, 2010). However, my study shows that for individuals with HCC, who are typically employed for creative work, creativity-weighted-quantity incentives elicit highly creative performance consistently over time, with no decline in overall performance. Sustained creative performance in repeated tasks is important for a range of creative jobs such as architects, designers, and researchers, all of whom need to be encouraged to perform creatively in the many projects they undertake.[5]

[5] By repeated performance of a creative task over time, I mean the repeated performance of a task by an employee in a relatively short time horizon. E.g. Designing a line of clothing that will include multiple designs [fashion designer], designing logos for several companies within a year [graphics designer], or designing housing plans for several houses in a year [architect]. There are various other factors that can affect the performance of an

My findings also add to the literature on the interplay between personal and contextual characteristics relating to creative work, as called for by Shalley and Gilson (2004). Finally, to my knowledge, my study is the first to measure the timing of creative ideas under multidimensional incentives in an experiment context. The time-sensitive findings highlight the importance of studying the effects over time in future studies on creativity.

The remainder of the chapter is organized as follows. In section 2.2, I develop the hypotheses, in section 2.3, I describe the method, followed by the results and findings in section 2.4. Finally, in section 2.5, I conclude with a discussion of the key implications of the study.

2.2. Background and hypothesis development

In this section, I first discuss the extant literature on creativity-weighted-quantity incentives. Following this, I develop hypotheses on the effect of such incentives on creative performance and overall performance, conditional on individuals' creative capacity.

2.2.1. Background: Incentivizing both creativity and quantity

In organizations, employees who need to be creative also need to perform on other task dimensions (e.g. achieve volume targets, reach financial and cost goals, perform other tasks that are more routine). Examples of such employees include architects, fashion designers, researchers, and employees of innovation or research laboratories. While creativity is an important performance dimension for such jobs, Grabner (2014) points out how creativity for creativity's sake is not desirable for most organizations. Unsworth (2018) emphasizes the importance of studying creativity in the context of multiple goals.

individual over their long careers. Personal traits and characteristics themselves can change over a longer period of time (Feist, 1998). While interesting, this is beyond the scope of this study.

If both creativity and quantity are important for an organization, then behavioral theories suggest that only quantity should be extrinsically rewarded because incentives for creativity are unnecessary, or sometimes even undesirable.[6] There is a long-held belief that rewards for creativity negatively affects creative performance by crowding out intrinsic motivation for the task (Amabile, 1985; Deci, 1971, 1975).[7] However, studies in psychology and management do not consider the existence of incentives for other dimensions of performance, which may already crowd out intrinsic motivation (Unsworth, 2018). Recent accounting research that considers multidimensional incentives that includes creativity finds that creativity incentives are generally undesirable. KRW propose that individuals, when faced with creativity-weighted-quantity incentives, prioritize the creativity objective over the quantity objective, but without success in improving creative performance.[8] However, the presence of incentives for other dimensions, such as quantity-only incentives, can already crowd out intrinsic motivation for the task. Therefore, the crowding-out effect of adding creativity incentives to a multidimensional incentive contract is likely minimal or irrelevant.

In contrast, economic theory clearly predicts that if individuals are incentivized for quantity and not creativity, self-interested individuals will focus solely on the quantity dimension, and as a result, will not focus on creativity at all (Holmstrom & Milgrom, 1991). In addition, the theory of learned industriousness suggests that repeated effort can be an

[6] Individuals are "extrinsically motivated to engage in a task if they view their task motivation as motivated primarily by external goals such as the promise of reward or the expectation of evaluation" (Amabile, 1985, p. 393). Conversely, intrinsic motivation occurs when individuals engage in a task for their own interest and involvement (Amabile, 1985; Deci, 1975).

[7] When extrinsic rewards (or punishments) are attached to performance, individuals focus on achieving those rewards (or avoiding punishment), and thus, intrinsic motivation for the task declines (Deci, 1971, 1975). This is referred to as the crowding-out effect.

[8] KRW were the first to study the effect of rewarding creativity, with and without quantity incentives. Their experiment participants constructed rebus puzzles on blank cards. Participants were compensated in one of four ways: fixed pay; incentives for quantity; incentives for creativity; or incentives for creativity-weighted-quantity. KRW show that participants under quantity-only incentives, produce as much highly creative output as those under creativity-weighted-quantity incentives, while producing significantly more output overall. This result, and results from subsequent studies (e.g. Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010) suggest that creativity-weighted-quantity incentives do not induce greater creative performance.

unpleasant sensation. The positive sensation from being rewarded can counter this unpleasant sensation and the resulting aversion to exerting effort (Eisenberger & Armeli, 1997). When rewards do not reflect high or low effort in a certain dimension, the desire to work towards that dimension reduces. If there are rewards for some task dimensions (e.g. quantity, cost control), but not for creativity, individuals will learn to respond to the rewards and ignore creativity. This argument is consistent with the literature on incomplete performance measures that highlights the dangers of only incentivizing some dimensions in a multidimensional setting (e.g. Feltham & Xie, 1994; Holmstrom & Milgrom, 1991; Roberts, 2010). Therefore, it is important to include creativity as part of the reward system.

While economic theory predicts that creativity incentives are required to induce creative effort in a multidimensional task environment, empirical evidence suggests that such incentives are ineffective (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010).[9] However, these studies do not consider the individuals' potential to produce creative outcomes (i.e. their creative capacity). In the following sections, I hypothesize the effect of creativity-weighted-quantity incentives on creative and overall performance, depending on individuals' creative capacity.

2.2.2. Creative performance under creativity-weighted-quantity incentives for individuals with different creative capacities

The capacity to be creative varies across individuals.[10] Using composite measures of creative performance, many studies find variations in creative potential and creative performance (e.g.

[9] The few exceptions to this are studies that look at specific situations in which creativity incentives are beneficial such as single-solution tasks, and peer-evaluated tasks (e.g. Charness & Grieco, 2018; Liu et al., 2019).

[10] There is a long-standing question on whether creative personality characteristics are innate (i.e. nature) or developed through environmental factors (i.e. nurture). The overarching belief is that both nature and nurture are complementary factors (Vernon, 1989). In this study, I refer to creative capacity as a trait that could be a result of nature, nurture, or a combination of both. Nevertheless, longitudinal studies find creative personality characteristics to be stable over time (Feist, 1998).

Barron & Harrington, 1981; Feist, 1998). Some common contributors to variation in creative capacities across individuals include intrinsic motivation based on early development (Deci, 1971), natural cognitive style (Kirton, 1976), and skills in creative thinking (Amabile, 1988). Empirical evidence indicates that such individual-level factors are predictive of employees' creative performance (e.g. Malik & Butt, 2017; Shalley et al., 2004).

Differences in creative capacity across individuals can lead to differences in how they respond to incentives. For individuals with LCC, creativity-weighted-quantity incentives can be stressful because the incentives demand creative performance that the individual has difficulty delivering. The capacity to perform is a determinant of the effectiveness of incentives (e.g. Locke, 1968; Locke, Shaw, Saari, & Latham, 1981). Effort may be elicited by the incentive but the desired outcome will not be achieved if there is a lack of ability/capacity (Locke, 1968).

In comparison, for individuals with HCC, creativity-weighted-quantity incentives can be effective because they are able to deliver performance on both creativity and quantity dimensions. Consequently, they have the capacity to make efficient trade-offs between the two. Furthermore, where individuals have the capacity to be creative, they are more likely to consider the incentives for creativity as a recognition of their abilities. In contrast, if individuals know they have the capacity to be creative, but they are not rewarded for it, economic theory predicts that they are less likely to be encouraged to be creative. Malik et al. (2015) find survey evidence of a strong positive relation between creativity incentives and creative performance when individuals' creative self-efficacy is high, but a negative relation when it is low. However, they do not consider incentives for dimensions other than creativity.[11] Kachelmeier and Williamson (2010) (hereafter referred to as KW) find that individuals with high creative

[11] In addition, the authors acknowledge that their survey data does not allow them to control for the type of creative work and the magnitude of the incentives.

ability do not improve creative performance under creativity-weighted-quantity incentives. However, their measure of creative ability is based on the choice of the incentive contract. Those who self-select into quantity-only incentives are considered to have low creative ability, and those who self-select into creativity-weighted-quantity incentives are considered to have high creative ability. However, this self-selection to incentive contracts is task specific and KW argue that it reflects how easily some initial creative ideas about the task come to the participants' minds. In contrast, I argue that, if creative capacity is explicitly measured as a personality characteristic unrelated to a specific task or incentives, then it will better capture individuals with HCC who are able to respond to creativity incentives. Hence, I predict that:

Hypothesis 1 [H1]: Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to higher creative performance for individuals with HCC, but not for individuals with LCC.

2.2.3. Overall performance under creativity-weighted-quantity incentives for individuals with different creative capacities

Creativity is not costless and so organizations need to carefully consider the extent to which it is incentivized. Thus, it is important to evaluate whether the inclusion of creativity in incentives has any negative effects on overall performance.

Under standard economic assumptions of rationality, individuals have complete control over their effort and subsequent outcomes, and they respond to incentive schemes to produce outputs that maximize their pay-off. That is, the individuals under quantity-only incentives will produce as many units as possible with no regard for creativity at all. And individuals under creativity-weighted-quantity incentives will produce outputs so that any loss in the quantity will be exactly offset by the increase in creativity.

However, it is unclear whether the assumption of rationality is appropriate in the context of creativity due to the unique features of creative tasks. First, the relation between input effort and creative output is unclear (Kachelmeier et al., 2008). Second, creativity can be a subjective evaluation, while quantity is objective. When a good performance measure is available for one dimension but not for the other, individuals generally tend to prioritize the measurable dimension (Roberts, 2010). However, KRW observe that when one of the performance dimensions is creativity, individuals tend to simplify the incentives by prioritizing creativity over other dimensions and only submit ideas that reach a self-imposed threshold level of creativity. Third, creative tasks are inherently connected to behavioral effects such as triggering intrinsic motivation (Shalley et al., 2004), and creating stress and debilitating pressure (Ariely et al., 2009; Byron, Khazanchi, & Nazarian, 2010). Together, these unique characteristics challenge the ability to make rational trade-offs between creativity and quantity.

Empirical research finds that when individuals are subject to creativity-weighted-quantity incentives they invest effort on creativity, but are not able to convert this effort to more creative output (Kachelmeier et al., 2008). This results in the wasted effort, lower quantity produced, and poorer overall performance. Based on this argument, if individuals are somehow able to improve creative output by increasing effort (e.g. Amabile et al., 1996), then they should be able to compensate for the loss in quantity produced. Thus, I argue that individuals with HCC can use their capacity to produce creative output if they invest effort on creativity. However, under quantity-only incentives, they will not invest effort in creativity, but maintain overall performance through producing a larger quantity. Therefore, creativity-weighted-quantity incentives, compared to quantity-only incentives, are unlikely to affect overall performance for individuals with HCC.

In contrast, in the process of trying to improve a performance dimension in which they lack aptitude, individuals with LCC under creativity-weighted-quantity incentives will waste

valuable effort they could otherwise invest towards achieving more quantity. Consequently, they will perform poorly in quantity with no noticeable improvement in creativity. That is, creativity-weighted-quantity incentives will likely lead to a lower overall performance compared to quantity-only incentives for individuals with LCC. Therefore, I predict that:

Hypothesis 2 [H2]: Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to lower overall performance for individuals with LCC, but not for individuals with HCC.

2.2.4. Performance under creativity-weighted-quantity incentives over time

For creative tasks that are performed repeatedly over time (e.g. fashion designing, architectural designing), it is important to consistently encourage creativity throughout. Prior studies on creativity are unable to capture the performance in repeated tasks relative to time with reasonable precision, which is a specific focus in my study.[12,13]

For HCC individuals, I expect sustained creative and overall performance over time because they have the creative capacity to respond to creativity-weighted-quantity incentives. Creative personality characteristics are generally stable over time (Feist, 1998) enabling consistently high creative performance from individuals with HCC. This also means that such individuals can consistently make desirable trade-offs between effort invested in creativity and quantity.[14]

[12] In this study, I refer to time as a relatively short-term construct. See footnote 5 for a detailed explanation.
[13] KW attempt to approximate the time taken per creative idea by defining the number of outputs up to the mean for all participants as "initial" and the remainder as "subsequent". However, this is a crude measure because the time spent per output could vary.
[14] Contrary to this, KW find that creativity-weighted-quantity incentives negatively affect subsequent performance of high-ability individuals. In their study, creative ability is inferred from individuals' self-selection into the incentive contract, a method that is specific to the task. KW acknowledge that: "For softer performance dimensions such as creativity, we find that individuals have reasonable self-insight into their *initial creative abilities* [emphasis added] but cannot necessarily anticipate the effort-inducing effects of the selected contracts in mapping this self-insight to sustainable aggregate performance" (Kachelmeier & Williamson, 2010, p. 1689). In contrast, I argue that the relatively stable characteristic of HCC enables individuals to effectively respond to creativity-weighted-quantity incentives consistently.

For individuals with LCC, it is unlikely that incentives can elicit a sustained increase in creativity. While creativity-weighted-quantity incentives may prompt these individuals to produce some creative outputs initially, this will not be sustained over time, because trying harder will not lead to further creative performance in the absence of strong creative capacity. For individuals with LCC, I argue that they are unable to respond to creativity-weighted-quantity incentives once they exhaust a limited stock of ideas. As a result, their creative performance may appear high initially, compared to LCC individuals with quantity-only incentives, but this will fall subsequently.[15] Consequently, the overall performance of individuals with LCC under creativity-weighted-quantity incentives is expected to be lower in the later period, compared to the initial period.

Hypothesis 3a [H3a]: For individuals with LCC, creativity-weighted-quantity incentives will lead to lower creative performance in the later period compared to the initial period.

Hypothesis 3b [H3b]: For individuals with LCC, creativity-weighted-quantity incentives will lead to lower overall performance in the later period compared to the initial period.

These performance effects over time mean that the severity of the negative effects of creativity-weighted-quantity incentives for individuals with LCC will be more evident in the later period. As a result, the differences between the HCC and LCC individuals predicted in H1 and H2 will also be more evident in the later period.

Hypothesis 4a [H4a]: The difference in the creative performance of HCC and LCC individuals predicted in Hypothesis 1 will be stronger in the later period compared to the initial period.

[15] This is similar to the effect observed by KW for the group self-selected to creativity-weighted-quantity incentives; individuals produce creative outputs that initially come to their mind but are not able to effectively respond to the incentives subsequently.

Hypothesis 4b [H4b]: The difference in the overall performance of HCC and LCC individuals predicted in Hypothesis 2 will be stronger in the later period compared to the initial period.

The predicted relations are summarized in Figure 1. Variable definitions and explanations of experiment conditions are provided in Appendix A.

2.3. Method

In this section, I first discuss the details of the experiment, including the task, design, participants, and procedure. Following this, I discuss the key dependent variables.

2.3.1. Task

The creative task used in the experiment is to construct rebus puzzles, similar to KRW. The experiment instrument is provided in Appendix B. All participants are instructed that both creativity and quantity are valued by the experimenters.

2.3.2. Design

To test the hypotheses, I design a 2x2 between-subject experiment. The first factor is the *incentive type*, manipulated as quantity-only incentives or creativity-weighted-quantity incentives. Quantity-only incentives are based on the number of puzzles produced, and creativity-weighted-quantity incentives are based on a creativity-weighted total score, which is the sum of creativity ratings of all puzzles produced per participant. I do not include a base group with no incentives (i.e. fixed-pay) because prior research already shows that having no incentives leads to lower creative and overall performance compared to conditions with incentives (e.g. Kachelmeier et al., 2008).

The second factor, *individual creative capacity*, is measured prior to introducing the experiment task (i.e. to capture non-task specific creative capacity) using a survey instrument

for creative self-efficacy by Tierney and Farmer (2002). This 3-item, 7-point scale measure has been used repeatedly in the prior literature (e.g. Baer, Oldham, & Cummings, 2003; Gong, Huang, & Farh, 2009; Richter, Hirst, Van Knippenberg, & Baer, 2012; Tierney & Farmer, 2002) and is found to be strongly positively correlated to creative performance (Tierney & Farmer, 2002, 2004). A sample item in the measure is 'I have confidence in my ability to solve problems creatively'.[16] This instrument has the advantage of brevity compared to other validated instruments in the literature such as the Creative Personality Scale (Gough, 1979), Five-Factor Model of personality (Costa Jr & McCrae, 2008), and Adaptor-Innovator inventory (Kirton, 1976), all of which have a large inventory of items that is impractical to use in an experiment setting. I use educational background, self-assessment of performance, and future career preferences, as additional measures to validate that the creative self-efficacy instrument sufficiently captures creative capacity. Using this self-reported measure, I partition the sample based on the median into HCC and LCC groups.

2.3.3. Participants

The experiment was conducted with 85 university students at an experimental economics laboratory. Subjects were randomly drawn from a registered pool of participants pre-screened to be born in countries with English as the first language (i.e. the United States, Canada, the United Kingdom, Australia, New Zealand or South Africa). Based on repeated failures to answer questions on how incentives are calculated at the beginning of the experiment, eight participants were excluded from the analysis.[17] These eight participants took five to seven attempts to answer three multiple choice questions correctly, indicating that they had poor

[16] There is a copyright on this measure by Tierney and Farmer (2002) that only allows the publication of the sample item.
[17] The questions required participants to calculate the incentive they would receive under different scenarios. There were three multiple choice questions per participant.

understanding of the incentive contract. In the final sample of 77 participants, 66 answered all questions correctly in the first try, and 11 answered all three questions in four attempts.[18]

2.3.4. Procedure

The experiment procedure is illustrated in Figure 2. All data was collected using a custom-developed web-based interface that embeds a drawing tool. Upon entering the laboratory, participants were asked to answer a pre-experiment questionnaire that included the measure to capture creative capacity.[19] Thereafter, they were randomly allocated to an incentive condition and introduced to the rebus puzzle task. Participants were not aware of incentive conditions other than their own. Incentives ranged from $5 to $45 with an average payment at about $25 within each incentive condition. After reading the instructions, they answered questions on their understanding of the incentive contract. Then they had a practice round of 3 minutes on the task, followed by the experiment task of 20 minutes. Finally, they answered a post-experiment questionnaire which included manipulation checks, questions on individual characteristics and preferences, and their performance in the task. The participants were paid after two weeks to allow time to collect creativity ratings.

[18] All the inferences relating to this study's predictions are unchanged by this exclusion. However, when the 8 participants with poor understanding are included, the direct effects found in the prior literature do not replicate statistically significantly, although directionally consistent. In order to maximize the comparability of the current study with prior literature, I conduct the analysis after excluding the 8 participants with poor understanding of the incentive contract.

[19] I took several measures to guard against a possible demand effect on creativity by asking participants to self-rate creative capacity before the experiment. First, I included some distracting demographic questions in the same questionnaire. Second, I added some distracting "nonsense" statements, unrelated to the experiment, before and after the creative capacity statements. These statements were intentionally designed to have very little effect on their performance in the task. An example of a "nonsense" statement is "I live walking distance from the University (see Appendix B for the experiment instrument). Third, in the post-experiment questionnaire, I directly asked whether answering the pre-experiment questionnaire affected their performance in the rebus puzzle task. Only 13 individuals answered "yes", explaining that the self-ratings on creativity affected their performance in some way. There was no statistical difference across conditions. Even if the self-rating did encourage individuals to focus more on creativity, this would work against the results because it will encourage all individuals to focus on creativity (not only those who are incentivized for it), which will reduce differences between the groups.

A panel of independent raters, recruited from the online crowdsourcing platform Prolific, rated the puzzles for creativity.[20] They were first introduced to the rebus puzzles similar to the participants of the experiment but were not made aware of the incentive conditions in the experiment. Thereafter, 10 raters each rated a random subset of about 164 puzzles from a total of 1,895 puzzles.[21] There were 15 groups of raters based on the subsets of puzzles leading to 150 raters in total. I paid each rater GBP 5.00 for the rating work that took, on average, 60 minutes. Following KW, 40 randomly selected puzzles were rated by all raters at the beginning, in the same order, to measure inter-rater reliability. For the first 40 puzzles, correlations between the average rating within each subset and the average of the other 14 subsets are significantly positive (between 0.80 and 0.95). The Cronbach's alpha for the first 40 puzzles rated by everyone is 0.98. This provides evidence that the creativity ratings are, on average, reasonably consistent across subsets. To check for reliability and consistency within subsets, I calculated the correlations between each rater and the average of the other nine raters in the same subset. These correlations range from -0.14 to 0.71, exhibiting some inconsistent raters compared to the rest of the group. To ensure reliability, I use a threshold of 0.2 minimum correlation, which is the lowest correlation used by KW.[22] Based on this threshold, I drop 14 raters which leaves a minimum (maximum) of 7 (10) raters per group.

2.3.5. Key dependent variables

To capture creative performance, I use the number of highly creative puzzles per participant, which is regarded as a meaningful measure in the prior literature (Kachelmeier et al., 2008).

[20] I pre-screened the raters to be first-language English speakers, based in the United Kingdom or the United States of America, above 18 years of age, a student (for comparability with prior studies), with at least a 90% approval rate of responses on previous submissions on Prolific (for response quality), and with a minimum of 10 prior submissions on Prolific (to allow for a reasonable basis to assess response quality).

[21] Subsets of puzzles were used to reduce rater fatigue so that rating quality can be maintained.

[22] Sensitivity checks for other thresholds, namely, excluding negatively correlated raters (2 raters dropped), excluding raters below 0.1 minimum correlation (5 raters dropped) were conducted. Inferences relating to the predicted findings are unchanged for these alternative thresholds.

This measure of creativity is suitable for firms that place more value on highly creative ideas and has little to no interest in (the lack of) mediocre ideas. I first calculate the creativity rating for each puzzle by averaging the ratings across the panel of raters. Then, I designate the 25% puzzles with the highest creativity ratings as "highly creative puzzles". The remaining 75% are "mediocre" puzzles.[23]

For completeness, I also include average creativity per participant, another commonly used, but less robust, measure. I calculate average creativity per participant as the average of the creativity ratings for all the puzzles submitted by each participant. However, this measure would be distorted if a participant were to submit fewer mediocre puzzles rather than more highly creative puzzles.[24] Because firms are typically interested in highly creative ideas rather than average performance, this measure is also less likely to be relevant to practice.

To capture overall performance, I weight the number of puzzles submitted by a participant by the creativity rating for each of those puzzles. This provides a measure of the ability to maximize performance on both creativity and quantity, as required in the instructions.

2.4. Results and discussion

In this section, I first verify the validity of the measure for creative capacity, followed by an analysis of the measures for creative performance. Then, I discuss the main results of the study, followed by some additional analyses based on the post-experiment questionnaire.

[23] The 25% threshold is used following prior studies (Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). Repeating the analysis using other thresholds such as the top 10% or the top 30% does not change the inferences relating to the predictions of this study. Generally, the higher the threshold for "highly creative puzzles", the stronger are the results. To facilitate comparability with prior literature, I use the 25% threshold in the analyses presented.

[24] An individual with the same number of highly creative puzzles as another individual, can appear to have better creative performance if she submits less mediocre puzzles. This "denominator effect" of Average Creativity was highlighted as a weakness by both KRW and KW. Hence, the number of highly creative puzzles is considered a more robust measure.

2.4.1. Verification of creative capacity

The factor analysis for the Tierney and Farmer (2002) 3-item measure used to capture creative capacity produces a single factor, *CC_Score*, with individual item loadings between 0.74 and 0.83. Cronbach's alpha is 0.7, which meets typical reliability thresholds (Peterson, 1994). These measures provide reasonable confidence on the internal validity of the measure.

After calculating the *CC_Score*, I split the sample based on its median (i.e. *High_CC* = 1 if above median, and 0 otherwise). The use of the median split allows an approximately equal distribution of individuals across the conditions. It is also statistically appropriate because the full range of the average of the three items are 3.3 to 7 (on a scale of 1-7), with a mean of 5.21 and a median of 5.33. Skewness of this average score is -0.22, and the skewness of the *CC_Score* is -0.21, both well within the typically accepted range (Kline, 2015).

To establish construct validity of the *CC_Score*, I use three additional measures from the post-experiment questionnaire: educational background, self-assessment of performance, and future career preferences. Although educational background is a noisy measure, I develop the indicator variable *CreativeArea* which equals 1 if it falls into a "creative area", and 0 otherwise.[25] The correlation between *CreativeArea* and *CC_Score* is positive and significant at 0.23. To the extent that individuals of high creative ability self-select into subject areas typically considered to be creative, this provides an indication that *CC_Score* is a reasonably valid measure of creative ability. I also ask participants to rate whether they think their rebus puzzles are likely to be more creative than an average participant's creativity (Kachelmeier & Williamson, 2010). This measure is positively and significantly correlated with *CC_Score* with a correlation value of 0.26. Finally, I asked participants to rate on a scale of 0 to 10, the degree

[25] I hand-code music, theatre, architecture, psychology, philosophy, humanities and applied subjects such as engineering and computer science as "creative areas". Science, pure mathematics, business, law and education are coded as "non-creative areas". This is of course subjective. However, several variations to the classification such as excluding engineering and computer science from *CreativeArea* do not affect the positive correlation.

to which they have a preference for a future career that incorporates novel problem solving, technical knowledge, and clearly defined procedures. Novel problem-solving has a 0.50 significant positive correlation with *CC_Score* while reliance on technical knowledge and reliance on clearly defined procedures are not significantly correlated. Together, these tests provide reasonable evidence as to the validity of the creative capacity measure.

I rule out a number of alternative explanations to the *CC_Score* that may conceivably drive results. First, I capture *Education*, which is a measure of the level of education of participants, to rule out the possibility of *CC_Score* simply reflecting ability. *Education* is coded as 1 if the participant is a post-graduate student at the university, and 0 otherwise.[26] The correlation between this measure and the *CC_Score* is 0.004 and insignificant. Second, I examine the relation between *CC_Score* and participants' self-assessment of their prior experience in rebus puzzles. The correlation between the two measures is 0.12 and insignificant, indicating that the measure for creative ability is broader, and findings are not simply driven by participants' prior experience with rebus puzzles.

2.4.2. Variation in creativity

I establish that creativity ratings exhibit a meaningful variation in both the number of highly creative puzzles and average creativity. Raters used the full range of ratings from 1 to 10, and the highest (lowest) average rating for a puzzle across the panel of raters is 9 (1.7). Figure 3a shows the distribution of highly creative puzzles, while Figure 3b shows the distribution of average creativity ratings per participant. These distributions show variation in both measures, supporting their use as meaningful dependent variables. Figure 4 provides examples of puzzles from different points in the distribution of creativity ratings.

[26] A post-graduate student is defined as any participant who has already completed at least a bachelor's degree or more. Of the 85 participants, 35 were post-graduate students.

2.4.3. Descriptive statistics

Descriptive statistics for the four experiment conditions are presented in Table 1. Collapsing the HCC and LCC groups, I find that, under creativity-weighted-quantity incentives, compared to quantity-only incentives, the number of puzzles produced is significantly lower (p-value = 0.09, one-tailed), average creativity is significantly higher (p-value = 0.08, one-tailed), and the number of highly creative puzzles is not significantly different (p-value = 0.26). These results replicate the findings in KRW. However, unlike KRW, I do not find a statistically significant difference for overall performance, although it is directionally lower under creativity-weighted-quantity incentives compared to quantity-only incentives (p-value = 0.18, one tailed).[27]

2.4.4. Hypotheses testing: Creative and overall performance

In this section I first discuss results relating to creative performance, followed by overall performance.[28] The results are illustrated in Figures 5a-5f.

2.4.4.1. Effects on *creative performance*

Table 2 Panel A presents ANOVA results for the number of highly creative puzzles, my primary measure of creative performance. In H1, I predict that creativity-weighted-quantity

[27] There are two possible reasons for this difference. First, while I use almost the same instructions as KRW, the participants verified their understanding of incentives in multiple-choice questions with no involvement of the experimenter. In KRW, the participants wrote down the incentive structure in their own words and the experimenter read through each and asked the participants to reattempt until it was clear the participant understood the incentive structure. I chose not to do this because the experimenter's involvement could wrongly imply that incentives reflect the experiment's performance requirements (which is not true in the case of quantity-only incentives). By making this choice, I may have diluted the emphasis on quantity in quantity-only incentives. Second, the self-rating of creative self-efficacy prior to the experiment could have created some demand effect for more creativity. Both these reasons could lead to comparatively lower focus on quantity for those in the quantity-only incentives condition, shrinking the differences between conditions for overall performance.

[28] To highlight the effects for each outcome variable, I present the results based on creative and overall performance separately. The discussion relating to both full duration and timing effects are within these performance dimensions. Therefore, the order of hypotheses discussed in the results section is different to the order followed in theory building. For clarity, a summary of results is presented in Table 6.

incentives will improve creative performance for the HCC group, but not for the LCC group. While I find no significant interaction at conventional levels of significance, at *p*-value = 0.11 (one-tailed) it is just shy of significance. Exploring the simple effects of creativity-weighted-quantity incentives for the HCC and LCC groups, I find that creativity-weighted-quantity incentives have a positive effect on the number of highly creative puzzles by the HCC group (*p*-value = 0.05, one-tailed), but not the LCC group (*p*-value > 0.40).

To examine the effects over time (H4a), I separately measure performance in the first half (FH) and the second half (SH) corresponding to the first/last 10 minutes of the experiment.[29] Table 2 Panels B and C repeat the main analysis for FH and SH separately, to examine results for H4a. Consistent with H4a, the effect predicted in H1 is stronger in the SH. In the FH (Panel B), all participants under creativity-weighted-quantity incentives produce more highly creative puzzles, irrespective of whether they are HCC or LCC (*p*-value = 0.08, one-tailed). In the SH (Panel C), there is a significant positive effect of creativity-weighted-quantity incentives on the number of highly creative puzzles produced by the HCC group (*p*-value = 0.07, one-tailed), and no difference for the LCC group (*p*-value = 0.16, one-tailed). Consistent with H4a, this provides evidence that creativity-weighted-quantity incentives lead to stronger differences in creative performance between the HCC and LCC groups in the later period.

Furthermore, Table 3 Panel A presents paired-sample t-test results on the number of highly creative puzzles between FH and SH within all conditions. As predicted in H3a, there is a significant reduction in the number of highly creative puzzles by the LCC group under creativity-weighted-quantity incentives in SH compared to FH (*p*-value = 0.05, one-tailed).

[29] While 10 minutes can seem short for a creative task in an organizational setting, 'time' is a relative concept in an experiment setting. Ten minutes is relatively long in my setting because constructing a rebus puzzle took about 63 seconds, on average. Other studies have made use of short experiment time durations to examine performance over time (e.g. Gneezy & List, 2006; Kachelmeier & Williamson, 2010).

There is no such difference in creative performance across time for the HCC group under creativity-weighted-quantity incentives (p-value = 0.92).

Overall, there is evidence that the HCC group has stronger creative performance under creativity-weighted-quantity incentives, compared to quantity-only incentives, in both the initial and later periods. Meanwhile, the LCC individuals under creativity-weighted-quantity incentives, compared to quantity-only incentives, produce some initial creative ideas, but this declines over time.

For completeness, I also conduct the analysis using average creativity as the measure of creative performance. Table 4 Panel A presents ANOVA results for average creativity. I do not find support for H1 with average creativity. Creativity-weighted-quantity incentives lead to better average creativity in all individuals (p-value = 0.08, one-tailed), with no difference between LCC and HCC groups. In the FH (Table 4 Panel B), average creativity of participants does not vary across different incentives (p-value = 0.11, one-tailed), but in the SH (Table 4 Panel C), the HCC participants with creativity-weighted-quantity incentives have greater average creativity than the rest of the group (p-value = 0.07, one-tailed). This supports H4a, because individuals with LCC are not able to consistently respond to creativity-weighted-quantity incentives, whereas individuals with HCC are able to do so over time. I find no significant difference in average creativity for the LCC group with creativity-weighted-quantity incentives over time (p-value = 0.47, one-tailed) (see Table 3 Panel B), which does not support H3a. Overall, the results on average creativity, at least directionally, provide insights similar to those using highly creative puzzles.

Finally, I compare these results with KW, who look at the effectiveness of self-selection into quantity-only and creativity-weighted-quantity incentives, when compared to random assignment, on creative performance. While KW argue that individuals with high creative

ability will self-select into creativity-weighted-quantity incentives, their definition of ability is based on the task at hand, and the immediate ideas that come to mind when the task is introduced.[30] In contrast, I use the *CC_Score* before the task is even introduced to participants, in order to capture a more stable personality characteristic.[31] Based on this conceptual difference, I do not expect to see the same results as KW in my experiment. Instead, my theory argues that higher ability in creativity will lead to better response to creativity-weighted-quantity incentives. Contrary to KW, my analysis shows that all individuals with creativity-weighted-quantity incentives demonstrate better creativity when compared to all individuals with quantity-only incentives (regardless of self-selected or randomly assigned) for the first eight puzzles (*p*-value = 0.09, one-tailed, for average creativity; and *p*-value = 0.04, one-tailed, for the number of highly creative puzzles).[32] The creativity rating for the very first puzzle is not significantly different across any of the cells. This is as expected because, in my study, there is no *ex ante* reason to expect individuals with stable HCC, compared to a mix of high and low ability individuals, to perform better only in the first few puzzles produced.

2.4.4.2. Effects on *overall performance*

Table 5 Panel A presents ANOVA results for overall performance of participants. There is no significant difference across incentive types or creative capacity. This is inconsistent with H2,

[30] KW explain that the "individuals contemplating the benefits of a creativity-weighted contract would likely reflect on how easily some creative ideas come to mind" (p. 1672) initially, a variant of the "availability heuristic" used by Vaughn (1999).

[31] To show that *CC_Score* reflects a stable creative trait, unrelated to immediate examples that come to a person's mind, I recapture the same measure for creative capacity after the experiment task (*CC_Score_Post*). If *CC_Score* is stable, it should not be affected by participants' performance in the task, which is fresh in their minds. As expected, the correlation between *CC_Score* and *CC_Score_Post* is positive and significant at 0.72. The correlation between *CC_Score_Post* and the number of highly creative puzzles is insignificant and below 0.2.

[32] I (imperfectly) replicate the conditions in KW using a few steps. To mimic the two groups that are randomly assigned to the two incentive contracts in KW, I pool individuals based on incentive contract types, regardless of creative capacity. To (imperfectly) mimic individuals that would have self-selected into creativity-weighted-quantity incentives, I use the HCC individuals under creativity-weighted-quantity incentives. Similarly, to (imperfectly) mimic individuals that would have self-selected into a quantity-only contract, I use the LCC individuals with quantity-only incentives. Based on these four conditions, I run an ANOVA to compare results with KW.

in which I expect lower overall performance for individuals with creativity-weighted-quantity incentives, compared to those incentivized for quantity only, in the LCC group.

For further analysis, on the effects of timing (H4b), I conduct ANOVA for FH and SH in Table 5 Panels B and C, respectively. In the FH, there is no significant difference between the conditions on overall performance (*p*-value = 0.45, one-tailed). In the SH, the HCC group has significantly better overall performance than the LCC group (*p*-value = 0.10, one-tailed) regardless of the incentive type. Although this is consistent with H4b, in that the differences in overall performance for the LCC and HCC groups will be more prominent in the later period, the results show that this is true for both incentive types, with no significant effect from the incentives.

However, the paired-sample t-test results in Table 3 Panel C indicates a significant drop in overall performance in the LCC group under creativity-weighted-quantity incentives in the SH, compared to the FH (*p*-value = 0.02, one-tailed) as predicted in H3b. There is no difference between the FH and SH overall performance of the HCC group regardless of incentives. This provides some evidence of inefficient trade-offs between creativity and quantity for LCC individuals under creativity-weighted-quantity incentives in the later period. In addition, there is an unpredicted drop in overall performance in the LCC group under quantity-only incentives (*p*-value = 0.03).

The drop in overall performance for the LCC group can be partially explained by the drop in creative performance in the SH (Section 2.4.4.1). In addition, paired-sample t-tests show a drop in the number of puzzles by the LCC group in the SH compared to the FH (untabulated), both under creativity-weighted-quantity incentives (*p*-value = 0.02, one-tailed) and quantity-only incentives (*p*-value = 0.02). ANOVA results for quantity (untabulated) show that the quantity produced is significantly lower for creativity-weighted-quantity incentives in

the SH, regardless of creative capacity (p-value < 0.10, one-tailed). There is no such difference in the FH (p-value > 0.10, one-tailed). Together, a drop in both creativity and quantity performance of the LCC group in the SH explains the drop in overall performance over time. In contrast, the HCC group appears to maintain overall performance over time in a more stable manner.

In summary, individuals with HCC under creativity-weighted-quantity incentives, compared to quantity-only incentives, have higher creative performance in both initial and later periods, without a negative effect on overall performance. Individuals with LCC under creativity-weighted-quantity incentives, when compared to quantity-only incentives, have more creative puzzles in the FH, but not in the SH. Their creative performance and overall performance decline over time as well. A summary of these results is presented in Table 6. Collectively, these results imply that, to encourage more creative performance while maximizing overall performance, creativity-weighted-quantity incentives are significantly more important for the HCC group than for the LCC group.

2.4.5. Additional analysis

2.4.5.1. Manipulation checks

All participants in the quantity-only incentive condition correctly indicated that their pay was affected by the number of puzzles they produced, but about 45% also incorrectly indicated that it was affected by a creativity rating. If these participants indeed behaved as if they were paid for creativity, the differences between the incentive manipulations will be weak, working against the predictions on creative performance. Similarly, all participants under creativity-weighted-quantity incentives correctly indicated that their pay was affected by a creativity

rating, but 25% incorrectly indicated that it was not affected by the number of puzzles.[33] These participants may have produced less puzzles, leading to lower overall performance from those under creativity-weighted-quantity incentives. Excluding the participants that failed these tests leads to stronger results relating to the hypotheses but loses a substantial number of participants.[34] In order to not overly infer results from a small number of participants, I keep all participants who failed the manipulation test in the analysis.

I also included a question to verify whether the participants understood that the experimenters valued both creativity and quantity, irrespective of how they were incentivized. Excluding participants that answered "no" or "uncertain" does not lead to changes in inferences.[35] I keep all participants in the reported analyses.

2.4.5.2. Demographic characteristics

Of the participants, 43% were male, 77% ranged between ages 18 and 24, 57% were undergraduate students, 31% had at least one completed undergraduate qualification and 12% had at least one post-graduate qualification. Gender, age and level of education distributions are not significantly different across the conditions. Due to the language-sensitive nature of the experiment task, all participants were screened to be first-language English speakers, with country of birth being the UK, USA, Canada, Australia, New Zealand or South Africa, using the experiment lab participant information system. To verify this screening, I ask participants

[33] It is difficult to compare this issue with prior studies that I closely follow (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010) because those studies do not report such manipulation checks in the papers.
[34] Only 51 observations remained for the analysis because 26 participants failed the manipulation test. The lowest number per condition is eight for the HCC group with quantity-only incentives. All analyses were repeated using these participants and results were either similar or stronger.
[35] Seven participants failed this understanding question leaving 70 observations for the analysis. The lowest number per condition is 16 for the HCC group with quantity-only incentives. All analyses were repeated using these participants and results were either similar or stronger. Results relating to H1 are significant when those who failed the understanding check are excluded, but only approaching significance when they are included. However, to be conservative, to be comparable to the procedures followed in prior literature, and to minimize the loss of participants, I keep all participants that failed this check in the analysis.

to self-report the information in the experiment instrument. Nine participants self-reported that they were either born outside of the specified countries or English was not their first language. Results are stronger and inferences unchanged when these participants are excluded. However, to preserve the number of participants, they are included in the analyses presented.

2.4.5.3. Individual characteristics and preferences

In the post-experiment questionnaire, I collected measures on participants' drawing skills, prior experience with rebus puzzles, ambiguity and risk aversion tendencies, and preferences for workplace structure to verify the robustness of the results.

A self-assessed measure of drawing skills on a scale of 0-10 was collected to ensure that the creativity ratings were not a mere reflection of good drawing. The correlation between drawing skills and creative performance is 0.22 and significant (p-value < 0.10). Predictably, *CC_Score* has a 0.23 significant correlation with drawing skills (p-value < 0.05), indicating that those who perceive themselves to be highly creative, also perceive themselves to have better drawing skills. To verify that results relating to creative capacity are not simply reflective of drawing skills, I use the measure as a covariate and find the inferences to be unchanged.[36]

Prior familiarity with rebus puzzles was also collected on a scale of 0-10. While it is possible that individuals with more awareness of their creative potential could also be more aware of creative tasks, there is no significant difference in the prior familiarity of rebus puzzles between the HCC and LCC groups (p-value > 0.50). However, surprisingly, there is a marginally significant difference between the incentive types within the HCC condition (p-

[36] The measure on drawing skills is significantly positively related to the number of highly creative puzzles (p-value < 0.05, one-tailed), average creativity (p-value < 0.05, one-tailed), and overall performance (p-value < 0.10, one-tailed). The significance of the interaction between incentive type and individual creative capacity is stronger for creative performance (p-value = 0.10, one tailed, for the number of highly creative puzzles; p-value = 0.03, one-tailed, for the number of highly creative puzzles in the SH; and p-value = 0.06, one-tailed, for average creativity in the SH).

value < 0.10). I redo the analysis assuming this was a randomization error and find the inferences of the results largely unchanged.[37]

Risk and ambiguity aversion of the participants were measured adapting instruments from Charness and Grieco (2018) to check if these interacted with creative performance. Individuals' risk-seeking behavior is not different across all conditions (*p*-values > 0.20). Within the HCC condition, those under creativity-weighted-quantity incentives are significantly more ambiguity averse than those under quantity-only incentives (*p*-value < 0.05). There is no significant difference in ambiguity aversion between other groups. The lower ambiguity tolerance should lead to lower performance from HCC individuals under creativity-weighted-quantity incentives, which works against the hypotheses. Using the ambiguity aversion measure as a covariate does not affect the inferences in this study.[38]

A rating for participants' preferences for structure in the workplace was collected to assess if this might confound differences between HCC and LCC. The measure is not different across conditions (*p*-values > 0.30).

2.4.5.4. Reflections about task performance

To capture intrinsic motivation for the task, the participants rated how much they enjoyed designing the rebus puzzles. Prior research in psychology discusses how incentives for creativity can lead to crowding out of intrinsic motivation, resulting in poor creative performance. As expected, participants with HCC appear to have significantly more intrinsic

[37] Prior experience in rebus puzzles is marginally significantly positively related to average creativity for the full duration (*p*-value < 0.10, one-tailed), and significantly related to average creativity in the FH (*p*-value < 0.05, one-tailed). It is not significantly related to all other dependent measures. The significance of the interaction between incentive type and individual creative capacity is weaker for creative performance in the SH (*p*-value = 0.07, one-tailed, for the number of highly creative puzzles; and *p*-value = 0.10, one-tailed, for average creativity).
[38] Ambiguity aversion is not significantly related any of the dependent variables. The significance of the interactions between incentive type and individual creative capacity in the ANCOVA are unchanged after including the ambiguity aversion measure as a covariate.

motivation than those with LCC (p-value < 0.10, one-tailed), but the type of incentives were not found to significantly affect intrinsic motivation for any of the groups (p-value > 0.50). Thus, it appears that if individuals receive some form of incentives, even if not for creativity, then intrinsic motivation is already crowded out. If so, intrinsic motivation is not significantly affected by incrementally adding creativity incentives.

In addition, participants rated how hard they worked to increase the quantity of puzzles produced and the creativity of the puzzles produced, separately. As expected, all participants indicate they worked hard to increase the quantity of puzzles produced, irrespective of the incentive condition (p-values > 0.20), because both incentive options included a measure of quantity. However, ratings of those participants under quantity-only incentives are directionally higher, and close to significance (p-value < 0.11, one-tailed). Next, participants with creativity-weighted-quantity incentives indicated that they worked hard to increase the creativity of puzzles, but this was not significantly different to the participants with quantity-only incentives (p-value < 0.17, one-tailed). Upon closer examination, the difference is significant for the HCC group (p-value < 0.10, one-tailed) but not for the LCC group (p-value > 0.50). This is consistent with the idea that, for the LCC group, incentives for creativity are not very meaningful because the individuals have little capacity to be creative.

2.5. Conclusion

In this study I find that individuals with HCC can be motivated using creativity-weighted-quantity incentives to perform more creatively, while successfully managing the desired trade-offs between creativity and quantity. Furthermore, the effectiveness of incentives is not limited to the initial period but is, instead, stable across the entire period.

In contrast, for individuals with LCC, creativity-weighted-quantity incentives are only effective in inducing some initial creativity, but this is not sustained, with creativity suffering

in the later period. Therefore, the only possible advantage of creativity-weighted-quantity incentives for individuals with LCC is that they may produce the same amount of creative outputs over a shorter time interval, because the incentives prompt them to focus on creativity. As such, creativity-weighted-quantity incentives may be useful for eliciting a 'burst' of creative ideas, but not for sustained creativity. In addition, creativity-weighted-quantity incentives could even be harmful to overall performance of individuals with lower creative capacity. This reflects the harmful effects of incentivizing performance dimensions for which individuals do not have the capacity.

This study contributes to the literature in several ways. Firstly, it provides evidence that the response to creativity-weighted-quantity incentives can vary based on individuals' creative capacity. For individuals with HCC, creativity-weighted-quantity incentives lead to more stable and higher creative performance over time. This finding is in contrast to the prior literature, which shows creativity-weighted-quantity incentives lead to poor creative performance over time (Kachelmeier et al., 2008), even despite individuals self-selecting into such contracts (Kachelmeier & Williamson, 2010).

Secondly, this is the first study, to my knowledge, to specifically capture the time taken for each creative idea, providing more precision on data for "initial" versus "later" creativity. Performance over time is key for employees who are required to produce a sustained body of creative works. As such, this study incorporates a timing variable to better mimic reality and generate better insights on creative performance on repeated creative work over time.

Thirdly, the findings add to the creativity literature in management and psychology on the interactions between individual and contextual factors in influencing creative performance. Shalley and Gilson (2004) identifies the need for research to consider such interactions between creative personality characteristics and rewards. The findings of this research will be a starting

point to understand interactions between complex reward structures and creative personality characteristics.

My study also informs practice. If a job requires high creative performance from individuals, it is likely that individuals with higher creative capacities will be hired for the task. In this case, I show that it is desirable to reward individuals for creativity, alongside efficiency. In comparison, if organizations are seeking creativity from employees that are hired for more routine tasks, and the creative capacity of those individuals is low, then creativity-weighted-quantity incentives will not be effective in eliciting sustained creativity. However, even for individuals with low creative capacity, creativity-weighted-quantity may be useful to encourage a small 'burst' of creativity, if that is what the organization desires.

Finally, perceived creative self-efficacy can be used as one successful way to sort individuals based on creative capacities, when compared with self-selection to incentive types that was previously found to be ineffective (Kachelmeier & Williamson, 2010). This type of artificial screening of individual creative abilities can be observed in the business world, such as in the case of assessment centers to screen employees based on personality characteristics (Diekmann & König, 2015; Hunter et al., 2012).

However, reasonable caution should be exercised when generalizing the inferences made in this study to practice. First, I collected data in a controlled laboratory setting with university student participants, which is different to more complex business environments. However, because I study creativity as a more fundamental personal characteristic (with a generic task that does not require any particular type of knowledge, experience, or business context), I do not expect the theoretical predictions to be different with experienced individuals in a business environment. It is, of course, possible that other contextual characteristics could lead to more differences within the findings of this research, which is an avenue for future studies.

Second, the strict selection requirements for the experiment task created barriers to compiling a large sample of participants and, thus, results of experiments are subject to limited statistical power. In my study, the pool of eligible potential participants is narrowed by the screening criteria, and the lack of immediate payment. Even if large participant numbers are available, the relatively high incentives offered (that are designed to ensure comparability to prior literature), and the cost of obtaining ratings for creativity also limit the scale of the experiment. Despite these issues, the experiment has participant numbers comparable to prior literature (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010).

In addition, the study deliberately only focuses on a single period of performance, which ensures that creativity evaluations and incentives will not act as feedback. This is significantly different to practice, where incentives are likely to affect subsequent performance in the task. While this aspect is not explicitly examined in this study, it is likely that when individuals with HCC receive incentives for their high creative performance, it will further encourage them to be creative. In contrast, based on learned industriousness theory, individuals with HCC who do not receive incentives for their high creative performance will focus more on quantity, subsequently (Eisenberger & Armeli, 1997). This will likely lead to stronger long-term effects for individuals with HCC. When individuals with LCC receive low incentives for their creative performance, it would likely encourage them to focus less on creativity and more on quantity. While this will not lead to greater creativity for the LCC, it could potentially lead to better trade-offs between creativity and quantity. Finally, individuals with LCC under quantity-only incentives will likely continue to focus more on quantity over time, with some of the output being creative based on their basic levels of ability. Collectively, all these arguments are unlikely to dispute the overall inference in this study that, for individuals with HCC, creativity-weighted-quantity incentives can be effectively used to enhance creative performance, while not harming overall performance.

This study also paves the way for future studies on the interplay between incentives for creativity as a part of a multidimensional incentive system. For instance, the previously discussed multi-period study could be conducted to explore whether there are feedback-related effects from incentives for individuals with varying creative capacities. Kachelmeier et al. (2019) use repeat task performances to study the effects of an idea incubation period, but they do not identify the effects of individual creative capacity, nor the effects of incentives performing the role of a feedback loop. Future studies could also look at how the interplay between creative capacity and other contextual characteristics such as spatial configurations, goal setting, and budgeting can affect creative performance.

Appendix

Appendix A: Variable definitions and experiment conditions

Dependent variable definitions	
Number of puzzles	The total number of puzzles submitted by a participant
Number of highly creative puzzles	The total number of puzzles considered to be "highly creative" by each participant. "Highly creative" puzzles are those with creativity ratings in the top 25% in the pool of all puzzles submitted. Creativity rating per puzzle is defined below.
Average creativity	The average of creativity ratings for all the puzzles submitted by each participant. Creativity rating per puzzle is defined below.
Overall performance	The number of puzzles submitted by each participant weighted for the creativity rating for each of those puzzles. Creativity rating per puzzle is defined below.
Creativity rating per puzzle	The average of the individual ratings for each puzzle by a panel of independent raters.
Independent variable definitions	
Creativity_Pay	An indicator variable that takes the value of 1 if the participant received creativity-weighted-quantity incentives, and 0 if participant received quantity-only incentives. See below for descriptions of experiment conditions.
High_CC	An indicator variable that takes the value of 1 if the participant belongs to the high creative capacity (HCC) group and 0 if participant belongs to the low creative capacity (LCC) group. See below for descriptions of experiment conditions.
Timing	
Full duration	The entire 20-minute experiment task.
First half (FH)	The first 10 minutes of the experiment task.
Second half (SH)	The second 10 minutes of the experiment task.

Experiment conditions	
Incentive type	
Quantity-only incentives	Incentives range from $5 to $45, based on the number of puzzles produced per participant.
Creativity-weighted-quantity incentives	Incentives range from $5 to $45, based on a creativity-weighted total score, which is the sum of creativity ratings of all puzzles produced per participant. Creativity rating per puzzle is defined above.
Individual creative capacity	
High creative capacity (HCC)	Participants with self-reported creative capacity above the median for the sample are designated as the HCC group. Creative capacity is measured using a 3-item, 7-point Likert scale by Tierney and Farmer (2002), leading to one factor score.
Low creative capacity (LCC)	Participants with self-reported creative capacity below the median for the sample are designated as the LCC group. Creative capacity is measured using a 3-item, 7-point Likert scale by Tierney and Farmer (2002), leading to one factor score.
Other	
Timing	
Initial period	Initial stages of working on a task
Later period	Later stages of working on a task

Appendix B: Experiment instrument

<Adapted from Kachelmeier et al. (2008)>

Ground Rules

Before describing the experiment, it is important to establish ground rules.

- **NO** COMMUNICATION **WITHIN OR OUTSIDE SESSIONS**

Please help us maintain control over the experiment by refraining from commenting or communicating with your fellow participants or with others who might be participating in this research. You will be working individually during this experiment. If you have any questions, please tap on the table and raise your hand. One of the experimenters will assist you.

- **NO** DECEPTION

We promise to carry out the experiment in the manner described in these instructions, with no deception of any form. **We will make the payment for this experiment at a later date (in about two weeks).** But we promise that your pay will be determined exactly as described in the rules explained later.

Part 1: About yourself

Please answer the following questions **about yourself**. Your answers will not affect the remainder of the experiment or the pay you receive.

Please fill-in the following information.

- Gender:
- Age:
- Country currently residing in:
- Country of birth:
- Is English your first language? Yes/No
- What is your major area of study?

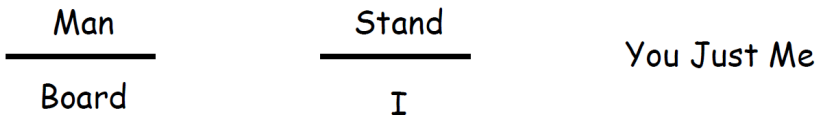
Using the following responses please indicate the extent to which you agree or disagree that each statement currently describes you.

	1 - Very strongly disagree	2 - Strongly disagree	3 - Disagree	4 - Neutral	5 - Agree	6 - Strongly agree	7 - Very strongly agree
a. I do my best work early in the workday.	○	○	○	○	○	○	○
b. [Statement 1 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
c. [Statement 2 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
d. [Statement 3 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
e. I prefer to live within walking-distance of my job.	○	○	○	○	○	○	○
f. I prefer working in an environment with fewer distractions.	○	○	○	○	○	○	○

What is the highest level of education you have completed?

Part 2: Task Instructions

In this research, you will be constructing "rebus puzzles". A rebus puzzle is a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase. Here are some examples:



These examples use the positioning of the words to create the riddle.

- The first is "man overboard", because "man" is over "board".
- The second is "I understand", for similar reasons.
- The third is "just between you and me" (i.e., "just" is between "you" and "me")

Task Instructions (continued)

Other rebuses use counting, different sizes, shapes, or positions to create the riddle, such as these examples:

Rebus Puzzle	Word-Phrase/ Solution
0 M.D. Ph.D. B.A.	Three degrees below zero
Funny Funny Words Words Words Words	Too funny for words (2 funny; 4 words)
pains	Growing pains
R R O A D S A D S	Cross Roads
CHAIR	High chair

Task Instructions (continued)

Still other rebuses use simple pictures, symbols, and diagrams as part of the riddle, like these two examples:

Rebus Puzzle	Word-Phrase/ Solution
	A hole in one
BB 	To be or not to be

---

What we would like you to do:

In this experiment, you will be presented with a blank drawing pad. Using the stylus provided to you, construct a single rebus puzzle in the drawing pad. Remember to type the word-phrase in the box below the drawing pad. Then Click "Submit". Once you submit a puzzle, it is considered finished and you will be redirected to a fresh drawing pad. Then proceed to design the next puzzle and repeat the submission process.

While we do not place any rules on the kind of rebus puzzles you can submit, we value both the ***number of different puzzles*** you can construct (i.e. **quantity**) and the ***creativity*** of those puzzles (i.e. puzzles that are **original ideas, innovative, and clever**).

You will have 20 minutes to construct your rebus puzzles.

Payment:

[Quantity-only incentives]

Your payment will be based on ***how many*** rebus puzzles you construct in 20 minutes. To determine this, we will count the number of rebus puzzles you construct.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) who constructs the ***most*** rebus puzzles in this group will receive **$ 45.00**. The participant (or participants, if tied) who constructs the ***fewest*** rebus puzzles in this group will receive **$ 5.00**. Everyone else in the group will receive an amount between $ 45.00 and $ 5.00, depending on the number of rebus puzzles constructed. This will result in an expected average payment of approximately **$ 25.00**. The more puzzles you construct, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

[Creativity-weighted-quantity incentives]

Your payment will be based on ***the creativity-weighted total score*** of all rebus puzzles you construct in 20 minutes.

To determine this score, we will ask a group of independent raters to rate the creativity of each puzzle you submit on a 1 to 10 scale, where "10" is the highest rating and "1" is the lowest rating. We will then ***add*** the creativity ratings of ***all*** puzzles you submit to arrive at the creativity-weighted total score. Thus, ***each puzzle you submit helps your total score, but higher rated puzzles count more***. For example, at the extreme, a puzzle rated 10 counts ten times as much as a puzzle rated 1.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) with the ***highest*** creativity-weighted total score in this group will receive **$ 45.00**. The participant (or participants, if tied) with the ***lowest*** creativity-weighted total score in this group will receive **$ 5.00**. Everyone else in the group will receive an amount between $ 45.00 and $ 5.00, depending on the individual creativity-weighted total score. This will result in an expected average payment of approximately **$ 25.00**. The higher your creativity-weighted total score, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

Checking your understanding:

Let's see if you understood the payment scheme:

[Quantity-only incentives]

Suppose your randomly assigned group has 4 participants. If you produce 8 puzzles, and the three others produce 6, 7, and 8 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If you produce 20 puzzles, and the three others produce 25, 30 and 30 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If you produce 15 puzzles, and the three others produce 6, 18 and 18 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

[Creativity-weighted-quantity incentives]

Suppose you produce 5 puzzles and a panel of independent raters awards them creativity scores as 4, 7, 2, 6, and 9 respectively. What is your creativity-weighted total score?

- 28
- 5.6
- 9

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 45, and the three others receive 35, 40, and 29 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 60, and the three others receive 88, 60, and 79 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 33, and the three others receive 35, 60, and 29 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

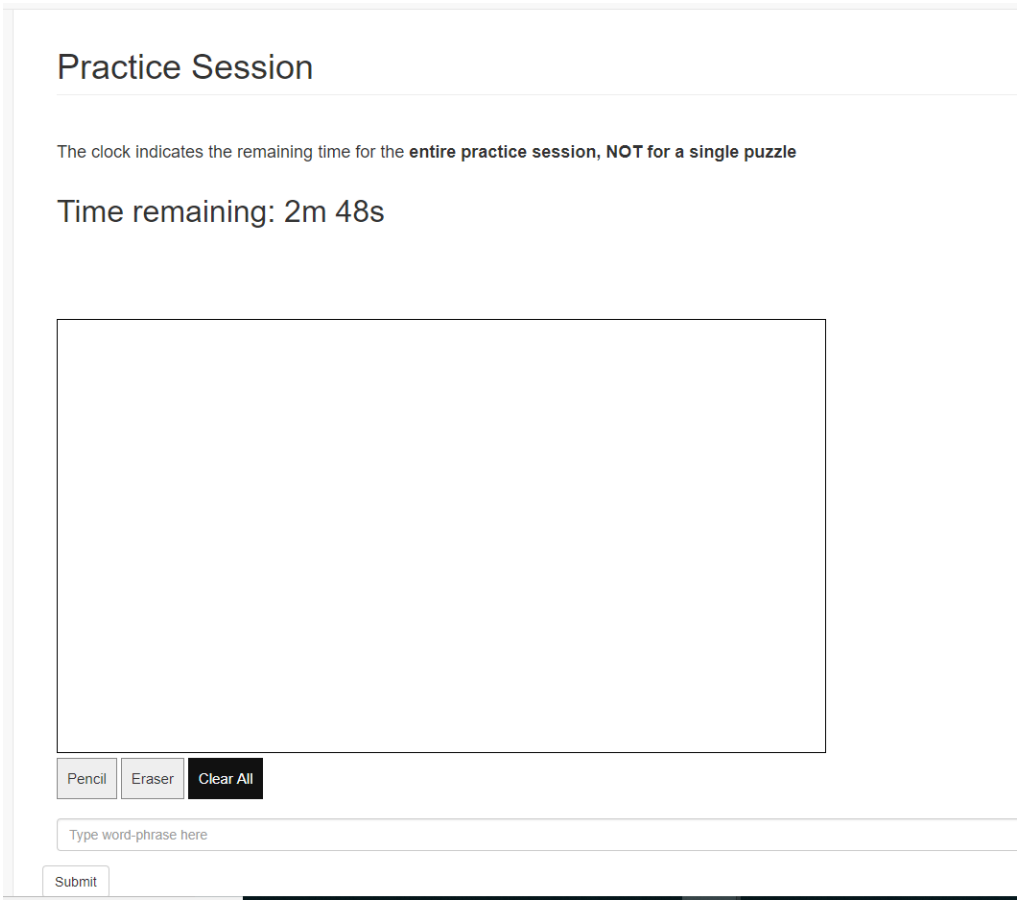
Practice Session

Now let us have a practice session. Performance in this section will **NOT count towards your pay**. It is simply an opportunity to familiarize yourself with the drawing interface and with the puzzle submission process.

In the next screen you will see the drawing pad. After constructing one puzzle on the drawing pad, type the word-phrase in the space provided and click "Submit". Then you will see a new drawing pad on which you can construct a new puzzle. Accordingly, you can construct and submit as many puzzles as you like.

You have **three minutes to practice**. There will be a clock on the screen that will count-down the 3 minutes for your reference.

Click "Start" to proceed.

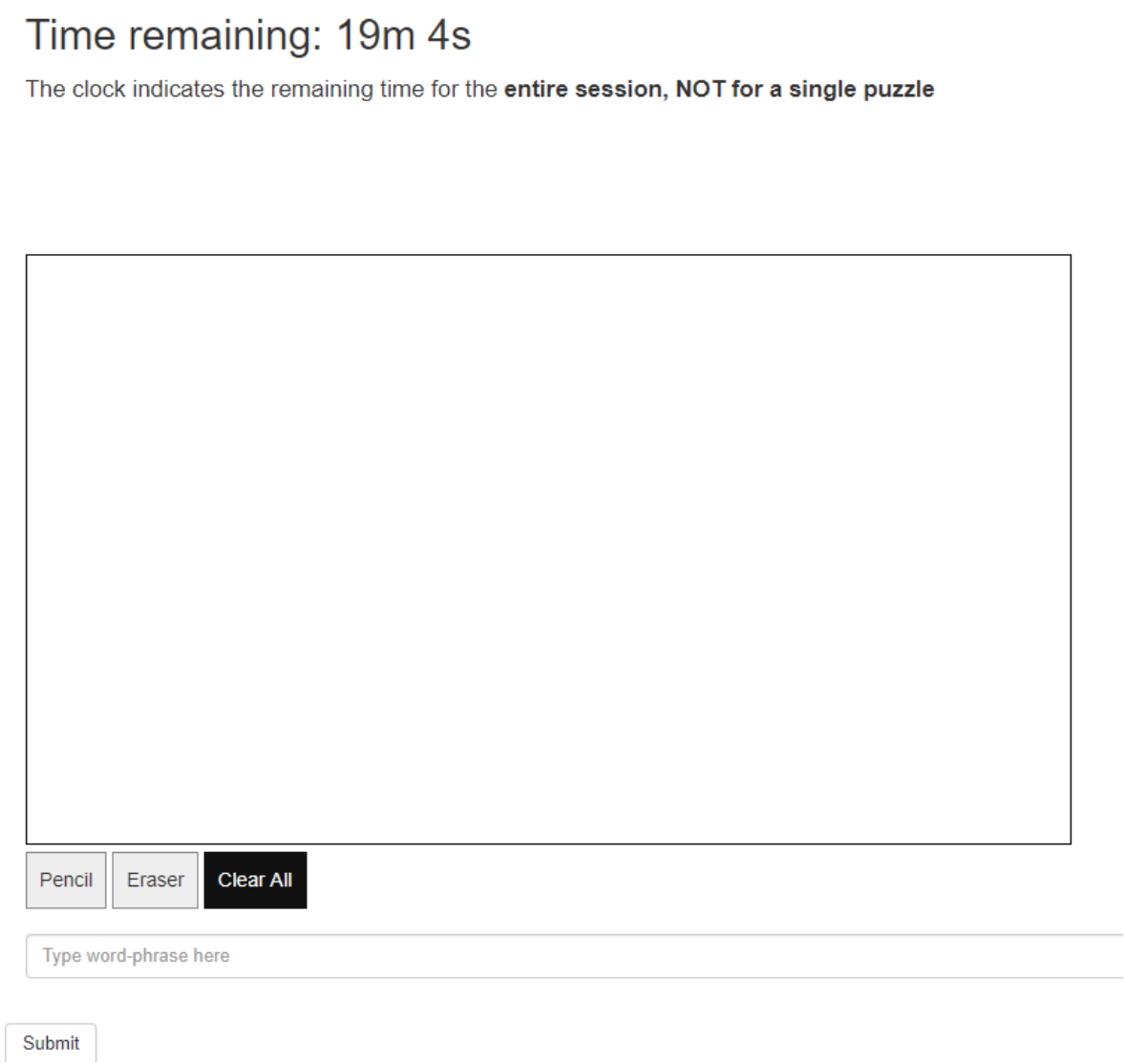


Main Session

Now let us start the experiment.

Performance in this section **WILL DETERMINE YOUR PAY** as described earlier. You have **20 minutes**. There will be a clock on the screen that will count-down the 20 minutes for your reference.

Please click "Start" to proceed.



Part 3: Some Questions

Please answer the next few questions **about the task you performed** in the experiment.

Your payment will not be affected by the answers provided. However, it is important that you answer these questions to complete the experiment.

Click "Next" to continue.

This question is **about your pay**. Do you agree with the following statements?

	Yes	No	Not Certain
The payment you will receive in this experiment is affected by a **creativity rating for each puzzle** you produced.	○	○	○
The payment you will receive in this experiment is affected by **how many puzzles** you produced.	○	○	○

This question is about the **instructions you read** for the experiment. Do you agree with the following statement?

	Yes	No	Not Certain
The instructions indicated that we value **both quantity and creativity** in the puzzles produced	○	○	○

This question is about **your performance on the rebus puzzle constructing task** in the experiment. Do you agree with the following statement?

	Yes	No
Answering the questions about myself at the beginning of the experiment affected my performance in the rebus puzzle constructing task	○	○

If yes, please briefly explain how:

To what extent do you agree or disagree with the following statements?

	Strongly Disagree	Disagree	Somewhat disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
	1	2	3	4	5	6	7
I enjoyed designing rebus puzzles in the experiment	○	○	○	○	○	○	○
My rebus puzzles were likely more creative than the average participant's in this study	○	○	○	○	○	○	○
I worked hard to increase the number of rebus puzzles I constructed	○	○	○	○	○	○	○
I worked hard to increase the creativity of the rebus puzzles I constructed	○	○	○	○	○	○	○

Please rate on the scale below the degree to which you were familiar with rebus puzzles before participating in this study.

Never Heard										Highly Familiar
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

What specific prior familiarity did you have with rebus puzzles? (select the most appropriate answer)

- I had no prior familiarity with rebus puzzles
- I had solved rebus puzzles in the past, but had not designed any
- I had designed rebus puzzles in the past, but had not solved any
- I was aware of rebus puzzles, but had never designed or solved any
- I had designed and solved rebus puzzles in the past

Please answer the next few questions **about yourself,** more generally.

Suppose you are given $100. You can invest any portion of this money in an investment opportunity. There is an **unknown chance** that this investment will be successful. If the investment is successful, you will be paid 2.5 times the amount invested. If the investment is not successful, you will lose the money you invested. You will retain any money you do not invest. How much of the $100 will you invest in this opportunity?

(Please provide an amount between 0 and 100 without decimal points.)

$ [ ]

Please rate the degree to which you consider yourself skilled in drawing.

Not Skilled										**Highly Skilled**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Using the following responses please indicate the extent to which you agree or disagree that each statement currently describes you.

	1 - Very strongly disagree	**2 - Strongly disagree**	**3 - Disagree**	**4 - Neutral**	**5 - Agree**	**6 - Strongly agree**	**7 - Very strongly agree**
[Statement 1 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
[Statement 2 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
[Statement 3 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○

The following question relates to your preferences for your **future career**.

Please indicate on the scale below the degree to which you would prefer the following characteristics in your ideal job.

Characteristic	**Low**									**High**
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Reliance on technical knowledge	○	○	○	○	○	○	○	○	○	○	○
Reliance on novel problem-solving	○	○	○	○	○	○	○	○	○	○	○
Reliance on clearly-defined procedures	○	○	○	○	○	○	○	○	○	○	○

People differ in the amount of structure they prefer in their working environment. What degree of structure do you prefer in your working environment?

Highly Unstructured										**Highly Structured**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

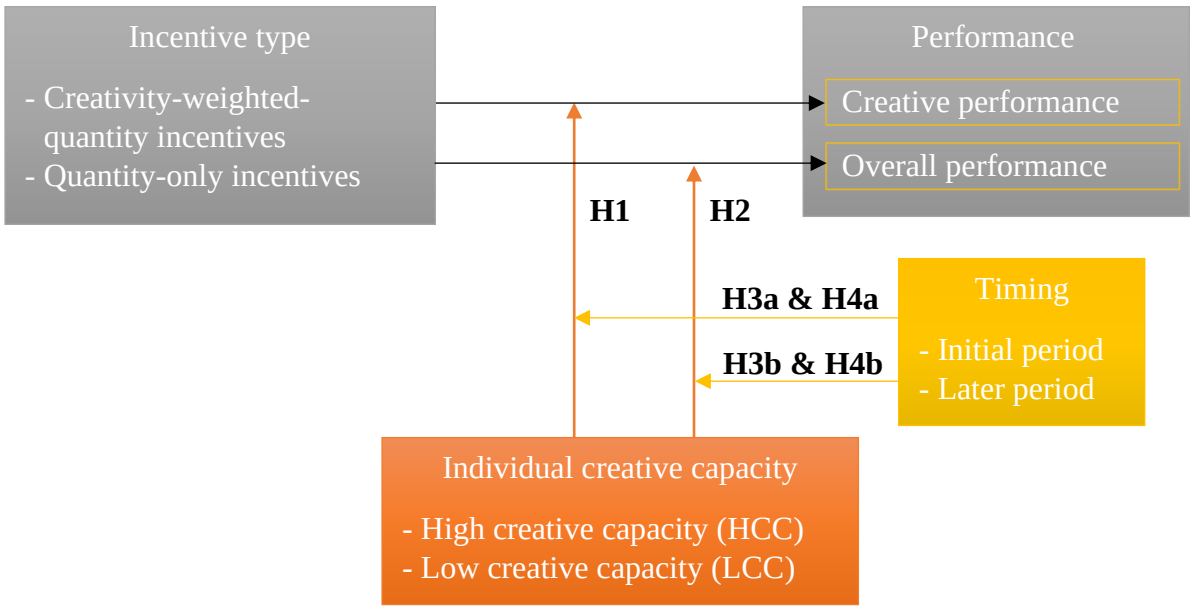
Please answer this question **about yourself.**

Suppose you are given $100. You can invest any portion of this money in an investment opportunity. There is **a 50% chance** that this investment will be successful. If the investment is successful, you will be paid 2.5 times the amount invested. If the investment is not successful, you will lose the money invested. You will retain any money you do not invest. How much of the $100 will you invest in this opportunity?

(Please provide an amount between 0 and 100 without decimal points.)

$ [ ]

Figures and tables



H1	Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to higher creative performance for individuals with HCC, but not for those with LCC.
H2	Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to lower overall performance for individuals with LCC, but not for those with HCC.
H3a	For individuals with LCC, creativity-weighted-quantity incentives will lead to lower creative performance in the later period compared to the initial period.
H3b	For individuals with LCC, creativity-weighted-quantity incentives will lead to lower overall performance in the later period compared to the initial period.
H4a	The difference in the creative performance of HCC and LCC individuals predicted in Hypothesis 1 will be stronger in the later period compared to the initial period.
H4b	The difference in the overall performance of HCC and LCC individuals predicted in Hypothesis 2 will be stronger in the later period compared to the initial period.

Figure 1: Predicted relations. This figure illustrates the predicted effects of individual creative capacity and timing on the relationship between incentive types and performance. See Appendix A for descriptions of experiment conditions.

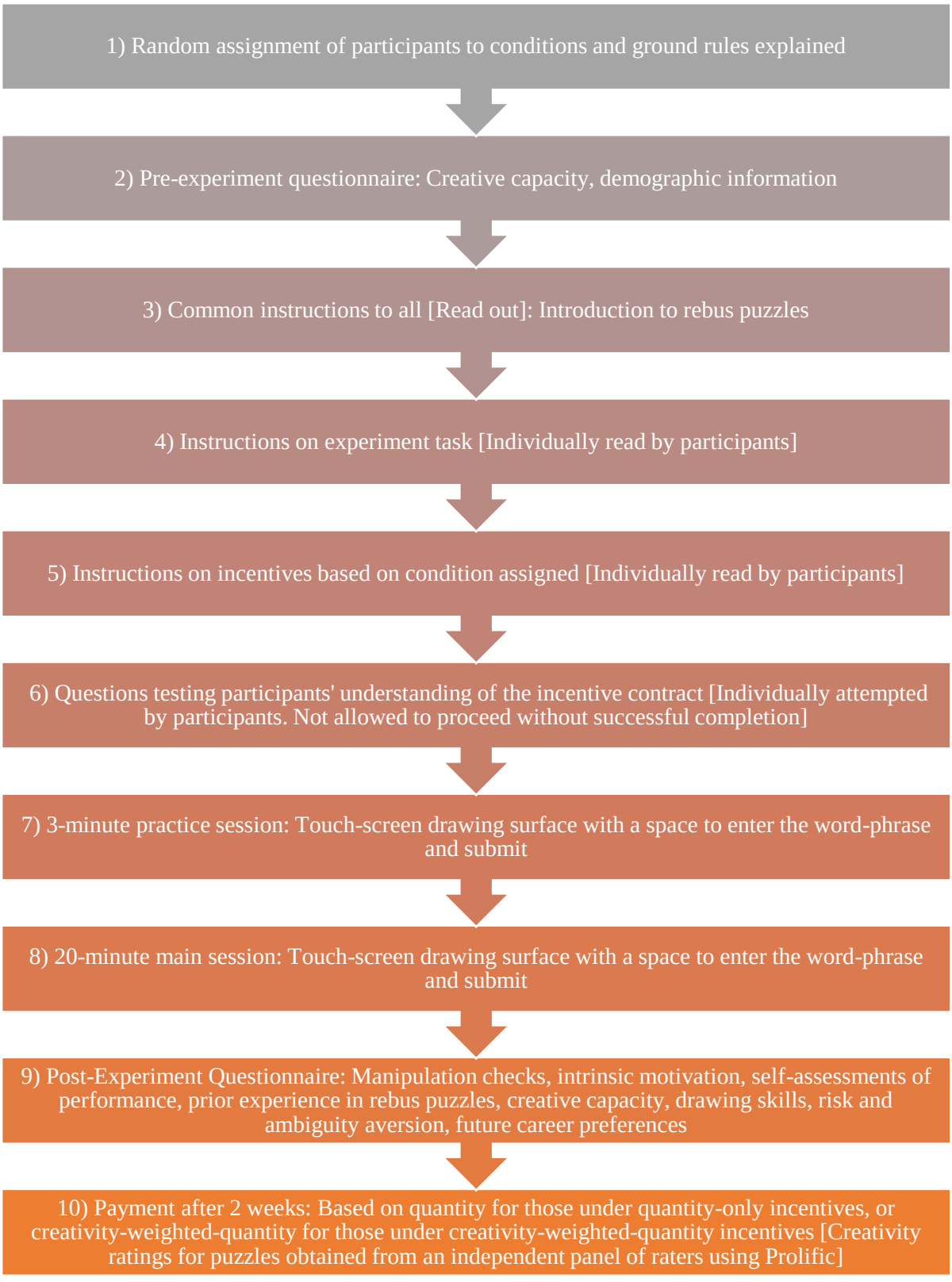
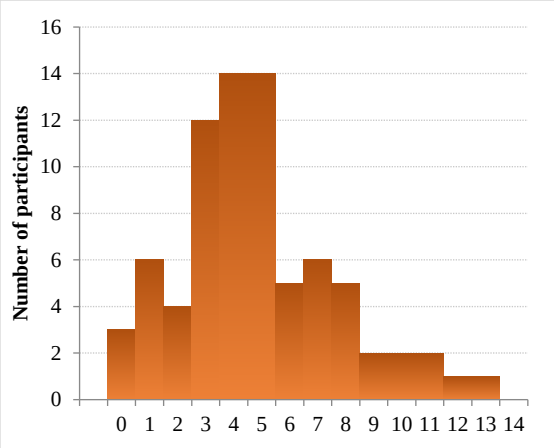


Figure 2: Experiment procedure. This figure illustrates the stepwise procedure followed to collect data.

(a) Number of highly creative puzzles



(b) Average creativity

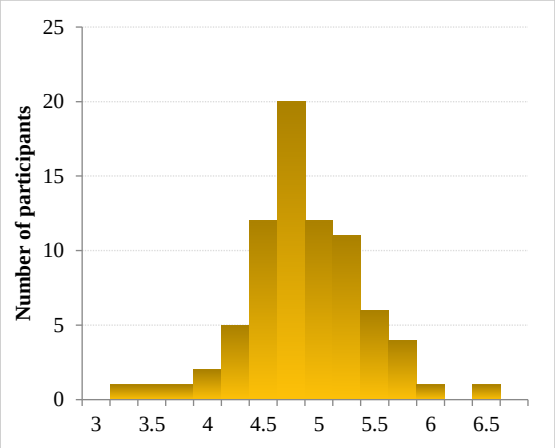


Figure 3a-b: Distribution of creative performance. 3a. Distribution of the number of highly creative puzzles. 3b. Distribution of average creativity per participant. See Appendix A for variable definitions.

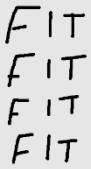
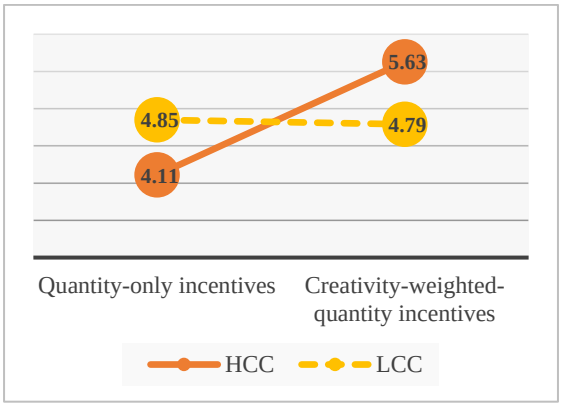
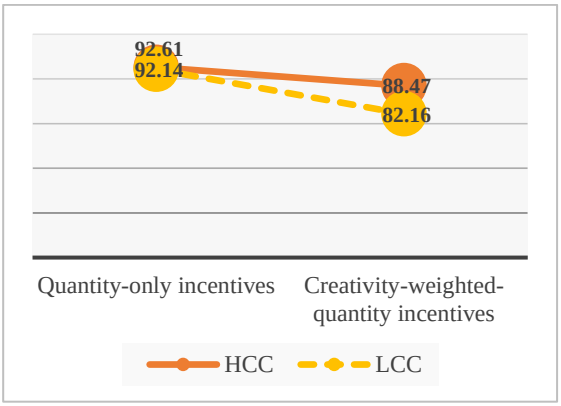
Word/Phrase: Hanging out  Creativity rating: 8.4 (in top 5%) ***** Condition: HCC/Quantity-only incentives	**Word/Phrase: Cough drops**  Creativity rating: 7.6 (in top 5%) ***** Condition: HCC/Creativity-weighted-quantity incentives
Word/Phrase: Twisted thoughts  Creativity rating: 5.5 (3rd quartile) ***** Condition: LCC/Quantity-only incentives	**Word/Phrase: Forfeit**  Creativity rating: 4.7 (2nd quartile) Condition: HCC/Creativity-weighted-quantity incentives
Word/Phrase: Think outside the box  Creativity rating: 3.9 (1st quartile) Condition: HCC/Quantity-only incentives	**Word/Phrase: Over the top**  Creativity rating: 2.4 (in bottom 5%) Condition: LCC/Creativity-weighted-quantity incentives

Figure 4: Examples of rebus puzzles from different points of the distribution of creativity ratings. This figure illustrates the stepwise procedure followed to collect data. See Appendix A for variable definitions and descriptions of experiment conditions. Puzzles designated as highly creative are marked with a ***.**

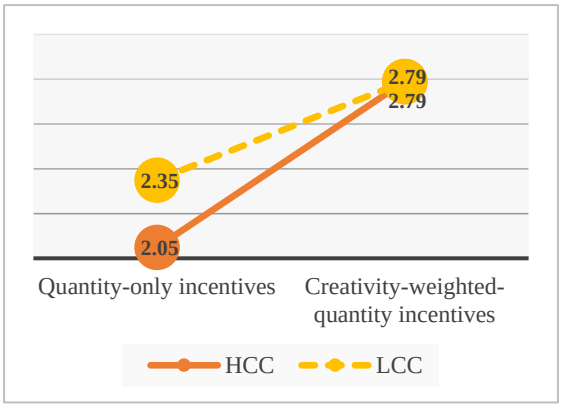
(a) Full duration: Creative performance (Number of highly creative puzzles)



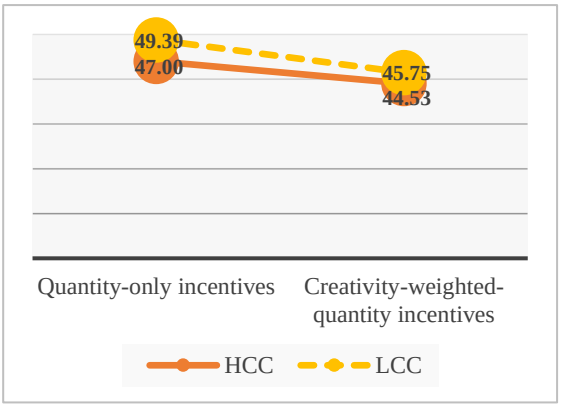
(b) Full duration: Overall performance



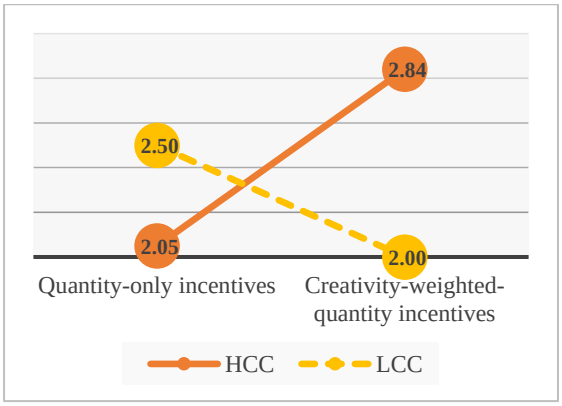
(c) First half: Creative performance (Number of highly creative puzzles)



(d) First half: Overall performance



(e) Second half: Creative performance (Number of highly creative puzzles)



(f) Second half: Overall performance

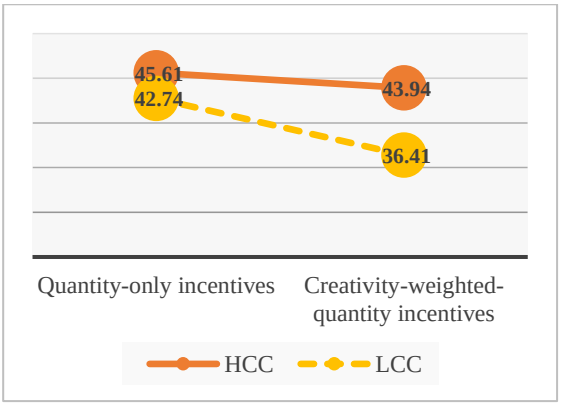


Figure 5a-f: Illustration of key results. This figure illustrates the effect of individual creative capacity and incentive type on performance over time. 6a-b. Results on creative and overall performance for the full duration. 6c-d. Results on creative and overall performance for the first half. 6e-f. Results on creative and overall performance for the second half. See Appendix A for variable definitions and descriptions on experiment conditions.

Table 1: Descriptive statistics

		HCC + Quantity-only incentives	HCC + Creativity-weighted-quantity incentives	LCC + Quantity-only incentives	LCC + Creativity-weighted-quantity incentives	Differences HCC versus LCC *p*-value	Differences Quantity-only versus Creativity-weighted-quantity *p*-value
		A	B	C	D	(A+B) vs (C+D)	(A+C) vs (B+D)
Number of participants	N	19	19	20	19		
Number of puzzles per participant	Mean	20.42	18.37	19.60	17.16	0.55	**0.09**
	Min	10.00	10.00	5.00	6.00	.	.
	Max	32.00	39.00	37.00	35.00	.	.
	SDev	5.92	5.89	8.91	7.85	.	.
Average creativity per participant	Mean	4.56	4.85	4.75	4.81	0.54	**0.08**
	Min	3.60	3.12	4.11	4.03	.	.
	Max	5.27	5.78	6.32	5.64	.	.
	SDev	0.44	0.73	0.47	0.48	.	.
The number of highly creative puzzles per participant	Mean	4.11	5.63	4.85	4.79	0.94	0.26
	Min	-	-	1.00	1.00	.	.
	Max	9.00	11.00	13.00	12.00	.	.
	SDev	2.40	3.02	2.64	3.21	.	.
Overall performance per participant	Mean	92.61	88.47	92.14	82.16	0.67	0.18
	Min	50.71	37.46	21.57	33.83	.	.
	Max	155.51	162.63	170.13	171.14	.	.
	SDev	26.91	26.31	39.87	38.44	.	.

Table 1: Descriptive statistics (continued)

First half (FH):	Mean	4.54	4.78	4.68	4.81	0.58	0.11
Average creativity	Min	3.56	2.34	3.78	3.87	.	.
per participant	Max	5.33	5.83	6.62	5.79	.	.
	SDev	0.53	0.94	0.60	0.47	.	.
FH: The number of	Mean	2.05	2.79	2.35	2.79	0.73	0.16
highly creative	Min	0.00	0.00	0.00	0.00	.	.
puzzles per	Max	5.00	6.00	7.00	8.00	.	.
participant	SDev	1.51	1.90	1.69	2.10	.	.
FH: Overall	Mean	47.01	44.53	49.39	45.75	0.67	0.24
performance per	Min	21.36	11.71	12.37	14.88	.	.
participant	Max	81.28	82.00	92.24	94.57	.	.
	SDev	15.27	17.53	20.56	22.00	.	.
Second Half (SH):	Mean	4.58	4.88	4.88	4.78	0.46	0.24
Average creativity	Min	3.64	3.59	4.23	402	.	.
per participant	Max	5.87	6.08	5.96	6.32	.	.
	SDev	0.55	0.67	0.48	0.64	.	.
SH: The number of	Mean	2.05	2.84	2.50	2.00	0.60	0.71
highly creative	Min	0.00	0.00	0.00	0.00	.	.
puzzles per	Max	5.00	7.00	6.00	5.00	.	.
participant	SDev	1.35	1.83	1.47	1.67	.	.
SH: Overall	Mean	45.61	43.94	42.74	36.41	0.20	0.16
performance per	Min	29.13	25.75	9.20	9.22	.	.
participant	Max	74.24	80.63	80.24	90.74	.	.
	SDev	13.76	13.21	21.28	20.41	.	.

This table presents descriptive statistics for the four experiment cells, and differences across the manipulation/measured variable. See Appendix A for variable definitions and descriptions on experiment conditions. The *p*-values for difference between HCC and LCC are two-tailed because main effects are not predicted. The *p*-values for differences between incentive types are one-tailed for the number of puzzles, average creativity and overall performance per participant, because prior literature provides clear directional predictions for these (Kachelmeier et al., 2008). The *p*-value for the difference between incentive types for the number of highly creative puzzles is two-tailed because prior literature does not provide basis for a clear directional prediction (Kachelmeier et al., 2008).

Table 2: ANOVA for the effect of individual creative capacity and incentives on the number of highly creative puzzles

Panel A: Full duration

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	10.33	1.29	0.13
High_CC	1	0.05	0.01	0.47
Creativity_Pay High_CC*	1	12.11	1.51	0.11
Error	73			
Simple effects				
Effect of *Creativity_Pay* given *High_CC*=1	1	22.13	2.97	**0.05**
Effect of *Creativity_Pay* given *High_CC*=0	1	0.04	<0.01	0.47

Panel B: First half (FH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	6.66	2.03	**0.08**
High_CC	1	0.43	0.13	0.36
Creativity_Pay High_CC*	1	0.43	0.13	0.36
Error	73			

Panel C: Second half (SH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	0.40	0.16	0.35
High_CC	1	0.75	0.30	0.29
Creativity_Pay High_CC*	1	8.00	3.16	**0.04**
Error	73			
Simple effects				
Effect of *Creativity_Pay* given *High_CC*=1	1	5.92	2.28	**0.07**
Effect of *Creativity_Pay* given *High_CC*=0	1	2.44	0.99	0.16

This table presents ANOVA results and simple effects on the number of highly creative puzzles. Panel A looks at performance in the full duration while Panels B and C focus on the first half and second half, respectively. See Appendix A for variable definitions and descriptions on experiment conditions. The *p*-values are one-tailed.

Table 3: Differences in performance over time: First half (FH) and second half (SH)

Panel A: Number of highly creative puzzles

Condition	n	FH	SH	Difference	
				t-stat	*p*-value
HCC + Quantity-only incentives	19	2.05	2.05	<0.01	1.00
HCC + Creativity-weighted-quantity incentives	19	2.79	2.84	- 0.10	0.92
LCC + Quantity-only incentives	20	2.35	2.50	- 0.38	0.71
LCC + Creativity-weighted-quantity incentives	19	2.79	2.00	1.71	**0.05**

Panel B: Average creativity

Condition	n	FH	SH	Difference	
				t-stat	*p*-value
HCC + Quantity-only incentives	19	4.54	4.58	- 0.26	0.80
HCC + Creativity-weighted-quantity incentives	19	4.78	4.88	- 0.73	0.47
LCC + Quantity-only incentives	20	4.68	4.88	- 1.45	0.16
LCC + Creativity-weighted-quantity incentives	19	4.81	4.78	0.27	0.40

Panel C: Overall performance

Condition	n	FH	SH	Difference	
				t-stat	*p*-value
HCC + Quantity-only incentives	19	47.01	45.61	0.55	0.59
HCC + Creativity-weighted-quantity incentives	19	44.53	43.94	0.16	0.88
LCC + Quantity-only incentives	20	49.39	42.74	2.34	**0.03**
LCC + Creativity-weighted-quantity incentives	19	45.75	36.41	2.27	**0.02**

Panels A, B and C of this table present results from paired sample t-tests on differences over time for creative and overall performance. See Appendix A for variable definitions and descriptions on experiment conditions. The *p*-values for the LCC group under creativity-weighted-quantity incentives are reported one-tailed because directional predictions are made in H3a and H3b. All other *p*-values are two-tailed.

Table 4: ANOVA for the effect of individual creative capacity and incentives on average creativity

Panel A: Full duration

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	0.59	1.99	**0.08**
High_CC	1	0.11	0.38	0.27
Creativity_Pay High_CC*	1	0.27	0.91	0.17
Error	73			

Panel B: First half (FH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	0.67	1.54	0.11
High_CC	1	0.14	0.32	0.29
Creativity_Pay High_CC*	1	0.05	0.12	0.36
Error	73			

Panel C: Second half (SH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	0.19	0.54	0.23
High_CC	1	0.19	0.55	0.23
Creativity_Pay High_CC*	1	0.81	2.34	**0.07**
Error	73			
Simple effects				
Effect of *Creativity_Pay* given *High_CC*=1	1	0.88	2.33	**0.07**
Effect of *Creativity_Pay* given *High_CC*=0	1	0.11	0.35	0.28

This table presents ANOVA results and simple effects on average creativity. Panel A looks at performance in the full duration while Panels B and C focus on the first half and second half, respectively. See Appendix A for variable definitions and descriptions on experiment conditions. The *p*-values are one-tailed.

Table 5: ANOVA for the effect of individual creative capacity and incentives on overall performance

Panel A: Full duration

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	959.61	0.85	0.18
High_CC	1	221.63	0.20	0.33
Creativity_Pay High_CC*	1	163.59	0.15	0.35
Error	73			

Panel B: First half (FH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	179.81	0.50	0.24
High_CC	1	62.73	0.17	0.34
Creativity_Pay High_CC*	1	6.64	0.02	0.45
Error	73			

Panel C: Second half (SH)

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	308.64	0.99	0.16
High_CC	1	520.17	1.68	**0.10**
Creativity_Pay High_CC*	1	104.33	0.34	0.28
Error	73			

This table presents ANOVA results on overall performance. Panel A looks at performance in the full duration while Panels B and C focus on the first half and second half, respectively. See Appendix A for variable definitions and descriptions on experiment conditions. The *p*-values are one-tailed.

Table 6: Summary of results

Hypotheses on the interactive effects of individual creative capacity: ANOVA results			
Hypothesis	Dependent construct	Predicted effect of creativity-weighted-quantity incentives	Result
H1	Full duration (FD): Creative performance	HCC: Positive LCC: No difference	HCC: Positive LCC: No difference (Interaction just shy of significance)
H2	FD: Overall performance	HCC: No difference LCC: Negative	HCC: No difference LCC: No difference
H4a	Second half (SH): Creative performance	HCC: Positive (Strong) LCC: No difference	HCC: Positive (Strong) LCC: No difference
H4b	SH: Overall performance	HCC: No difference LCC: Negative (Strong)	HCC: No difference LCC: No difference
Hypotheses comparing performance in FH to SH: Paired-sample t-test results			
Hypothesis	Dependent construct	Predicted difference between FH and SH for creativity-weighted-quantity incentives	Result
H3a	Creative performance	LCC: Negative	LCC: Negative
H3b	Overall performance	LCC: Negative	LCC: Negative

This table presents a summary of results from ANOVA and paired-sample t-tests against hypothesized predictions. See Appendix A for variable definitions and descriptions on experiment conditions.

Chapter 3
Multidimensional incentives for different types of creative tasks

Abstract

Using two experiments, I examine the effect of creativity and quantity incentives on the performance of creative tasks with pre-specified problems, and the influence of incentive strength on this relationship. I manipulate creative task type (pre-specified problems or not), incentive type (quantity-only or creativity-weighted-quantity), and incentive strength (weak or strong). Contrary to theoretical expectations, I find that creativity-weighted-quantity incentives negatively affect creative and overall performance for tasks with pre-specified problems, although, predictably, this effect is lower when incentives are weaker than when they are stronger. This shows that creativity-weighted-quantity incentives are not just ineffective, as found in the prior literature, but can impede creative performance in tasks with pre-specified problems. Additionally, weaker incentives may alleviate some of the negative effects of incentives on performance. These results highlight the need to consider creative task types and incentive strengths when designing incentives, and to conduct further research on creative tasks with pre-specified problems.

Key words: Closed tasks, creativity, quantity, incentives, incentive strength

"… *be creative, take smart risks and move with speed*"
– Working at Walmart (Walmart Inc., 2018)

3.1. Introduction

Creativity has become essential to modern organizations as they seek to remain competitive in rapidly changing business environments (Amabile & Khaire, 2008; Unsworth, 2018). Creativity is defined as the production of novel, potentially useful ideas (Amabile et al., 1996; Malik & Butt, 2017; Shalley et al., 2004), and is at the heart of innovation.[1]

Recent studies in accounting are beginning to shed light on how incentives can influence individuals' motivation to be creative in order to improve firm performance. An important focus of this work is achieving a balance between creativity and efficiency (e.g. Chang & Birkett, 2004; Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010).[2] Early studies find that financial incentives for "efficient creativity", in the form of creativity-weighted-quantity incentives, do not improve creativity or the joint performance on creativity and efficiency (hereafter referred to as "overall performance") (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010). In Chapter 2, I show that this relationship differs based on the creative capacity of individuals. However, most prior work, including my study in Chapter 2, consider creativity as a broad construct, and there are repeated calls to decompose the construct into distinct types of creativity (e.g. Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015). This study contributes to this effort by examining the

[1] Innovation is defined as the development and successful implementation of novel and useful ideas (Amabile et al., 1996; Malik & Butt, 2017; Shalley et al., 2004). Hence creativity is a necessary first step for innovation. While I view creativity and innovation as distinctive constructs, I will refer to relevant literature from both areas throughout the chapter because creativity is an integral part of innovation.

[2] Efficiency is defined as getting the most amount of outputs from the least amount of inputs (e.g. Robbins & Coulter, 2012). Production quantity is often used to capture efficiency because holding inputs constant, higher outputs mean lower inputs per unit of output (i.e. efficiency). Of course, the desired degree of trade-offs between creativity and/or the volume of creative production will depend on the organizational strategy (Kachelmeier et al., 2008). Some firms' competitive strategy may place greater weight on creativity and little on efficiency, while some others may put more weight on efficiency. Regardless of the degree of creativity requirement, most organizations will need some level of trade-off between volume of production and creativity.

effectiveness of incentives in eliciting creative and overall performance in two different types of creative tasks.

Specifically, I investigate open and closed creative tasks (e.g. Charness & Grieco, 2018; Unsworth, 2001; Unsworth & Luksyte, 2015). In open tasks, employees need to both discover the problem and find a creative solution (e.g. employees need to discover the next revolutionary innovation to advance mobile communication). In contrast, in closed tasks, the problem is pre-specified, and employees need to find a creative solution to this pre-specified problem (e.g. employees need to find a creative solution to reduce the weight of a mobile device while not compromising functionality).

While, in practice, some creative tasks are open, it is speculated that many creative tasks are closed (Unsworth, 2001). Examples include the work of architects and graphic designers who are often given task specifications, or employees in innovation/ research and development (R&D) labs who are engaged in finding solutions to problems in products and processes.[3] Moreover, in many organizations all employees are encouraged to be creative in solving problems arising in their daily work.

Despite the prevalence of closed creative tasks in organizations, most prior accounting studies examining the effectiveness of creativity incentives on performance, including my study in Chapter 2, focus on relatively open tasks (e.g. Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). Although Liu et al. (2019) examine the differences between open and closed creative tasks, their closed task has only one correct solution, and thus, no judgment on the "creativity" of the solution is necessary. In contrast, in

[3] It is difficult to designate entire creative professions as open or closed, because features of work can differ within professions. For instance, an architect engaged in urban planning will likely face more open tasks compared to an architect designing commercial buildings with pre-specified requirements. Hence, in this study, I focus on the unique characteristics of open/closed tasks, and these can be related to creative roles on a case by case basis.

practice, closed tasks often have a range of possible creative solutions (e.g. a logo design). I extend prior work by studying the effectiveness of incentives on creative and overall performance for *closed tasks with multiple possible solutions*, compared to open tasks.

Closed tasks are naturally constrained by the scope of the defined problem, and therefore, individuals will be more focused (Charness & Grieco, 2018; Unsworth & Luksyte, 2015). Furthermore, the task will be less ambiguous and search procedures will be more analyzable (Perrow, 1970). I expect these task attributes to impact the efficacy of incentives for creativity. Specifically, for closed tasks, I expect creativity-weighted-quantity incentives to lead to higher creative performance and overall performance, than quantity-only incentives.

I use two experiments to study the effects of incentives on performance in a creative task. The first experiment task is similar to Kachelmeier et al. (2008) (hereafter referred to as "KRW"), but with modifications to test my hypotheses. There are two main between-subject manipulations: *incentive type* (quantity-only or creativity-weighted-quantity); and *creative task* type (open or closed). For the open task, I use the same task as KRW while for the closed task I use a modified version of the same task.[4]

The key result from Experiment 1 is that, contrary to my theory-based expectation for the closed task, creativity-weighted-quantity incentives lead to significantly *lower* creative and overall performance, than quantity-only incentives. On close scrutiny, both my experiment and the only other study that finds a negative effect of incentives for closed creative tasks (i.e. Ariely et al., 2009) use high performance-based pay, while all other studies use low performance-based pay.

[4] In KRW, the experiment participants construct rebus puzzles – riddles in which words and/or diagrams are used to represent a familiar word or phrase – during a 20-minute period.

Therefore, in Experiment 2, I examine whether incentive strength, defined as the relative value of performance-based variable pay compared to fixed pay, can help explain my Experiment 1 findings on closed tasks. Based on behavioral theories on the negative effects of high-powered incentives such as "choking" under the pressure (Baumeister, 1984), I expect weaker incentives to reduce the negative effect of incentives on performance. To test this, I design a 2x2 between-subject experiment in which *incentive strength* (high or low) and *incentive type* (quantity-only or creativity-weighted-quantity) are manipulated. I only use a closed task in Experiment 2 because the purpose is to explore the unexpected finding for closed tasks in Experiment 1.

The main result from Experiment 2 is that under creativity-weighted-quantity incentives, individuals make more efficient trade-offs between creativity and quantity when incentive strength is weaker. However, regardless of incentive strength, creative and overall performance are still lower with creativity-weighted-quantity incentives than with quantity-only incentives. Thus, although the strength of incentives partially explain why I find a strong negative relationship between creativity-weighted-quantity incentives and overall performance for closed tasks, it still does not explain why the relationship is not positive, as theory on closed tasks predicts. In addition, incentive strength does not explain the negative relationship between creativity-weighted-quantity incentives and creative performance. In summary, incentive strength provides a partial explanation for my findings in Experiment 1.

My findings contribute to the management accounting literature on incentivizing creativity. Prior accounting studies suggest that including creativity measures in multidimensional compensation schemes does not affect creative performance (e.g. Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). While this discourages the need for creativity-weighted-quantity incentives in practice, there is no

apparent negative effect of incentives on the creative performance in open tasks. My finding that creativity-weighted-quantity incentives negatively affect creative performance in closed tasks extends these prior findings.

By studying the effect of incentives for closed tasks, I contribute to understanding a creativity setting that differs in ways that are important to both practice and theory. My study responds to the repeated calls for more studies that unpack the blanket term of creativity (e.g. Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015). While a few other studies have considered creativity incentives for open and closed tasks (e.g. Charness & Grieco, 2018; Cools, Stouthuysen, & Van den Abbeele, 2017; Liu et al., 2019), to my knowledge, none have examined the effectiveness of creativity-weighted-quantity incentives.

Finally, the study also examines the effect of incentive strength for closed tasks to explore the unexpected findings on closed tasks in Experiment 1 and to reconcile mixed findings in the prior literature. I show that incentive strength is an important factor that can lead to differences in performance, although this alone does not fully explain the differences between my predictions and findings for Experiment 1. Nevertheless, it provides a starting point to understand factors affecting differences in incentive effectiveness and opens avenues for future research.

Section 3.2 discusses the literature on types of creativity and rewards, hypotheses development, method and results of Experiment 1. Section 3.3 discusses hypotheses, method and results for Experiment 2. Section 3.4 provides some additional insights on how organizations can manage mediocre solutions employees produce in response to important closed problems. Section 3.5 concludes.

3.2. Experiment 1: The effect of incentivizing both creativity and quantity for creative tasks with pre-specified problems

3.2.1. Background and hypothesis development

In this section, I first introduce two distinct types of creative tasks: open and closed tasks. Then, I develop hypotheses on the effectiveness of creativity-weighted-quantity incentives for these two types of creative tasks.

3.2.1.1. Types of creativity

Recent creativity literature recognizes that "creativity" is a broad term for a range of behaviors, and that thus far, distinct types of creativity have received insufficient attention (e.g. Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015). As a result, studies in accounting and elsewhere have begun to examine differences in the use and effectiveness of control mechanisms for distinct types of creativity (e.g. Charness & Grieco, 2018; Cools et al., 2017; Gilson, Lim, D'Innocenzo, & Moye, 2012).

One such important distinction between two common, but quite different, types of creative tasks, is open and closed problems.[5] In open tasks, the individual has autonomy to identify the problem as well as find a solution (Unsworth, 2001; Unsworth & Luksyte, 2015).[6]

[5] Some other characteristics that are used to identify distinct types of creativity include the nature of the creative outcome (e.g. radical and incremental creativity) (Madjar, Greenberg, & Chen, 2011), and the driver to be creative (e.g. internal and external drivers to creativity) (Unsworth, 2001; Unsworth & Luksyte, 2015). However, a classification based on creative outcome does not consider the process of arriving at that outcome, which is what incentives have the potential to influence. Furthermore, the driver to be creative is usually external to the employee if the organization implicitly or explicitly desires creativity. Therefore, in this study I focus on the *process* of being creative, based on the nature of the problem posed.

[6] Unsworth (2001) develops a framework to unpack creativity with the open versus closed nature of the problem as one of the dimensions. She also considers a second dimension, namely the driver for creative engagement that is either external (i.e. required) or internally motivated (i.e. self-determined). Ultimately, based on these two dimensions (i.e. open/closed nature of problem and internal/external driver of engagement) Unsworth (2001) defines four types of creativity: Expected, Responsive, Proactive and Contributory. However, the second dimension (i.e. driver for creativity engagement) is not relevant to the current study because the focus is on a context where organizations desire creativity from employees (i.e. external driver to be creative). Therefore, for simplicity, I use the terms "open task" and "closed task". However, in Unsworth (2001) terminology this is synonymous to "Expected Creativity" and "Responsive Creativity", respectively.

This type of creativity represents unconstrained thinking outside the box with no obvious underlying ex-ante goal or direction (Charness & Grieco, 2018). Innovation tournaments in organizations where employees self-discover potential problems/opportunities and find creative solutions are a prime example for this type of creativity. Examples of organizations holding such tournaments include Deloitte, the Gates Foundation, and Google (Terwiesch & Ulrich, 2011). Along with this greater autonomy in open tasks, comes greater uncertainty and more complex decision-making (Unsworth & Luksyte, 2015).

In contrast, a closed task involves developing a creative solution to a specified problem (Unsworth, 2001; Unsworth & Luksyte, 2015). Such problems have a specific and delineated goal (Charness & Grieco, 2018). Hence, the employee need not engage in a search for problems, and instead, the creativity lies in finding a solution to the specified problem. Such creativity is common among creative professionals who are required to be creative with little autonomy in choosing the task specifications (e.g. architects and graphic designers), and even in traditional manufacturing/service firms (e.g. employees in R&D labs who are required to find creative solutions to specified business problems). Rockness and Shields (1984) identify three types of R&D tasks, namely, basic research, applied research, and development activities. Both applied and development activities are defined to have specific commercial objectives or ends (i.e. closed tasks). In contrast, basic research is defined as discoveries for the advancement of knowledge (i.e. an open task).[7]

The significance of distinguishing open and closed tasks in creative contexts is also implied in studies using a different typology. For instance, the open/closed classification

[7] Some employees in R&D divisions are found to prefer closed tasks. For example, in Bailyn (1985), quotes from interviewed R&D employees include "*It is not easy to find good problems*" or "*Supervisors don't tell [us] what to do, but I would prefer to be told. People want to be told; they desire strong management*" (p. 133). Bailyn (1985) finds that these employees desire discretion in the process of solving the problems that they are assigned, rather than discretion in discovering problems. This indirectly indicates the common prevalence of closed creative tasks.

provides a more nuanced understanding of the subtle degrees of analyzability of search procedures among non-routine tasks as identified by Perrow (1970). Furthermore, the difference between open and closed tasks (Unsworth, 2001) mirrors the distinction between having and not having strategic autonomy (Bailyn, 1985).[8]

No problem, however, is purely open or closed. Outside of a creativity context, a true closed problem is one that has a known method for solving while an open problem is one that has no constraints whatsoever (Unsworth, 2001). However, in a creativity context, a closed task will not have a known method for solving because, for there to be creativity, there inherently needs to be room for subjectivity. Additionally, open tasks will not be completely unconstrained due to practical constraints (Unsworth, 2001). Furthermore, the same problem can be perceived as open or closed by different individuals (Unsworth & Luksyte, 2015). Nevertheless, the degree to which a problem is closed/open can vary for different work roles.

The focus of my study, however, is not to designate creativity-related work roles as either open or closed tasks. The focus is to investigate whether the differences between the two types of creative tasks lead to differences in the efficacy of performance incentives.

3.2.1.2. Rewarding distinct types of creativity

Some past studies have examined incentives for different types of creativity. Comparing constrained (i.e. closed) and unconstrained (i.e. open) creative tasks, Charness and Grieco (2018) find that creativity incentives facilitate creativity in constrained tasks, but not in unconstrained tasks. However, neither the incentives nor the performance outcomes studied by them consider overall performance. Liu et al. (2019) find that financial incentives are negatively related to creative performance on open tasks but are positively related to creative

[8] Bailyn (1985) defines strategic autonomy in an R&D Laboratory setting as the freedom to decide one's own research direction.

performance on closed tasks with unique solutions. However, a task with a unique solution is a special case because there is inherently no variation in the creativity of the outcome, and there is no need for a subjective evaluation of the outcome's creativity. This is not generally observed in most creative work (e.g. designers, architects).[9] Furthermore, studies have not addressed how the incentive structure can guide individuals to make trade-offs between creativity and efficiency of outcomes for closed tasks compared to open tasks.

3.2.1.3. *Creative performance* in closed tasks under creativity-weighted-quantity incentives

Studies on open tasks find that creative performance (defined as number of highly creative ideas) for those with creativity-weighted-quantity incentives is similar to those with quantity-only incentives (Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). It is possible that this result reflects the open nature, and not just the creative demands, of the task. Open tasks require a broad focus due to task discovery and task solving, and individuals proceed with the problem in a relatively unplanned manner (Charness & Grieco, 2018). In addition, even if individuals realize that a trade-off should be made between investment of resources for creativity and efficiency (as communicated by the incentive structure), in the case that they haven't "discovered" a problem, they are unable to make such trade-offs. Hence, the presence of incentives for creativity is unlikely to improve creative performance.

[9] In addition to incentives, other control mechanisms have been studied in the past. Using case studies, Cools et al. (2017), find that creative firms looking for creative ideas with minimal restrictions (i.e. open tasks) use their budgets in a more interactive way, while creative firms looking for creative solutions to pre-specified problems (i.e. closed tasks) use their budgets in a more diagnostic way. Abernethy and Brownell (1997) also studied the impact of different types of control on overall performance in an R&D setting, and find that for routine tasks characterized by low variety and well-established techniques for performing tasks, accounting controls (e.g. standard costing, budgeting) appear more effective compared to personnel controls (e.g. selecting/training the "right" person) and behavioral controls.

Conversely, a closed task is pre-specified, and therefore, individuals are better equipped to act upon cost-benefit trade-offs relating to creativity. That is, they can produce a solution that is creative to the best of their ability for the given pre-specified problem and move on to the next pre-specified problem. This is unlikely in an open task, because there is no pre-specified problem to "move-on" to. In addition, in a closed task, uncertainty is reduced because there is no need for task discovery (Unsworth & Luksyte, 2015). Due to this lower uncertainty, Unsworth and Luksyte (2015) posit that rewards will help to narrow the focus on the goal, and this will lead to more successful completion of closed creative tasks. Charness and Grieco (2018) find creativity is facilitated by incentives in constrained/closed tasks. However, they only consider creativity incentives in isolation, with no consideration for quantity incentives. Liu et al. (2019) find that incentives for the number of outputs leads to better creative performance compared to fixed pay. However, due to the nature of the task in their study (i.e. a problem with a unique solution), the incentives do not include separate dimensions for creativity and quantity.

On the other hand, theory on task complexity posits that the greater the number of constraints, the greater the task complexity (Campbell, 1988; Earley, 1985). Because open tasks have less restrictions compared to closed tasks, this suggests that creativity-weighted-quantity incentives may have a detrimental effect on the creative performance on closed tasks, compared to open tasks. McGraw & McCullers (1979) find that rewards have a detrimental effect on performance for problem-solving tasks that need to break the mental set (i.e. divergent pattern of solving), while there is no such effect for problem-solving tasks with an established mental set (i.e. established pattern of solving). Thus, even for closed tasks, the form of the task can lead to differences in how individuals respond to creativity-weighted-quantity incentives.

Despite these tensions, previous empirical studies on closed tasks indicate that creativity incentives will be more effective for closed tasks. Therefore, I predict that:

Hypothesis 1 [H1]: Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to *greater creative performance* for closed tasks, but not for open tasks.

3.2.1.4. *Overall performance* in closed tasks under creativity-weighted-quantity incentives

Overall performance is a combination of the volume of production and the creativity of the products. Therefore, overall performance depends on both creative performance and the quantity of ideas produced.

It is previously established that for open tasks, overall performance of those with quantity-only incentives is significantly greater than the overall performance of those with creativity-weighted-quantity incentives (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010). That is, including creativity measures in the incentive structure fails to achieve the desired trade-offs between creativity and efficiency. In KRW, it is concluded that the time wasted by those incentivized for creativity-weighted quantity is simply a reduction in the submission of mediocre ideas. That is, their experiment participants under creativity-weighted-quantity incentives seem to spend more time on task discovery rather than on finding a solution.[10] When this need is removed (i.e. closed task), individuals that are incentivized for

[10] In KRW there is no direct measure to observe whether the participants in the creativity-weighted incentive group wasted time discovering problems (i.e. coming up with word phrases that can lead to a good rebus puzzle), or on solving the problem (i.e. designing a rebus puzzle for a word phrase they have already thought of). An additional piece of information on discarded incomplete puzzles, however, provides some indication on which of these two reasons could be the case. KRW reports that in their experiment, "the very few instances of discarded puzzles are restricted to incomplete efforts" (p.350). This provides some indication that failed attempts at designing a puzzle (solving) is low, which means that more effort is wasted on task discovery.

creativity-weighted quantity will waste less time on task discovery, and thereby increase the quantity of ideas produced.

Furthermore, using the concept of analyzability of the search function of a task (Perrow, 1970), the search for solutions in a closed task is relatively more analyzable compared to an open task because the restrictions provide some limits and perhaps a starting point to generate ideas. Therefore, for closed tasks, less time will be wasted by those incentivized for creativity-weighted quantity, increasing the quantity of ideas produced.

Alternatively, as argued in relation to creativity incentives, the closed task can have greater complexity due to the constraints created (i.e. the pre-specified problem). This can create additional challenges for the individuals to produce creative solutions, compared to the lack of restrictions in an open task. If this is the case, overall performance under creativity-weighted-quantity incentives can create additional pressures that can lead to more time spent on each task, and thus, reduce the quantity produced, as well as the overall performance.

Based on prior empirical findings, I posit the more likely scenario that closed tasks reduce the time spent on task discovery and are thus likely to be more efficient under creativity-weighted-quantity incentives. Combining the increased efficiency in closed tasks under creativity-weighted-quantity incentives with the higher creative performance predicted in H1, I predict that,

Hypothesis 2 [H2]: Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to *greater overall performance* for closed tasks, compared to open tasks.

The relationships explored in the hypotheses are illustrated in Figure 1. Variable definitions and explanations of experiment conditions are provided in Appendix A.

3.2.2. Method

3.2.2.1. Task and design

To examine the above predictions, I match my research design as closely as possible to the experiment design of KRW and Kachelmeier and Williamson (2010). The participants design rebus puzzles – "a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase" (Kachelmeier et al., 2008, p. 350) (See Appendix B for the experiment instrument).

I use a 2x2 between-subject experiment. The first manipulation is the *creative task type*: open task or closed task. The task involves designing rebus puzzles from scratch for the open task, similar to KRW. For the closed task, phrases/words that have potential to be creatively demonstrated as rebus puzzles are provided to the participants. This reflects the "pre-specified problem" feature of a closed task.

The second manipulation is the *incentive type*: quantity-only or creativity-weighted-quantity. To operationalize this, I use incentives that are as close as possible to KRW (adjusting for exchange rate and rounding differences). This enables me to replicate the basic findings of KRW. Specifically, for quantity incentives, the number of puzzles produced are the basis for the incentives. Creativity-weighted-quantity incentives are operationalized as the number of puzzles weighted for the creativity rating of each puzzle, as judged by a panel of independent raters. Regardless of the incentives, all participants are instructed that the experimenters value both creativity and quantity of the rebus puzzles submitted.

Creativity is measured by independent individuals who rate the creativity of each puzzle on a scale of 1 to 10, where "10" is the highest possible rating and "1" is the lowest possible

rating.[11,12] In all cells, incentives range from $5-$45 with the average payment maintained at about $25 across incentive conditions.

3.2.2.2. Participants

The experiment was conducted at an experimental economics laboratory at a large university, using a custom-developed web-based interface that embeds a drawing tool. Subjects were randomly drawn from a registered pool of participants pre-screened to be born in countries with English as the first language.[13] There were 92 university student participants, and of these, two participants' responses were deemed not useable, because they produced puzzles that were simply copies of the instructional examples. In addition, based on the repeated failed attempts to answer questions on the incentive structure at the beginning of the experiment, three were excluded from the analysis. The questions required participants to calculate the incentive they would receive under different scenarios presented. There were three multiple choice questions per participant, and 67 participants answered all of them correctly on the first try. There were 20 participants who answered all three questions in four attempts. The remaining three participants took five to six attempts, indicating that they possibly were not paying attention, or did not understand the incentive structure and simply got through the questions by random chance. I consider this group to have poor understanding of the incentive structure and remove

[11] Creative ratings have been used extensively in the prior literature, especially in experiments, with reasonable reliability. Past studies report high Cronbach's Alpha for inter-rater reliability for such ratings (e.g. 0.86 in Kachelmeier et al. (2008); 0.79-0.89 in Kachelmeier and Williamson (2010); 0.77 in Amabile (1982)). Other empirical measures such as the number of patents and investment in R&D have also been used in archival studies (e.g. Balkin, Markman, & Gomez-Mejia, 2000; Oldham & Cummings, 1996), but these tend to be more reflective of innovation as a whole (rather than creativity specifically).

[12] There could be a potential concern that creativity in the closed task (i.e. puzzle only) is different to the creativity in the open task (i.e. puzzle and word-phrase). While this is inevitable to the research question posed and prior studies have also been subject to this weakness (e.g. Charness & Grieco, 2018), I attempt to address this concern by obtaining additional ratings for puzzles from the open task where only the puzzle is rated, as a robustness test. Inferences are unchanged based on these ratings.

[13] Countries included were the United States, Canada, the United Kingdom, Australia, New Zealand and South Africa. In addition, participants from prior experiments with a rebus puzzle task held in the laboratory were screened out.

them from the sample. The results presented here are based on the performance of the remaining 87 participants, but the inferences made are unchanged even if all 90 participants are included.

3.2.2.3. Procedure

The experiment procedure is illustrated in Figure 2. Upon entering the laboratory, I randomly allocated participants to one of the four conditions. Participants were introduced to the respective version of the rebus puzzle task. No participant was aware of the other conditions in the experiment. After reading instructions, the participants answered questions testing their understanding of the incentive structure. Then they had a practice round of three minutes followed by the experiment task of 20 minutes. Finally, they answered a post-experiment questionnaire which included manipulation checks, questions on individual characteristics and preferences (i.e. drawing skills, prior experience with rebus puzzles, creative self-efficacy, ambiguity and risk aversion tendencies, and preferences for workplace structure), and questions on reflections of the task performed (i.e. intrinsic motivation, effort to increase quantity, effort to increase creativity). Afterwards, the participants left the laboratory with the promise of being paid in about two weeks.

I obtained creativity ratings for the puzzles using participants recruited from the online crowdsourcing platform, Prolific. First, I pre-screened the raters to be first-language English speakers, based in the United Kingdom or the United States of America, above 18 years of age, a student (for comparability with prior studies), with at least a 90% approval rate of responses on previous submissions on Prolific (for response quality), and with a minimum of 10 prior submissions on Prolific (to allow for a reasonable basis to assess response quality). There were 120 raters, each performing the task for a little over 60 minutes. All raters were paid GBP 5.00 per hour, pro rata. To reduce rater fatigue and maintain rating quality, I created subsets of

approximately 220 puzzles, and assigned raters randomly to one of the subsets. The raters received an introduction to the experiment task without information on the incentive schemes. As raters needed to be made aware of whether the word-phrase was provided (i.e. closed task) or not (i.e. open task), the two types of puzzles were rated separately.[14] The puzzles were presented in random order and each puzzle was rated by 10 raters. Of the 120 raters, 27 have a correlation coefficient below 0.2, the minimum correlation threshold used by Kachelmeier and Williamson (2010). Therefore, I exclude these ratings in the analyses.[15] To allow for inter-rater reliability, all raters rated the same 40 puzzles at the beginning, depending on whether they were assigned puzzles from the open task or the closed task. Based on these common ratings, Cronbach's Alpha was calculated as a measure of inter-rater reliability. For puzzles from the open task, this was 0.90 and for puzzles from the closed task this was 0.95. Both values exceed the common thresholds for inter-rater reliability (Kachelmeier et al., 2008; Peterson, 1994). Once the creativity ratings were obtained, I calculated the incentive payments for each participant, and made cash payments.

3.2.2.4. Key dependent variables

I calculate measures on creative and overall performance following prior literature (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010).

I capture creative performance using the number of highly creative puzzles submitted by a participant, the measure considered most meaningful in the prior literature (Kachelmeier et al., 2008). Creativity rating for each puzzle is the average rating from the panel of raters. Highly

[14] Because this may lead to concerns about a difference in the dependent variable for the two conditions, I also obtained ratings from additional 50 raters on the puzzles from the open task *assuming* that the word-phrases were provided. These ratings are highly correlated with the original ratings and the findings are not affected.
[15] I selected this threshold based on Kachelmeier and Williamson (2010). Sensitivity checks for using all raters, excluding negatively correlated raters (10 raters dropped), excluding raters below 0.1 minimum correlation (18 raters dropped), excluding the lowest correlated rater from each rater group (12 raters dropped) were conducted. Findings relating to the predicted relationships are unchanged for these alternative specifications.

creative puzzles are those puzzles with a creativity rating falling into the top 25% of all puzzles in the experiment. The remaining 75% are "mediocre" puzzles. The results presented use the 25% threshold following KRW, but sensitivity analyses for top 50%, top 30%, and top 20% creative puzzles result in similar inferences.

I also include average creativity per participant as a measure of creative performance, although prior literature criticizes how this measure can be distorted to appear high by fewer mediocre puzzles, rather than more highly creative puzzles. Despite this criticism, the measure has been commonly used in early studies, and thus, I also include it for completeness. Average creativity per participant is calculated as average of the creativity ratings for all the puzzles submitted by each participant.

I measure overall performance, which reflects the trade-off between creativity and efficiency, using creativity-weighted quantity – a combination of quantity and creativity. Quantity is the number of puzzles submitted by a participant in the 20-minute experiment task. Creativity-weighted quantity is the number of puzzles weighted for the creativity rating for each of those puzzles. This measures participants' ability to maximize overall performance.

3.2.3. Results and discussion

3.2.3.1. Variation in creativity

First, I verify whether there is meaningful variation in the creativity ratings, in terms of both the number of highly creative puzzles and average creativity. In the sample, raters used the full range of ratings from a minimum of 1 to a maximum of 10. The highest average rating for a puzzle across all raters is 8.8 and the lowest is 1.4. Figure 3a shows the distribution of highly creative puzzles across individuals and Figure 3b displays the distribution of average creativity

ratings at the individual level. As illustrated, there is variation in both measures, supporting the use of the measures as dependent variables.

3.2.3.2. Descriptive statistics

The spread of participants across the four experiment conditions and some measures of performance are presented in Table 1. There were 2,165 puzzles produced in total, with the maximum number of puzzles per participant at 58, and the minimum at 7. At a univariate level, there are differences in performance across experiment conditions for the quantity of puzzles produced, average creativity, the number of highly creative puzzles, and overall performance.

Apart from these measures, I collect some demographic and other characteristics in the post-experiment questionnaire. These will be discussed later under Section 3.2.3.4.

3.2.3.3. Hypothesis testing: Creative and overall performance

3.2.3.3.1. Effects on *creative performance*

Table 2 presents the ANOVA results for the number of highly creative puzzles per participant and the results are illustrated in Figure 4a. The number of highly creative puzzles for the open task is not significantly different across the two incentive types, replicating the findings by KRW. However, in the closed task, creativity-weighted-quantity incentives significantly *reduces* (p-value < 0.05) the number of highly creative puzzles submitted. This finding does not support H1, and in fact, provides results in the opposite direction to the hypothesized relationship.

While my finding is consistent with the argument that trying harder to be creative will not necessarily improve creativity for an open task (Kachelmeier et al., 2008), it suggests that trying harder *impedes* creative performance for the closed task. That is, if individuals are

simply incentivized for quantity (but not for creativity), they will not especially try to be creative, but ironically, they will produce more creative ideas. The findings are consistent with the alternative theory that imposing additional creativity incentive constraints on an already-constrained closed task can lead to performance anxiety and lower creative performance. However, this is still inconsistent with the majority of the prior empirical findings for closed creative tasks.

It is unlikely that the results are a manifestation of design issues of the study for several reasons. First, my results on creativity for open tasks replicate prior findings in the literature (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010). Second, the significant findings, albeit in the opposite direction to the predictions, indicate that the test does not suffer from an issue of low statistical power. Third, tests on differences across the groups in terms of demographics and personal characteristics are not significant, indicating there is no obvious reason to suspect that randomization of participants to experiment conditions failed. Therefore, in Experiment 2, I investigate one possible reason for the differences in findings compared to prior research on closed creative tasks (see section 3.3).

For completeness, I also use average creativity to measure creative performance. Table 3 presents the ANOVA results for average creativity. Average creativity is significantly higher under creativity-weighted-quantity incentives compared to quantity-only incentives for both open and closed tasks (p-value < 0.01). This replicates the finding of KRW, which uses an open task, that average creativity is higher when compensation includes creativity. However, there is an insignificant interaction between compensation and the task type, which does not support H1. Apart from this, average creativity is significantly higher for the closed task compared to the open task (p-value = 0.01), reflecting the inherent difference between the tasks and their potential for creativity. However, it is not very meaningful to compare the level of creativity

across the two tasks because the potential to be creative is different across the two. Overall, the results from this analysis replicate findings in the prior literature but do not provide much insight into the differences in incentive effectiveness between open and closed creative tasks.

3.2.3.3.2. Effect on *overall performance*

The ANOVA results for overall performance are presented in Table 4 with the graphical illustration in Figure 4b.

First, for the open task, I find that the mean overall performance under quantity-only incentives is higher than that of the creativity-weighted-quantity incentives, but the difference is not significant (p-value > 0.40). Hence, I fail to fully replicate the finding by KRW that overall performance is significantly higher for quantity-only incentives in the open task. The main driver behind this result is the difference in quantity produced in the open task. In KRW, the mean number of puzzles submitted was 28.55 (17.50) for quantity-only incentives (creativity-weighted-quantity incentives), compared to 22.19 (18.74) in my experiment. It is not clear why there is a larger difference for the number of puzzles only under quantity-only incentives. One possible reason could be that the maximum speed of drawing achieved on a digital screen (in my study) is lower than the speed achieved on pen and paper (in KRW), leading to a reduction in the power of results in my study. However, this conjecture is not tested.

For the closed task, overall performance is significantly lower for creativity-weighted-quantity incentives, compared to quantity-only incentives (p-value < 0.01). Hence, I do not find support for H2, and, instead, find significant results opposite to the hypothesized direction. This finding follows from the previous finding of lower creative performance coupled with a

significantly lower quantity (p-value < 0.01) under creativity-weighted-quantity incentives compared to quantity-only incentives (See Table 1 for cell means of quantity).

Similar to the findings for creative performance, my findings for overall performance are unexpected. That is, for the closed task, individuals are significantly worse in making trade-offs between creativity and quantity, under creativity-weighted-quantity incentives, compared to quantity-only incentives. The results are again consistent with the theory that the additional incentive constraints to an already constrained task can lead to worse performance. But this is not consistent with prior empirical findings. To explore some possible reasons for these differences in results, I first conduct some additional analysis on these results, and then, run a second experiment (See Section 3.3).

3.2.3.4. Additional analysis

3.2.3.4.1. Manipulation checks

It is possible that the results are affected by a failure in the experiment manipulations. In the post-experiment questionnaire participants answered questions about incentives and the task. First, 93% of participants correctly identified whether they were provided with word-phrases in the experiment. Excluding the failures does not significantly affect the results.

Second, all participants with creativity-weighted-quantity incentives indicated that their pay was affected by a creativity rating, but 16% incorrectly indicated that it was not affected by the number of puzzles they produced. Similarly, all participants in the quantity-only incentive condition indicated that their pay was affected by the number of puzzles they produced, but 50% also indicated that it was affected by a creativity rating, even if this was not the case.[16] If these participants who failed tests on the incentives behaved as if creativity

[16] It is difficult to compare this issue with prior studies that I closely follow (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010) because those studies do not report such manipulation checks in the papers.

affected their incentives when it did not, the differences between the incentive manipulations will be weak. This works against the prediction that creative and overall performance will be different based on the incentives. When I exclude the participants that failed these tests I find similar results for the hypotheses but lose a substantial number of participants.[17] In order to avoid over inferring results from a small number of participants, I keep all participants who failed the manipulation test in the analysis.

Third, I included a question to verify participants' understanding that both quantity and creativity were valuable to the experimenters, irrespective of the incentives. There were nine participants who either answered either "no" or "uncertain" to this. Excluding these participants does not lead to changes in inferences.[18] As before, I keep all participants in the reported analyses.

3.2.3.4.2. Complexity/difficulty of open and closed tasks

A key concern when comparing open and closed tasks is the possibility that differences in task complexity and difficulty (Campbell, 1988; Earley, 1985; Adams, Bessant, & Phelps, 2006) may confound the effect of open and closed nature of the tasks.

Predicting the differences in complexity of the two tasks is not straightforward. On the one hand, closed tasks have a well-defined outcome, which leads to the lack of need for task discovery. A well-specified desired outcome is important in reducing the complexity of a task (Campbell, 1988). On the other hand, the closed task has an additional restriction compared to

[17] 29 participants failed the manipulation test leaving only 58 observations for analysis. The lowest number of participants per condition is 7 for the open task with quantity-only incentives. All analyses were repeated using these participants and results were similar.

[18] Nine participants failed this understanding question leaving 78 observations for the analysis. The lowest number per condition is 18. All analyses were repeated using these participants and results were either similar or stronger. However, to be conservative, to be comparable to the procedures followed in prior literature, and to minimize the loss of participants, I keep all participants that failed the understanding check in the analysis.

the open task (i.e. the puzzle should not only be creative, but also match the phrase/word provided). In general, the greater the number of constraints, the greater is the complexity (Campbell, 1988; Earley, 1985). However, if one constraint facilitates the achievement of the other, the complexity could be reduced (Campbell, 1988).

To test if findings on the number of highly creative puzzles for the closed task was driven by greater task complexity (Campbell, 1988; Earley, 1985; Adams et al., 2006), I ask participants to rate how complex they perceived the experiment task was on an 11-point Likert scale. However, participants assigned to the open task rated the tasks as significantly *more* complex than those assigned to the closed task (p-value < 0.02), indicating that this is unlikely to explain the finding on closed tasks.

In addition, the participants assigned to both closed and open tasks rated the difficulty of constructing a creative rebus puzzle for each word-phrase they submitted on an 11-point Likert scale. Descriptive statistics of the self-assessed difficulty ratings are presented in Table 5. There is no significant difference in the difficulty ratings between conditions (p-values > 0.30), providing some assurance that results are not driven by task difficulty.

3.2.3.4.3. Personal characteristics

I find no significant difference in the spread of participants across conditions on gender, area of study, self-estimated drawing skills, a measure of ambiguity aversion, a measure of risk aversion, a measure of intrinsic motivation for the task, and previous rebus puzzle experience (p-values > 0.15). Level of education, an indicator variable coded 1 for a tertiary qualification and 0 otherwise, is significantly lower for the creativity-weighted-quantity incentives (p-value = 0.09). In addition, the average age of the participants is significantly higher in the open task with quantity incentives compared to conditions with creativity-weighted-quantity incentives

(*p*-values < 0.08). Including these variables as covariates does not change the inferences.[19] In addition, I find that participants' perception of "openness" of the task (Unsworth & Luksyte, 2015) was significantly larger for the open task than for the closed task (*p*-value < 0.01), reflecting the intended manipulation of tasks.

In Chapter 2, I find that individuals with high creative capacity will have significantly higher creative performance under creativity-weighted-quantity incentives for an open task. In a sub-sample analysis, I replicate this finding on creative capacity for the open task (Table 6 Panel A). For the closed task (Table 6 Panel B), there is a significant drop in the number of highly creative puzzles under creativity-weighted-quantity incentives for the individuals with low creative capacity but not for those with high creative capacity. However, regardless of creative capacity, individuals with creativity-weighted-quantity incentives have poorer overall performance, compared to quantity-only incentives. Overall, this indicates that for the closed task, creativity-weighted-quantity incentives are ineffective, regardless of individual creative capacity. However, caution is required in making strong inferences given the small sample size.

3.3. Experiment 2: The effect of strength of incentives for creative tasks with pre-specified problems

3.3.1. Background and hypothesis development

In Experiment 1, I find that creativity-weighted-quantity incentives for the closed task lead to significantly lower creative and overall performance (compared to quantity-only incentives).

[19] In an ANCOVA with the level of education as a covariate, education is not significantly related to average creativity, the number of highly creative puzzles or overall performance (*p*-values > 0.30 two-tailed). The significance of the interactions for creativity incentives and closed tasks are similar (*p*-value = 0.24 for average creativity, *p*-value = 0.03 for the number of highly creative puzzles and *p*-value < 0.01 for overall performance, all one tailed). Age is also not significantly related to average creativity, the number of highly creative puzzles or overall performance (*p*-values > 0.20 two-tailed). The significance of the interactions for creativity incentives and closed tasks are similar (*p*-value = 0.32 for average creativity, *p*-value = 0.03 for the number of highly creative puzzles and *p*-value = 0.01 for overall performance, all one tailed).

Prior studies on closed tasks find both a positive effect of creativity incentives on performance (e.g. Charness & Grieco, 2018; Liu et al., 2019), and a negative effect (Ariely et al., 2009). However, direct comparisons across these studies and my Experiment 1 are difficult because of different incentive structures and task characteristics. Notably, the studies that find a positive relationship use lower performance-based variable pay compared to fixed pay, than the studies, including my Experiment 1, that find a negative relationship. Therefore, in Experiment 2, I examine whether the strength of incentives partly explains the mixed findings in the prior literature for closed creative tasks.[20]

Historically, research in psychology has examined the effects of stimulation strength. Albeit using punishments rather than rewards, Yerkes and Dodson (1908), found that the strength of a stimulus is positively related to habit formation only up to some point, after which, it suffers, especially for more difficult tasks. Hebb (1955) argues that a stimulation to a mild (strong) degree may lead to desirable (undesirable) outcomes. Baumeister (1984) suggests that very large incentives may lead to "choking under pressure". Following this early research, many studies explore a range of related concepts such as intensity of motivation, drive level, reward magnitude and generally find evidence that performance is affected in a non-linear manner (Teigen, 1994).

When incentives are strong, those with creativity-weighted-quantity incentives face high financial stakes if they are not creative. This can result in "choking" under the pressure (Baumeister, 1984), which can be counter-productive, leading to less creativity. Similarly, with weak incentives, individuals will not "choke" and will be better able to make the trade-offs

[20] While Ariely et al. (2009) argue that the magnitude of incentives causes choking behavior for creative tasks, they simultaneously manipulate both the strength of incentives, (i.e. the relative value of performance-based variable pay compared to fixed pay), and scale of incentives (i.e. the average incentive values across experiment conditions are not held constant), limiting the ability to isolate the effect of incentive strength. Furthermore, they use a creative task with one unique solution which is relatively uncommon in practice.

between creativity and quantity. They will not waste effort by inefficiently trying to be creative, and instead, also focus on increasing quantity. Therefore, I predict that for a closed task:

Hypothesis 3 [H3]: Weak incentives will alleviate the negative relationship between creativity-weighted-quantity incentives and *creative performance*.

Hypothesis 4 [H4]: Weak incentives will alleviate the negative relationship between creativity-weighted-quantity incentives and *overall performance*.

Figure 5 illustrates these hypotheses. Variable definitions and explanations of experiment conditions are provided in Appendix A.

3.3.2. Method

I design a second 2x2 between-subject experiment similar to Experiment 1, using only a closed task (i.e. constructing rebus puzzles to pre-specified phrases/words). The first manipulation, *incentive type*, is the same as Experiment 1 (i.e. quantity-only incentives or creativity-weighted-quantity incentives). The second manipulation, *incentive strength*, is either high performance-based-incentives ($5-$45, same as Experiment 1) or low ($20-$30, new to Experiment 2), keeping the average pay per cell at about $25. See Appendix C for the changes in the experiment instrument compared to Experiment 1.[21]

I conducted the experiment at the same experimental economics laboratory as Experiment 1 and randomly drew participants from the same subject pool, with the same qualifying criteria.[22] Due to limitations in the participant pool, I was only able to recruit 59

[21] Apart from changes relating to incentive strength, additional questions were added to Experiment 2 relating to preoccupation with the reward, pressure, stress, time-pressure, and public and private self-consciousness. This is to capture potential psychological effects of strong/weak incentives. Two questions from Experiment 1 irrelevant to Experiment 2 were dropped (i.e. self-rating of task difficulty, self-rating on the open-closed nature of the task).
[22] In addition to pre-screens for being born in English-speaking countries, participants from Experiment 1 and participants from other studies using the same task were excluded.

participants.[23] Of those, two did not design rebus puzzles to match the word-phrases provided, and thus, are not useable. In addition, based on the repeated failed attempts to answer questions on the incentive structure at the beginning of the experiment, I exclude four more participants from the analysis, leaving a final sample of 53 participants.[24,25]

Given the limited sample size, I prioritized collecting data for the weak incentive conditions as these are the new conditions in Experiment 2, compared to Experiment 1. To obtain an acceptable sample size, I merged data on closed tasks in Experiment 1 with data from Experiment 2. The experiment procedure was identical for both experiments (See Figure 2).

Creativity ratings for the 1,627 puzzles were obtained using the same process as described for Experiment 1, leading to a total of 140 raters, each rating approximately 160 puzzles. Of the 140 raters, 12 with a correlation coefficient below 0.2 were dropped in the results calculation.[26] Inter-rater reliability is high, with Cronbach's Alpha at 0.97, based on the 40 initial puzzles rated by all raters.

[23] Because of the nature of the task, proficiency in the English language is important for a fair comparison between subjects. The University has a large proportion of international students from countries where English is not the first language, and this significantly limited the available pool of participants. After excluding participants in Experiment 1, and one other related study by the same group of researchers, there was only a small group of potential participants remaining. Using a different research facility (e.g. at another university) was considered unsuitable because of possible differences in the population characteristics.

[24] Like in Experiment 1, the questions required participants to calculate the incentives for three hypothetical scenarios. 41 participants answered all the multiple-choice questions correctly in the first try and 12 participants answered with one failed attempt. The remaining four participants took five to ten attempts, indicating that they were not paying attention, or did not understand the incentive structure. Like in Experiment 1, the results are presented after excluding these four participants. The findings are unchanged even if all 59 are included.

[25] Excluding the two participants that incorrectly identified the range of incentives they could possibly earn strengthens the findings. A quarter of the participants with creativity-weighted-quantity incentives incorrectly indicated that their pay was not affected by the number of puzzles they produced, and 59% under quantity-only incentives also incorrectly indicated that their pay was affected by a creativity rating. Five participants answered either "no" or "uncertain" to the question on whether they were instructed that the experimenters valued both quantity and creativity. Excluding the participants that failed these tests leads to a substantial loss in the number of participants without a significant difference to the findings. I keep all participants in the main analysis.

[26] This threshold was selected based on Kachelmeier and Williamson (2010). Sensitivity checks for other thresholds, namely, including all raters, excluding negatively correlated raters (4 raters dropped), excluding raters below 0.1 minimum correlation (7 raters dropped) were conducted. Inferences relating to the predicted findings are unchanged for these alternative thresholds.

3.3.3. Results and discussion

3.3.3.1. Pooling data from Experiment 1

Due to the lower sample size in Experiment 2 (see footnote 23), I combine data on closed tasks from Experiment 1 with data from Experiment 2 data for the results analysis. Before doing this, I compare the key performance measures of the two samples (Table 7 Panel A). There are no statistically significant differences in any of the performance measures between comparable conditions of Experiments 1 and 2 (p-values > 0.25). This is not surprising because the experiment instructions and the procedure were identical, the laboratory and the pool from which participants were drawn was the same, and the experimenter was unchanged. I also find no significant differences across the samples for gender, age, level of education, self-estimated drawing skills, risk aversion, intrinsic motivation, and previous rebus puzzle experience (p-values > 0.11). However, the average ambiguity aversion of the group with strong creativity-weighted-quantity incentives is higher in Experiment 1 compared to Experiment 2 (p-value = 0.01). Including ambiguity aversion as a covariate does not significantly alter the results.[27] In order to further verify that pooling data does not affect the findings, I conduct all the analyses with Experiment 2 data only, and all the inferences are supported.

3.3.3.2. Descriptive statistics

The descriptive statistics of participants' performance are presented in Table 7 Panel B. There were 2,936 puzzles produced in total, 1,627 from Experiment 2 and 1,309 from Experiment 1. At a univariate level, quantity-only incentives, compared to creativity-weighted-quantity incentives, lead to higher number of puzzles, lower average creativity, higher number of highly

[27] In the ANCOVA analysis, ambiguity aversion is not significantly related to creative or overall performance (p-values > 0.50, two tailed). All the findings of the study are unchanged.

creative puzzles, and higher overall performance (p-values < 0.01).[28] The effects are observed even with the reduced sample from Experiment 2. These replicate results from Experiment 1 that, for closed tasks, creativity-weighted-quantity incentives (compared to quantity-only incentives) lead to lower creative and overall performance. Univariate differences between weak and strong incentive conditions are not significant (p-values > 0.35).

In addition to these measures, I collect demographic and other characteristics in the post-experiment questionnaire. I find no significant difference in the spread of participants across the four conditions based on area of study, self-estimated drawing skills, ambiguity aversion, risk aversion, intrinsic motivation, and previous rebus puzzle experience. However, the weak quantity-only incentives group is significantly older, with higher educational qualifications (p-values < 0.10), while the weak creativity-weighted-quantity incentives group has significantly fewer males, significantly less preference for structure in the work place and significantly less confidence in their own creativity (p-values < 0.10). Including gender, age, level of education, preference for workplace structure and self-assessment of creative performance as covariates does not change the inferences of results.[29]

3.3.3.3. Hypothesis testing: Creative and overall performance

3.3.3.3.1. Effect on *creative performance*

Table 8 Panel A presents the ANOVA results for the number of highly creative puzzles per participant and results are illustrated in Figure 6a. The number of highly creative puzzles is

[28] In the creativity ratings for Experiment 2, raters have used the full range of ratings from a minimum of 1 to a maximum of 10 (untabulated). The highest average rating for a puzzle across all raters is 8.8 and the lowest is 1.1. The mean rating per puzzle is 4.7 with a standard deviation of 1.2, indicating that there is variation in the creativity ratings.

[29] In the ANCOVA analyses, male participants have a significantly lower number of highly creative puzzles (p-value < 0.10, two tailed), and having a tertiary qualification is significantly negatively related to average creativity (p-value < 0.05, two tailed). None of the other covariates are significantly related to creative or overall performance. Importantly, all the inferences of the main results are unchanged after including the covariates in the analyses.

significantly lower under creativity-weighted-quantity incentives (compared to quantity-only incentives) regardless of the strength of incentives. While this replicates the finding from Experiment 1 that for a closed task, creativity-weighted-quantity incentives significantly reduce the number of highly creative puzzles (*p*-value = 0.02, one-tailed), I find no moderating effect from incentive strength (i.e. H3 is not supported).

The takeaway is that, the drop in the number of highly creative ideas under creativity-weighted-quantity incentives is not explained by the strength of incentives. If individuals are simply incentivized on quantity, they will still produce significantly more highly creative ideas. The failure to detect an effect from incentive strength could reflect that weak incentives were "not sufficiently weak" in my experiment design. I am unable to test this proposition.

Table 8 Panel B presents the ANOVA results on average creativity, for completeness. Average creativity is significantly higher under creativity-weighted-quantity incentives compared to quantity-only incentives irrespective of incentive strength (*p*-value < 0.01, one-tailed). This replicates the finding in Experiment 1 that average creativity is higher when compensation includes creativity. However, against H3, the average creativity across weak and strong incentives is virtually the same with no interaction between compensation and incentive strength.

3.3.3.3.2. Effect on *overall performance*

The ANOVA results for overall performance are presented in Table 8 Panel C with the graphical illustration in Figure 6b. Overall performance is significantly lower under creativity-weighted-quantity incentives (compared to quantity-only incentives) for weak incentives (*p*-value = 0.04, one-tailed) as well as strong incentives (*p*-value < 0.01, one-tailed). However, the slopes are different, as shown in the significant interaction (*p*-value = 0.03, one-tailed). That

is, the decrease in overall performance with creativity-weighted-quantity incentives is significantly less when the incentives are weak, compared to strong. This supports H4, which predicted that incentive strength will negatively moderate the relationship between creativity-weighted-quantity incentives and overall performance.

However, it remains that regardless of the strength of incentives, in a closed task with multiple possible solutions, creativity-weighted-quantity incentives (compared to quantity-only incentives) have a significantly negative effect on creative and overall performance. While weakening the strength of incentives does significantly dampen this decrease, creative performance and overall performance are still significantly better under quantity-only incentives. Although it can be argued that, perhaps, even weaker incentives may have led to better performance, such a linear extrapolation is not straightforward because Ariely et al. (2009) finds that very small incentive magnitudes can also lead to poor creative performance.

In the post-experiment questions for Experiment 2, participants also rated, on a scale of 1 to 10, whether they found themselves "thinking about consequences of the reward" they might receive. The purpose of this variable, *Preoccupied*, was to examine whether strong incentives lead to preoccupation of potential rewards, which, in turn, leads to lower performance. I find that individuals in the weak-incentives conditions were significantly less preoccupied with the potential reward compared to individuals in the strong-incentives conditions (p-value = 0.04).[30] However, I do not find evidence that *Preoccupied* mediates the relationship between creativity-weighted-quantity incentives, and overall performance for weak or strong pay conditions (p-values > 0.20). Including *Preoccupied* as a covariate in the ANOVA strengthens the finding that overall performance is reduced under creativity-

[30] *Preoccupied* is only available for data from Experiment 2, which reduces the sample size to 53, with only 18 observations for the strong-incentives conditions. Therefore, this analysis is only meant as a descriptive exploration without strong reliance on statistical inferences.

weighted-quantity incentives (compared to quantity-only incentives) for strong incentives (p-value < 0.01), but not for weak incentives (p-value = 0.30).

Overall, in Experiment 2, I find that weaker incentives can reduce, but not eliminate, the detrimental effect of creativity-weighted-quantity incentives on the efficient trade-offs between creativity and quantity for closed tasks. This finding partially explains the mixed results in the prior literature. For instance, Liu et al. (2019) find a positive relationship between incentives and creative performance, with maximum performance-based incentives about 50% of total pay. Similarly, Charness and Grieco (2018) find a positive relationship using maximum performance-based incentives about 60% of total pay. In contrast, Ariely et al. (2009) find a negative relationship with extreme incentives, where incentive-based pay comprises 100% of total pay (i.e. no base pay). In Experiment 1 of the study, I too find a negative relationship using very high maximum performance-based incentives of approximately 90% of total pay. While these studies are not directly comparable (i.e. there are significant differences in task characteristics and incentive structure), together with the results from Experiment 2, they provide evidence that incentive strength may explain some differences in overall performance.

3.4. The problem of mediocre solutions

In both experiments in this study, I show that for the closed task, quantity-only incentives (compared to creativity-weighted-quantity incentives) lead to more creative puzzles as well as more mediocre puzzles (i.e. puzzles that are not designated as "highly creative"), regardless of incentive strength differences. It can be argued that, all problems specified by an organization (i.e. closed tasks) are important, and thus, having many such problems solved in a mediocre fashion may be harmful.

One way to address this challenge in organizations is to re-allocate issues that are solved in a mediocre fashion by one individual to another. This will give the opportunity for a second person to attempt the issue and potentially solve it more creatively. If not, the rotation can continue. This strategy will only be effective if all the employees can produce highly creative ideas at least some of the time. While studying this was not an initial intention of the experiment, my data offers the opportunity to provide preliminary insights into the feasibility of this strategy. Specifically, I check whether mediocre performance and high performance are spread across individuals in a way such that an individual with a mediocre idea for one word-phrase has the potential to have a highly creative idea for another word-phrase. Table 9 presents some descriptive statistics that this may be the case. In Experiment 1, over one third of the participants had at least one minimum-rated puzzle (coded as the puzzle with the lowest creativity score for a given word-phrase), but over 78% of that group also had at least one maximum-rated puzzle (coded as the puzzle with the highest creativity score for a given word-phrase). The comparative percentages in Experiment 2 are 81% and 65%, respectively. Panel B provides more descriptive statistics on a possible nonlinear, concave, relationship between the number of minimum- and maximum-rated puzzles for participants. These statistics indicate that there is potential to rotate closed tasks among individuals in order to get the greatest number of highly creative solutions. However, additional costs of employing such tactics will also need to be considered.

3.5. Conclusion

In this study, I examine the effectiveness of creativity-weighted-quantity incentives for closed tasks. While prior studies in accounting find that creativity-weighted-quantity incentives are ineffective for open tasks (Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010), other studies argue that creativity is not a blanket term, and effectiveness of incentives may differ

based on the type of creative task (Unsworth, 2001; Unsworth & Luksyte, 2015). Recent studies in economics and accounting have found that explicit incentives for creativity are desirable for closed tasks, but not for open tasks (Charness & Grieco, 2018; Liu et al., 2019).

I extend this literature by comparing multidimensional incentives for open and closed tasks. Specifically, I look at quantity-only incentives and creativity-weighted-quantity incentives. Contrary to expectations, I find that for closed tasks, using multidimensional incentives can be harmful, compared to using quantity-only incentives. Multidimensional incentives both impede production quantity as well as lead to a reduction in creative performance. In contrast, under quantity-only incentives, a focus on quantity leads to high production output as well as highly creative outputs.

In a follow-up experiment, I also examine the effect of incentive strength on the observed relationships between incentive structure and performance outcomes. I find that the negative effect of creativity-weighted-quantity incentives for overall performance on closed tasks is reduced when the incentives are weaker. However, there is no such effect of incentive strength on creative performance. That is, under weaker creativity-weighted-quantity incentives, individuals make comparatively better trade-offs between creativity and quantity, but not necessarily have better creativity. However, quantity-only incentives always lead to better creative and overall performance compared to creativity-weighted-quantity incentives, irrespective of incentive strength.

My findings contribute to the literature as follows. First, I identify differences in incentive effects for different types of creative tasks and incentive strengths. Prior studies find that the harmful effects of creativity-weighted-quantity incentives lie in the inefficient trade-offs between quantity and creativity caused, but not in creative performance per se (e.g. Kachelmeier et al., 2008; Kachelmeier & Williamson, 2010). However, this study presents a

type of closed task where creativity-weighted-quantity incentives can negatively influence both creative and overall performance. Understanding the effectiveness of incentives for such different creative tasks is vital because creative tasks vary in practice (Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015).

Second, I attempt to reconcile some mixed findings in the literature on the effectiveness of creative incentives for closed tasks. Specifically, some studies find a positive relationship between creative incentives and performance on a creative task, while others, including results from Experiment 1 in this study, show a negative relationship. I contribute to the literature by showing that incentive strength is a partial reason for these mixed findings. This leaves room for future studies to explore other contributory factors for the mixed findings.

One limitation of my study is that it does not consider what can be done about mediocre ideas. If a firm has pre-specified issues requiring creative solutions, quantity-only incentives may not be desirable because of the larger number of mediocre solutions. A potential solution would be to rotate tasks among individuals, or let multiple individuals undertake each task, in order to achieve highly creative outcomes. Descriptively, I find that this is possible, because individuals who perform least creatively on some problems perform most creatively on others. Another limitation is that I am only able to partially explain the reasons for the unexpected findings in Experiment 1. In addition, I only have the minimum number of participants due to practical reasons, and the inherent limitations in generalizing findings from a laboratory experiment remain. However, it is unlikely that any of these concerns significantly lessen the value of the key finding that creativity-weighted-quantity incentives can be detrimental to performance on closed tasks, and weaker incentives can reduce this effect.

Appendix

Appendix A: Variable definitions and experiment conditions

Dependent variable definitions	
Number of puzzles	The total number of puzzles submitted by a participant
Number of highly creative puzzles	The total number of puzzles considered to be "highly creative" by each participant. "Highly creative" puzzles are those with creativity ratings in the top 25% in the pool of all puzzles submitted. Creativity rating per puzzle is defined below.
Average creativity	The average of creativity ratings for all the puzzles submitted by each participant. Creativity rating per puzzle is defined below.
Overall performance	The number of puzzles submitted by each participant weighted for the creativity rating for each of those puzzles. Creativity rating per puzzle is defined below.
Creativity rating per puzzle	The average of the individual ratings for each puzzle by a panel of independent raters.
Independent variable definitions	
Creativity_Pay	An indicator variable that takes the value of 1 if the participant received creativity-weighted-quantity incentives, and 0 if participant received quantity-only incentives. See below for descriptions of experiment conditions.
Closed_Task	An indicator variable that takes the value of 1 if the participant performed a closed task, and 0 if participant performed an open task. See below for descriptions of experiment conditions.
Weak_Pay	An indicator variable that takes the value of 1 if the participant is offered weak incentives, and 0 if the participant is offered strong incentives. See below for descriptions of experiment conditions.

Experiment conditions	
Incentive type	
Quantity-only incentives	Incentives based on the number of puzzles produced per participant.
Creativity-weighted-quantity incentives	Incentives range from $5 to $45, based on a creativity-weighted total score, which is the sum of creativity ratings of all puzzles produced per participant. Creativity rating per puzzle is defined above.
Task type	
Open task	Constructing rebus puzzles from scratch
Closed task	Constructing rebus puzzles for word-phrases provided by the experimenters
Incentive strength	
Strong incentives	Incentives range from $5 to $45, with an approximate average of $25 per participant.
Weak incentives	Incentives range from $20 to $30, with an approximate average of $25 per participant.

Appendix B: Experiment instrument used in Experiment 1

<Adapted from Kachelmeier et al. (2008)>

Ground Rules

Before describing the experiment, it is important to establish ground rules.

- NO COMMUNICATION WITHIN OR OUTSIDE SESSIONS

Please help us maintain control over the experiment by refraining from commenting or communicating with your fellow participants or with others who might be participating in this research. You will be working individually during this experiment. If you have any questions, please tap on the table and raise your hand. One of the experimenters will assist you.

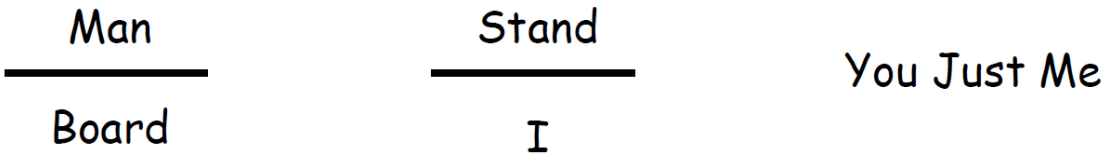
- NO DECEPTION

We promise to carry out the experiment in the manner described in these instructions, with no deception of any form. **We will make the payment for this experiment at a later date (in about two weeks)**. But we promise that your pay will be determined exactly as described in the rules explained later.

Instructions

Task

In this research, you will be constructing "rebus puzzles". A rebus puzzle is a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase. Here are some examples:

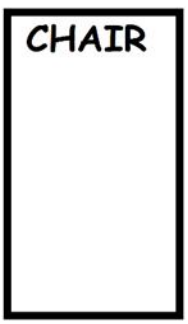


These examples use the positioning of the words to create the riddle.

- The first is "man overboard", because "man" is over "board".
- The second is "I understand", for similar reasons.
- The third is "just between you and me" (i.e., "just" is between "you" and "me")

Instructions (continued)

Other rebuses use counting, different sizes, shapes, or positions to create the riddle, such as these examples:

Rebus Puzzle	Word-Phrase/ Solution
0 M.D. Ph.D. B.A.	Three degrees below zero
Funny Funny Words Words Words Words	Too funny for words (2 funny; 4 words)
pains	Growing pains
R R O A D S A D S	Cross Roads
	High chair

Instructions (continued)

Still other rebuses use simple pictures, symbols, and diagrams as part of the riddle, like these two examples:

Rebus Puzzle	Word-Phrase/ Solution
	A hole in one
BB 	To be or not to be

What we would like you to do:

PART 1:

[Open task]

In this experiment, you will be presented with a blank drawing pad. Using the stylus provided to you, construct a single rebus puzzle in the drawing pad. Remember to type the word-phrase in the box below the drawing pad. Then Click "Submit". Once you submit a puzzle, it is considered finished and you will be redirected to a fresh drawing pad. Then proceed to design the next puzzle and repeat the submission process.

While we do not place any rules on the kind of rebus puzzles you can submit, we value both the ***number of different puzzles*** you can construct (i.e. **quantity**) and the ***creativity*** of those puzzles (i.e. **puzzles that are original ideas, innovative, and clever**).

You will have 20 minutes to construct your rebus puzzles.

[Closed task]

In this experiment, you will be presented with a blank drawing pad with a word-phrase at the top. Using the stylus provided to you, construct a single rebus puzzle in the drawing pad to match the word-phrase provided. Remember to type the word-phrase (appearing at the top) in the box below the drawing pad. Then click "Submit". Once you submit a puzzle, it is considered finished and you will be redirected to a fresh drawing pad with a new word-phrase at the top. Then proceed to design the next puzzle and repeat the submission process.

While we do not place any rules on the kind of rebus puzzles you can submit, we value both the ***number of different puzzles*** you can construct (i.e. **quantity**) and the ***creativity*** of those puzzles (i.e. puzzles that are **original ideas, innovative, and clever**).

You will have 20 minutes to construct your rebus puzzles.

Payment:

[Quantity-only incentives]

Your payment will be based on ***how many*** rebus puzzles you construct in 20 minutes. To determine this, we will count the number of rebus puzzles you construct.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) who constructs the ***most*** rebus puzzles in this group will receive **$ 45.00**. The participant (or participants, if tied) who constructs the ***fe*****west** rebus puzzles in this group will receive **$ 5.00**. Everyone else in the group will receive an amount between $ 45.00 and $ 5.00, depending on the number of rebus puzzles constructed. This will result in an expected average payment of approximately **$ 25.00**. The more puzzles you construct, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

[Creativity-weighted-quantity incentives]

Your payment will be based on ***the creativity-weighted total score*** of all rebus puzzles you construct in 20 minutes.

To determine this score, we will ask a group of independent raters to rate the creativity of each puzzle you submit on a 1 to 10 scale, where "10" is the highest rating and "1" is the lowest rating. We will then ***add*** the creativity ratings of ***all*** puzzles you submit to arrive at the creativity-weighted total score. Thus, ***each puzzle you submit helps your total score, but higher rated puzzles count more***. For example, at the extreme, a puzzle rated 10 counts ten times as much as a puzzle rated 1.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) with the ***highest*** creativity-weighted total score in this group will receive **$ 45.00**. The participant (or participants, if tied) with the ***lowest*** creativity-weighted total score in this group will receive **$ 5.00**. Everyone else in the group will receive an amount between $ 45.00 and $ 5.00, the individual creativity-weighted total score. This will result in an expected average payment of approximately **$ 25.00**. The higher your creativity-weighted total score, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

Checking your understanding:

Let's see if you understood the payment scheme:

[Quantity-only incentives]

Suppose your randomly assigned group has 4 participants. If you produce 8 puzzles, and the three others produce 6, 7, and 8 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If you produce 20 puzzles, and the three others produce 25, 30 and 30 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If you produce 15 puzzles, and the three others produce 6, 18 and 18 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

[Creativity-weighted-quantity incentives]

Suppose you produce 5 puzzles and a panel of independent raters awards them creativity scores as 4, 7, 2, 6, and 9 respectively. What is your creativity-weighted total score?

- 28
- 5.6
- 9

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 45, and the three others receive 35, 40, and 29 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 60, and the three others receive 88, 60, and 79 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 33, and the three others receive 35, 60, and 29 respectively, what is the payment you will receive?

- $ 5
- $ 45
- An amount between $ 5 and $ 45

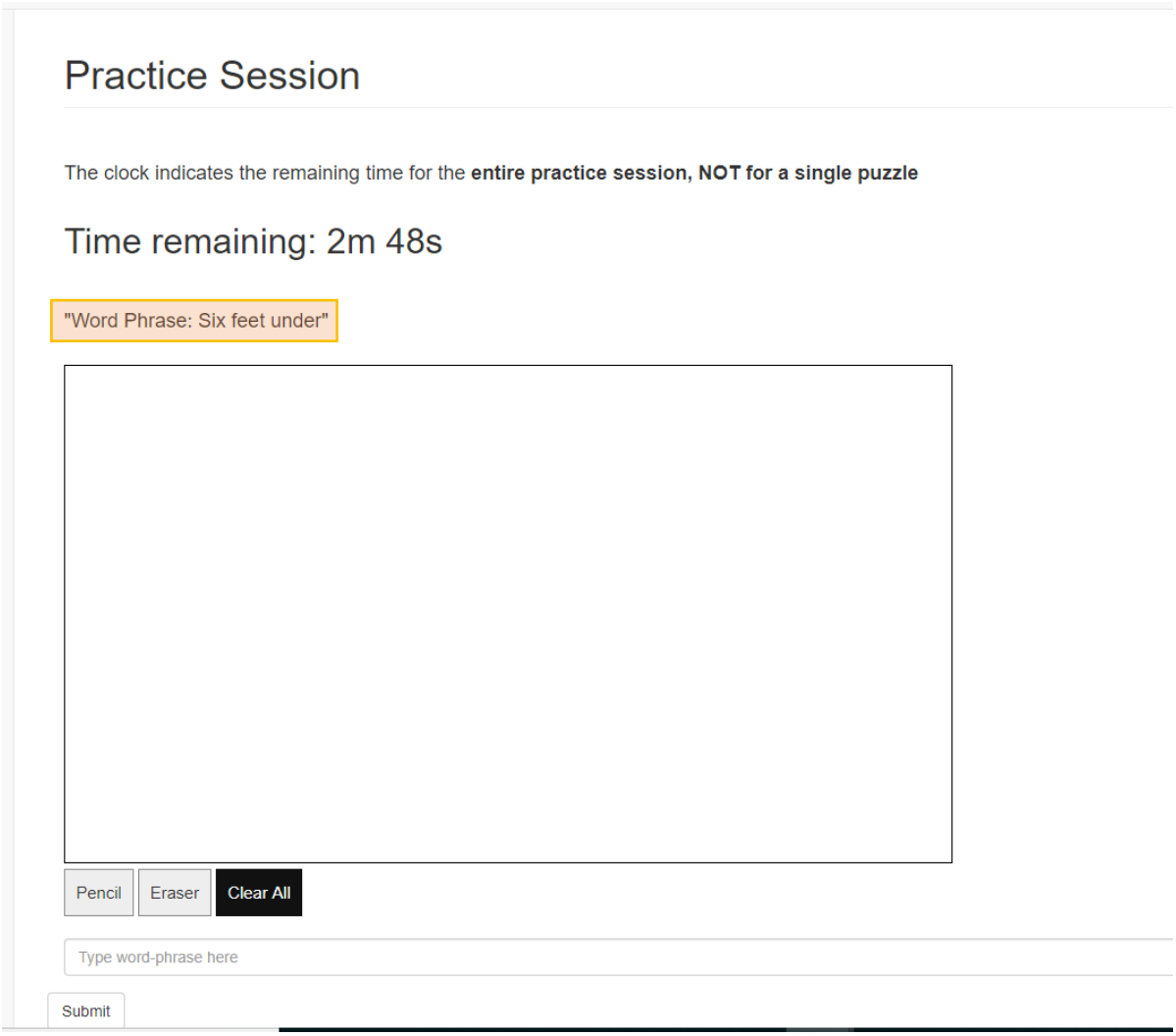
Practice Session

Now let us have a practice session. Performance in this section will **NOT count towards your pay**. It is simply an opportunity to familiarize yourself with the drawing interface and with the puzzle submission process.

In the next screen you will see the drawing pad. After constructing one puzzle on the drawing pad, type the word-phrase in the space provided and click "Submit". Then you will see a new drawing pad on which you can construct a new puzzle. Accordingly, you can construct and submit as many puzzles as you like.

You have **three minutes to practice**. There will be a clock on the screen that will count-down the 3 minutes for your reference.

Click "Start" to proceed.

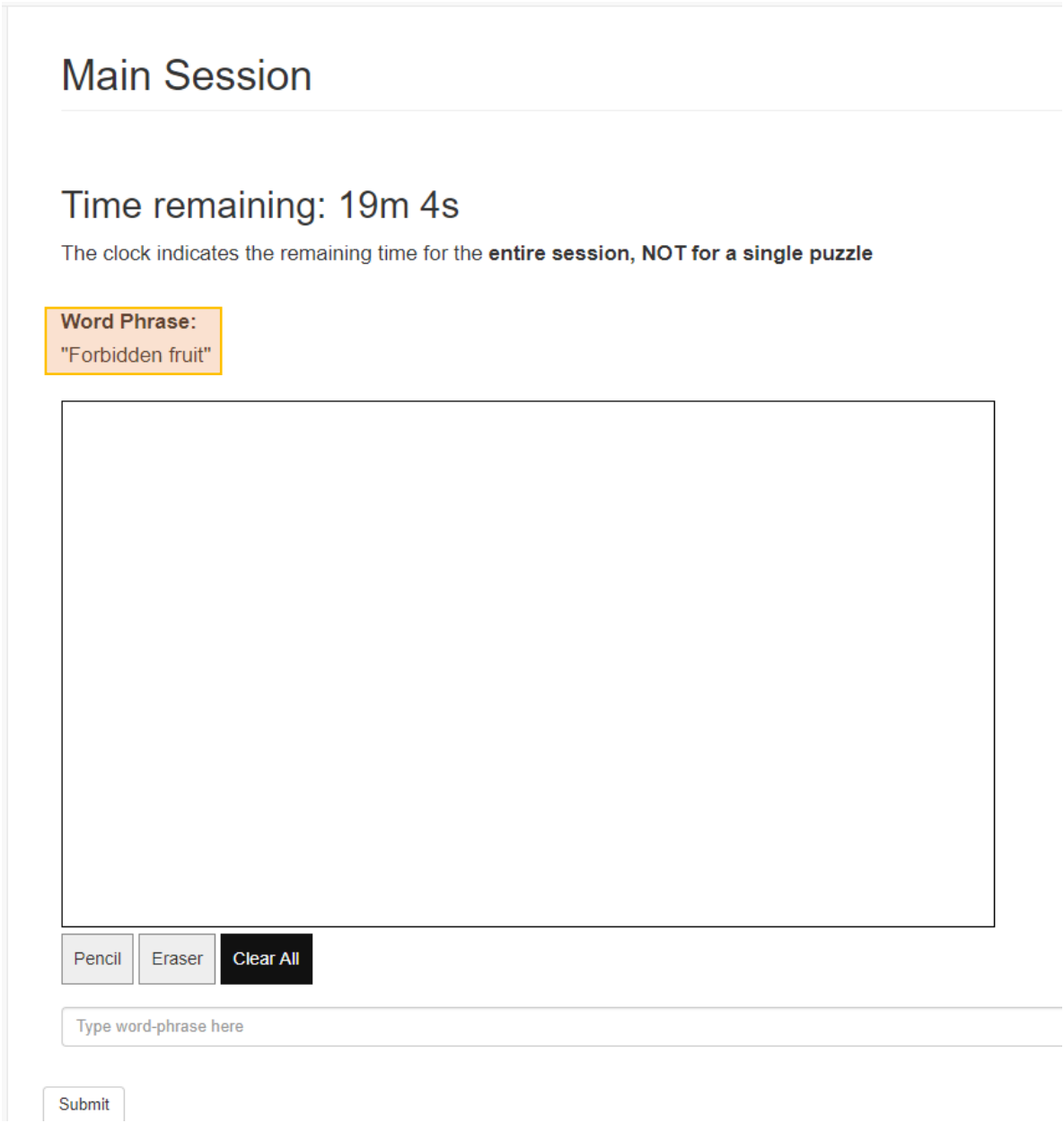


Main Session

Now let us start the experiment.

Performance in this section **WILL DETERMINE YOUR PAY** as described earlier. You have **20 minutes**. There will be a clock on the screen that will count-down the 20 minutes for your reference.

Please click "Start" to proceed.



Instructions

PART 2:

In this section, you will be asked some questions. Your payment will not be affected by the answers provided. However, it is important that you answer these questions to complete the experiment.

Click "Next" to continue.

Please answer the next few questions <u>about the task you performed</u> in the experiment.

Do you agree with each of the following statements?

	Yes	**No**	**Not Certain**
In the experiment, I was provided with a word-phrase for which I had to construct a rebus puzzle.	○	○	○
In the experiment, I had to generate a word-phrase for which I had to construct a rebus puzzle.	○	○	○

This question is **about your pay**. Do you agree with the following statements?

	Yes	**No**	**Not Certain**
The payment you will receive in this experiment is affected by a ***creativity rating for each puzzle*** you produced.	○	○	○
The payment you will receive in this experiment is affected by ***how many puzzles*** you produced.	○	○	○

This question is about the **instructions you read** for the experiment. Do you agree with the following statement?

	Yes	**No**	**Not Certain**
The instructions indicated that we value **both quantity and creativity** in the puzzles produced	○	○	○

Please rate on the scale below, how complex you believe the task in the experiment was.

Not Complex										**Highly Complex**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Please rate **how difficult** it is to construct **creative** rebus puzzles for the following word phrases you worked on. Remember**, a creative rebus puzzle** is one that is an **original, innovative and clever idea.**

	Very Easy									**Very difficult**
Word phrase you worked on	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
[Insert word-phrase 1 by participant]	○	○	○	○	○	○	○	○	○	○	○
[Insert word-phrase 2 by participant]	○	○	○	○	○	○	○	○	○	○	○
…	○	○	○	○	○	○	○	○	○	○	○

To what extent do you agree or disagree with the following statements?

	Strongly Disagree	**Disagree**	**Somewhat disagree**	**Neither Agree nor Disagree**	**Somewhat Agree**	**Agree**	**Strongly Agree**
	1	**2**	**3**	**4**	**5**	**6**	**7**
I enjoyed designing rebus puzzles in the experiment	○	○	○	○	○	○	○
My rebus puzzles were likely more creative than the average participant's in this study	○	○	○	○	○	○	○
I worked hard to increase the **number** of rebus puzzles I constructed	○	○	○	○	○	○	○
I worked hard to increase the **creativity** of the rebus puzzles I constructed	○	○	○	○	○	○	○

Open" and "Closed" are two types of problems.

- In an **open problem**, individuals both identify a problem, and find a solution to the identified problem
- In a **closed problem**, individuals find solutions to a pre-specified problem

However, no problem is purely closed or open. Most problems can be placed on a range from closed to open.

Please rate the degree to which you believe that the task you performed in this study was open or closed.

Closed Problem										**Open Problem**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Please rate on the scale below the degree to which you were familiar with rebus puzzles **before participating in this study**.

Never Heard										**Highly Familiar**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

What specific prior familiarity did you have with rebus puzzles? (select the most appropriate answer)

- I had no prior familiarity with rebus puzzles
- I had solved rebus puzzles in the past, but had not designed any
- I had designed rebus puzzles in the past, but had not solved any
- I was aware of rebus puzzles, but had never designed or solved any
- I had designed and solved rebus puzzles in the past

Please answer the next few questions <u>about yourself, more generally</u>.

Using the following responses please indicate the extent to which you agree or disagree that each statement currently describes you.

	1 - Very strongly disagree	2 - Strongly disagree	3 - Disagree	4 - Neutral	5 - Agree	6 - Strongly agree	7 - Very strongly agree
[Statement 1 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
[Statement 2 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○
[Statement 3 from the creative self-efficacy instrument by Tierney and Farmer (2002)]	○	○	○	○	○	○	○

Suppose you are given $100. You can invest any portion of this money in an investment opportunity. There is an **unknown chance** that this investment will be successful. If the investment is successful, you will be paid 2.5 times the amount invested. If the investment is not successful, you will lose the money you invested. You will retain any money you do not invest. How much of the $100 will you invest in this opportunity?

(Please provide an amount between 0 and 100 without decimal points.)
$ [ ]

People differ in the amount of structure they prefer in their working environment. What degree of structure do you prefer in your working environment?

Highly Unstructured									Highly Structured
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Please rate the degree to which you consider yourself skilled in **drawing**.

Not Skilled										**Highly Skilled**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Please fill the following information about yourself.

- Gender:
- Age:
- Country currently residing in:
- Country of birth:
- Is English your first language? Yes/No
- Current area of study:
- What is the highest level of education you have completed?

Suppose you are given $100. You can invest any portion of this money in an investment opportunity. There is a 50% chance that this investment will be successful. If the investment is successful, you will be paid 2.5 times the amount invested. If the investment is not successful, you will lose the money invested. You will retain any money you do not invest. How much of the $100 will you invest in this opportunity?

(Please provide an amount between 0 and 100 without decimal points.)
$ [ ]

Appendix C: Changes in the experiment instrument for Experiment 2 compared to Experiment 1

What we would like you to do:

PART 1:

[Closed task only]

In this experiment, you will be presented with a blank drawing pad with a word-phrase at the top. Using the stylus provided to you, construct a single rebus puzzle in the drawing pad to match the word-phrase provided. Remember to type the word-phrase (appearing at the top) in the box below the drawing pad. Then click "Submit". Once you submit a puzzle, it is considered finished and you will be redirected to a fresh drawing pad with a new word-phrase at the top. Then proceed to design the next puzzle and repeat the submission process.

While we do not place any rules on the kind of rebus puzzles you can submit, we value both ***the number of different puzzles*** you can construct (i.e. **quantity**) and the ***creativity*** of those puzzles (i.e. puzzles that are **original ideas, innovative, and clever**).

You will have 20 minutes to construct your rebus puzzles.

[Strong incentives] [Weak incentives]

Payment:

[Quantity-only incentives]

Your payment will be based on ***how many*** rebus puzzles you construct in 20 minutes. To determine this, we will count the number of rebus puzzles you construct.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) who constructs the ***most*** rebus puzzles in this group will receive **$ 45.00 [30.00]**. The participant (or participants, if tied) who constructs the ***fewest*** rebus puzzles in this group will receive **$ 5.00 [20.00]**. Everyone else in the group will receive an amount between $ 45.00 [30.00] and $ 5.00 [20.00], depending on the number of rebus puzzles constructed. This will result in an expected average payment of approximately **$ 25.00**. The more puzzles you construct, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

[Creativity-weighted-quantity incentives]

Your payment will be based on ***the creativity-weighted total score*** of all rebus puzzles you construct in 20 minutes.

To determine this score, we will ask a group of independent raters to rate the creativity of each puzzle you submit on a 1 to 10 scale, where "10" is the highest rating and "1" is the

lowest rating. We will then ***add*** the creativity ratings of ***all*** puzzles you submit to arrive at the creativity-weighted total score. Thus, ***each puzzle you submit helps your total score, but higher rated puzzles count more***. For example, at the extreme, a puzzle rated 10 counts ten times as much as a puzzle rated 1.

Upon entering the experiment, you have been randomly assigned to a group of participants. The participant (or participants, if tied) with the ***highest*** creativity-weighted total score in this group will receive **$ 45.00 [30.00]**. The participant (or participants, if tied) with the ***lowest*** creativity-weighted total score in this group will receive **$ 5.00 [20.00]**. Everyone else in the group will receive an amount between $ 45.00 [30.00] and $ 5.00 [20.00], depending on the number of rebus puzzles constructed. This will result in an expected average payment of approximately **$ 25.00**. The higher your creativity-weighted total score, the more money you will make.

We promise to make the payment, calculated as described above, in about two weeks.

[Strong incentives] [Weak incentives]

Checking your understanding:

Let's see if you understood the payment scheme:

[Quantity-only incentives]

Suppose your randomly assigned group has 4 participants. If you produce 8 puzzles, and the three others produce 6, 7, and 8 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

Suppose your randomly assigned group has 4 participants. If you produce 20 puzzles, and the three others produce 25, 30 and 30 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

Suppose your randomly assigned group has 4 participants. If you produce 15 puzzles, and the three others produce 6, 18 and 18 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

[Creativity-weighted-quantity incentives]

Suppose you produce 5 puzzles and a panel of independent raters awards them creativity scores as 4, 7, 2, 6, and 9 respectively. What is your creativity-weighted total score?

- 28
- 5.6
- 9

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 45, and the three others receive 35, 40, and 29 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 60, and the three others receive 88, 60, and 79 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

Suppose your randomly assigned group has 4 participants. If your work receives a creativity-weighted total score of 33, and the three others receive 35, 60, and 29 respectively, what is the payment you will receive?

- $ 5 [20]
- $ 45 [30]
- An amount between $ 5 [20] and $ 45 [30]

PART 2:

[New questions]

This question is based on the **instructions you read** about your payment for the experiment. What are the **minimum and maximum** amounts **you** could receive for today's performance?

- $ 20 minimum, $ 30 maximum
- $ 5 minimum, $ 45 maximum
- $ 25 fixed (no more, no less)

To what extent do you agree or disagree with the following statements?

	Strongly Disagree	Disagree	Somewhat disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
	1	2	3	4	5	6	7
While constructing puzzles, I found myself thinking about the payment I might receive	○	○	○	○	○	○	○
I felt a lot of pressure to be creative when constructing rebus puzzles	○	○	○	○	○	○	○
Performing the task in the experiment was stressful	○	○	○	○	○	○	○
I found it easier to perform the task towards the beginning of the 20 minutes, compared to the end.	○	○	○	○	○	○	○

Please rate on the scale below, **how difficult** you believe the task in the experiment was.

Not Difficult										**Very Difficult**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
○	○	○	○	○	○	○	○	○	○	○

Using the following responses please indicate the extent to which you agree or disagree that each statement currently describes you.

	1 - Very strongly disagree	**2 - Strongly disagree**	**3 - Disagree**	**4 - Neutral**	**5 - Agree**	**6 - Strongly agree**	**7 - Very strongly agree**
I generally pay attention to how I feel internally about my own behavior	○	○	○	○	○	○	○
I generally pay attention to what others would think of my behavior	○	○	○	○	○	○	○

[Removed questions]

Please rate **how difficult** it is to construct **creative** rebus puzzles for the following word phrases you worked on. Remember**, a creative rebus puzzle** is one that is an **original, innovative and clever idea.**

	Very Easy									**Very difficult**
Word phrase you worked on	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
[Insert word-phrase 1 by participant]	○	○	○	○	○	○	○	○	○	○	○
[Insert word-phrase 2 by participant]	○	○	○	○	○	○	○	○	○	○	○
…	○	○	○	○	○	○	○	○	○	○	○

Please rate on the scale below, how complex you believe the task in the experiment was.

Not Complex										Highly Complex
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Open" and "Closed" are two types of problems.

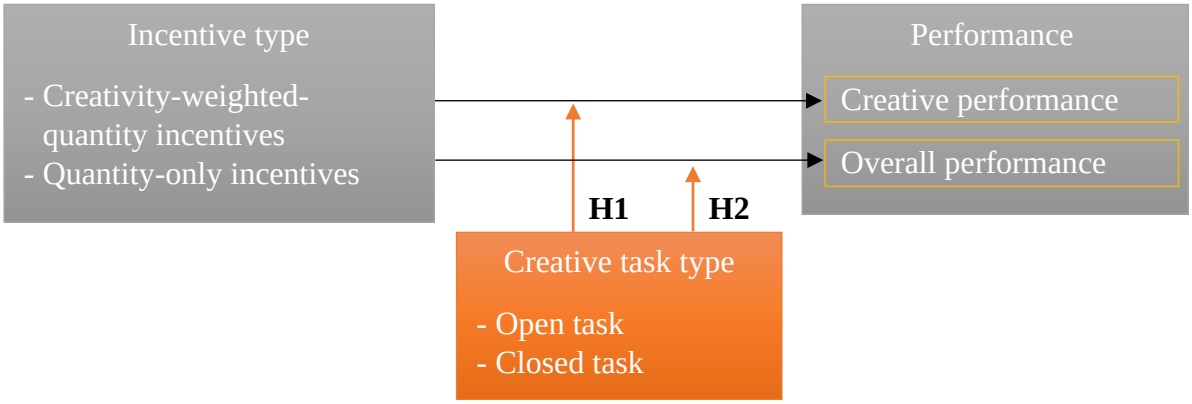
- In an **open problem**, individuals both identify a problem, and find a solution to the identified problem
- In a **closed problem**, individuals find solutions to a pre-specified problem

However, no problem is purely closed or open. Most problems can be placed on a range from closed to open.

Please rate the degree to which you believe that the task you performed in this study was open or closed.

Closed Problem										Open Problem
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Figures and tables



H1	Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to *greater creative performance* for closed tasks, but not for open tasks.
H2	Creativity-weighted-quantity incentives, relative to quantity-only incentives, will lead to *greater overall performance* for closed tasks, compared to open tasks.

Figure 1: Experiment 1 expectations. This figure illustrates the predicted effects of creative task type on the relationship between incentive types and performance. See Appendix A for descriptions of experiment conditions.

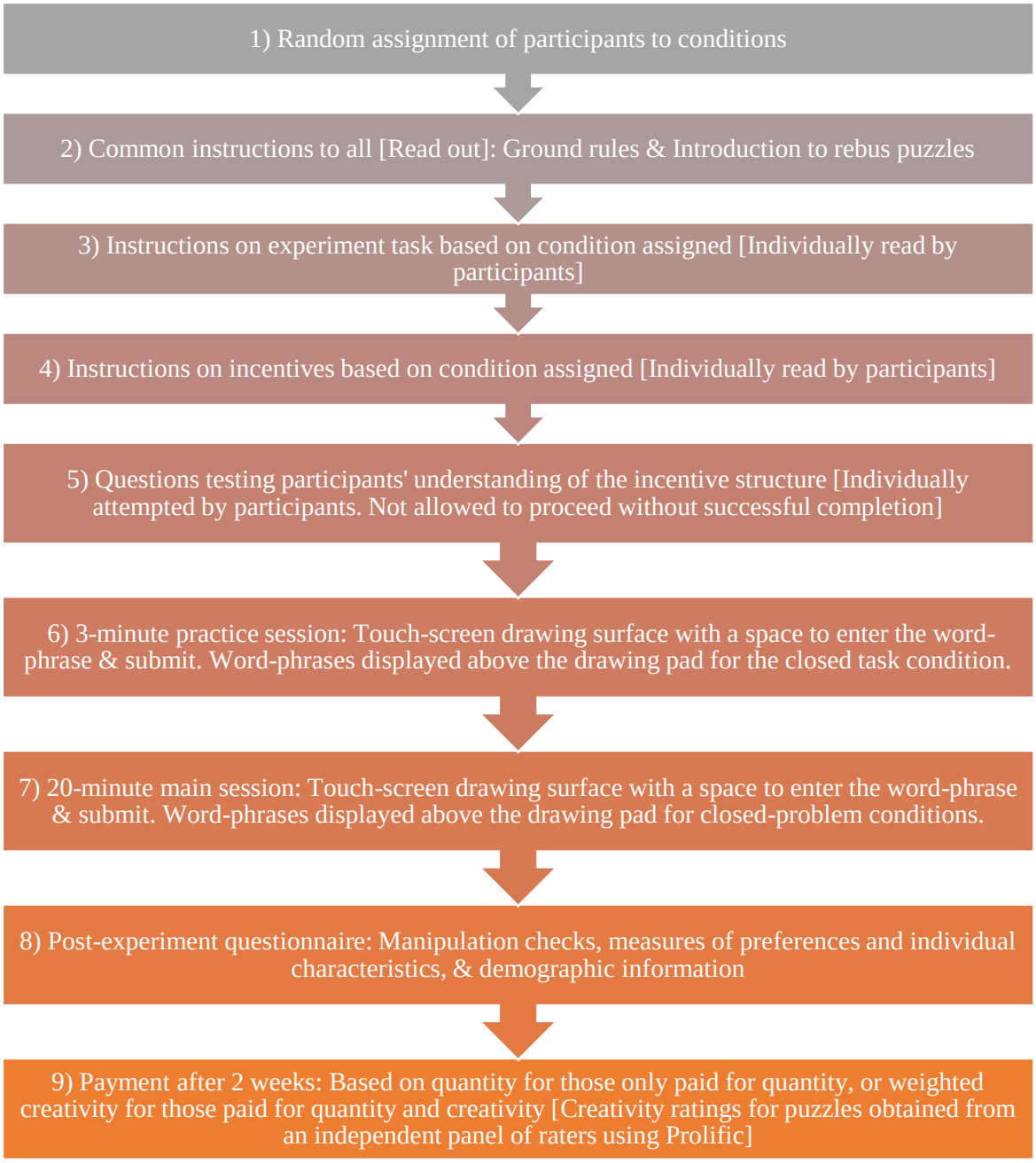
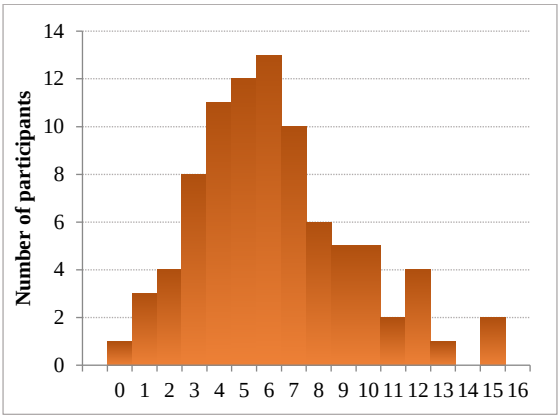


Figure 2: Experiment procedure. This figure illustrates the stepwise procedure followed to collect data in both Experiments 1 and 2.

(a) Number of highly creative puzzles



(b) Average creativity

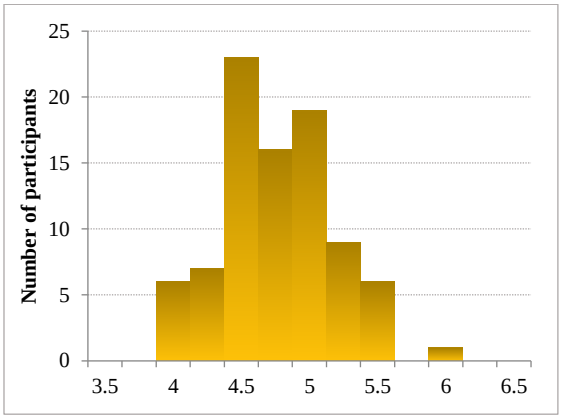
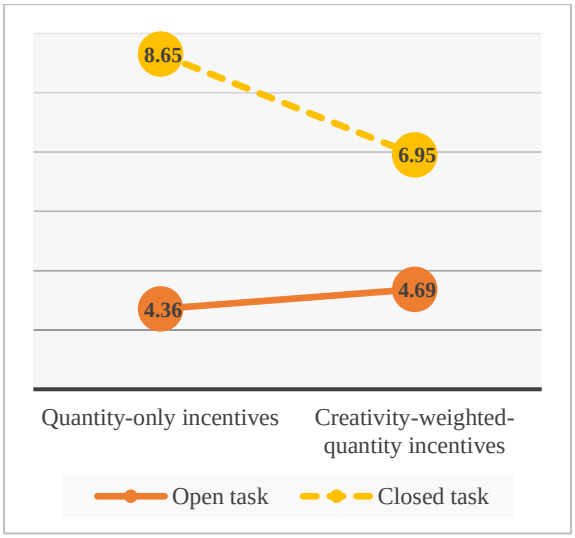


Figure 3a-b: Distribution of creative performance in Experiment 1. 3a. Distribution of the number of highly creative puzzles. 3b. Distribution of average creativity per participant. See Appendix A for variable definitions.

(a) Creative performance: Number of highly creative puzzles



(b) Overall performance

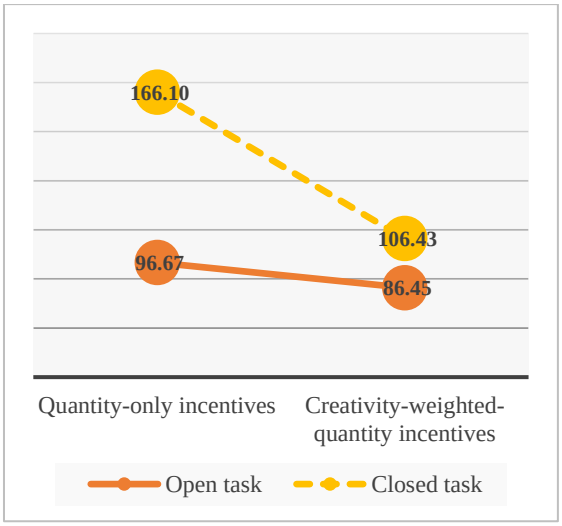
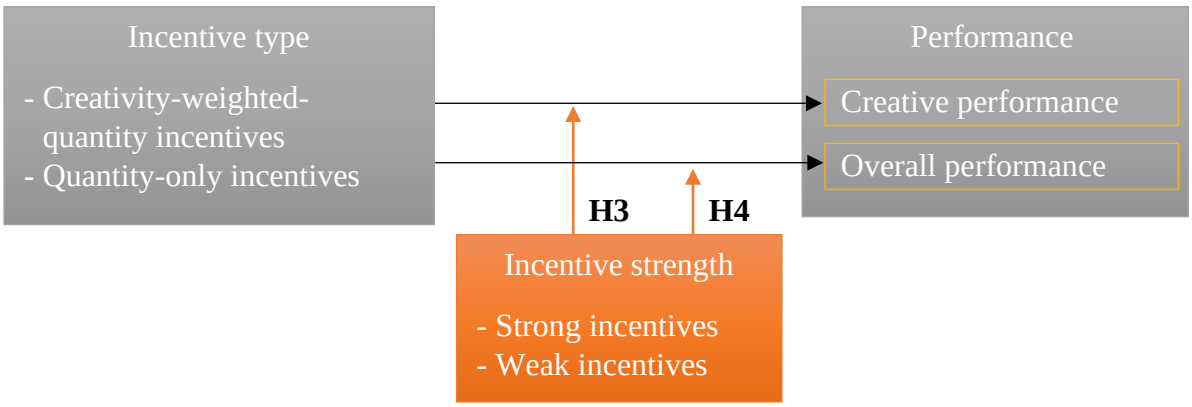


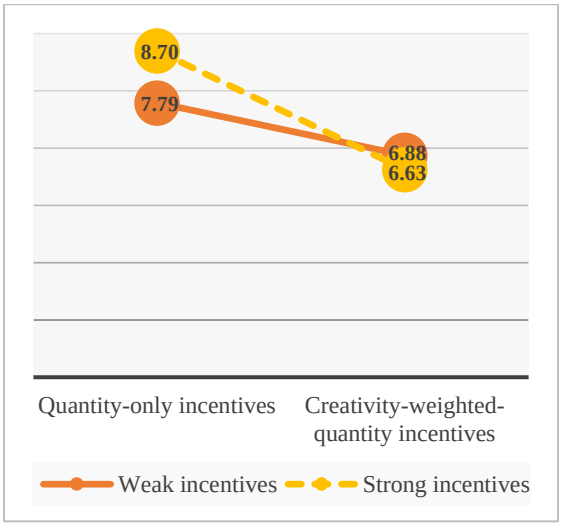
Figure 4a-b: Illustration of Experiment 1 results. This figure shows the effect of creative task type and incentive type on performance. 4a. Results on the mean number of highly creative puzzles. 4b. Results on mean overall performance. See Appendix A for variable definitions and descriptions on experiment conditions. For all groups, incentives ranged from $5 to $45, with an approximate average of $25 per participant.



H3	Weak incentives will alleviate the negative relationship between creativity-weighted-quantity incentives and *creative performance*.
H4	Weak incentives will alleviate the negative relationship between creativity-weighted-quantity incentives and *overall performance*.

Figure 5: Experiment 2 expectations. This figure illustrates the predicted effects of incentive strength on the relationship between incentive types and performance in closed tasks. See Appendix A for descriptions of experiment conditions**.**

(a) Creative performance: Number of highly creative puzzles



(b) Overall performance

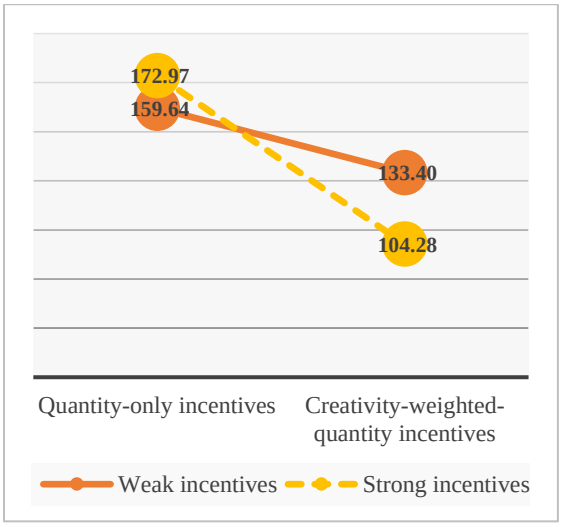


Figure 6a-b: Illustration of Experiment 2 results. This figure shows the effect of incentive strength and incentive type on performance in a closed task. 6a. Results on the mean number of highly creative puzzles. 6b. Results on mean overall performance. See Appendix A for variable definitions and descriptions on experiment conditions.

Table 1: Descriptive statistics for Experiment 1

		Open task + Quantity-only incentives	Open task + Creativity-weighted-quantity incentives	Closed task + Quantity-only incentives	Closed task + Creativity-weighted-quantity incentives	Differences Open versus Closed (*p*-value)	Differences Quantity-only versus Creativity-weighted-quantity (*p*-value)
		A	B	C	D	(A+B) vs (C+D)	(A+C) vs (B+D)
Number of participants	N	21	21	23	22		
Number of puzzles per participant	Mean	22.19	18.57	35.78	22.09	**<0.01**	**<0.01**
	Min	8.00	7.00	20.00	9.00	.	.
	Max	37.00	38.00	58.00	33.00	.	.
	SDev	8.90	7.33	10.60	7.89	.	.
Average creativity per participant	Mean	4.36	4.69	4.66	4.86	**0.01**	**<0.01**
	Min	3.79	3.93	4.02	4.06	.	.
	Max	5.24	5.29	5.14	5.99	.	.
	SDev	0.39	0.37	0.29	0.48	.	.
Number of highly creative puzzles per participant	Mean	4.29	4.81	8.65	6.95	**<0.01**	0.33
	Min	0.00	1.00	3.00	1.00	.	.
	Max	10.00	8.00	15.00	13.00	.	.
	SDev	2.31	2.02	3.35	2.75	.	.
Overall performance per participant	Mean	96.67	86.45	166.10	106.43	**<0.01**	**<0.01**
	Min	34.07	33.74	101.68	44.36	.	.
	Max	178.50	168.71	275.47	155.79	.	.
	SDev	34.07	31.84	47.89	35.99	.	.

This table presents descriptive statistics for the four experiment cells in Experiment 1, and differences across the two manipulations. See Appendix A for variable definitions and descriptions on experiment conditions. Incentives ranged from $5 to $45, with an approximate average of $25 per participant. The *p*-values for differences are two-tailed because prior literature does not provide basis for a clear directional prediction.

Table 2: ANOVA for the effect of task type and incentives on the number of highly creative puzzles in Experiment 1

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	7.48	1.05	0.31
Closed_Task	1	230.22	32.18	**<0.01**
Creativity_Pay Closed_Task*	1	26.79	3.75	**0.06**
Simple effects				
Effect of *Creativity_Pay* given *Closed_Task* =0	1	2.88	0.61	0.527
Effect of *Creativity_Pay* given *Closed_Task* =1	1	32.41	3.43	**0.036**

This table presents ANOVA results and simple effects on the number of highly creative puzzles for Experiment 1. See Appendix A for variable definitions and descriptions on experiment conditions. Incentives ranged from $5 to $45, with an approximate average of $25 per participant. The *p*-values for differences are two-tailed.

Table 3: ANOVA for the effect of task type and incentives on average creativity in Experiment 1

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	1.55	10.21	**<0.01**
Closed_Task	1	1.25	8.24	**0.01**
Creativity_Pay Closed_Task*	1	0.09	0.57	0.45

This table presents ANOVA results on average creativity for Experiment 1. See Appendix A for variable definitions and descriptions on experiment conditions. Incentives ranged from $5 to $45, with an approximate average of $25 per participant. The *p*-values are two-tailed.

Table 4: ANOVA for the effect of task type and incentives on overall performance in Experiment 1

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	26,521.49	17.02	**<0.01**
Closed_Task	1	43,410.97	27.85	**<0.01**
Creativity_Pay Closed_Task*	1	13,275.70	8.52	**<0.01**
Simple effects				
Effect of *Creativity_Pay* given *Closed_Task* =0	1	1,096.94	0.85	0.404
Effect of *Creativity_Pay* given *Closed_Task* =1	1	40,033.28	22.17	**<0.01**

This table presents ANOVA results and simple effects on overall performance for Experiment 1. See Appendix A for variable definitions and descriptions on experiment conditions. Incentives ranged from $5 to $45, with an approximate average of $25 per participant. The *p*-values are two-tailed.

Table 5: Descriptive statistics on the perceived difficulty of constructing creative puzzles for each word-phrase in Experiment 1

	Open task + Quantity-only incentives	Open task + Creativity-weighted-quantity incentives	Closed task + Quantity-only incentives	Closed task + Creativity-weighted-quantity incentives
Number of participants	21	21	23	22
Number of puzzles rated for difficulty	458	383	795	485
Average puzzle-difficulty per participant	4.17	4.48	4.59	4.27
Minimum puzzle-difficulty per participant	1.67	1.48	1.61	1.32
Maximum puzzle-difficulty per participant	7.19	7.90	9.13	7.91
Number of highly difficult puzzles per participant	3.81	4.90	8.87	5.05

This table presents descriptive statistics on the perceived difficulty of constructing rebus puzzles for word-phrases. *Difficulty* is a self-rating (on a scale of 0-10) by participants on each word-phrase for which they submitted a puzzle. *Average puzzle difficulty per participant* is the average of the difficulty ratings for all the puzzles by each participant. *Minimum puzzle difficulty per participant* is the lowest difficulty rating for a puzzle provided by each participant. *Maximum puzzle difficulty per participant* is the highest difficulty rating for a puzzle provided by each participant. *Number of highly difficult puzzles per participant* is the total number of puzzles by each participant classified as "highly difficult". "Highly difficult" puzzles are those with self-ratings for difficulty in the top 25% in the pool of all puzzles rated for difficulty. See Appendix A for descriptions on experiment conditions.

Table 6: The effect of individual creative capacity on incentive effectiveness in Experiment 1

Panel A: Open task

Dependent variable	Incentive condition	N		Means	
		HCC	LCC	HCC	LCC
Number of puzzles	Quantity-only	12	9	25.33 [1]	18.00
	Creativity-weighted-quantity	14	7	20.07 [1]	15.57 [1]
Number of highly creative puzzles	Quantity-only	12	9	4.58	3.89
	Creativity-weighted-quantity	14	7	5.71 [2]	3.00
Overall performance	Quantity-only	12	9	109.62 [3]	79.40
	Creativity-weighted-quantity	14	7	94.79 [3]	69.76

Panel B: Closed task

Dependent variable	Incentive condition	N		Means	
		HCC	LCC	HCC	LCC
Number of puzzles	Quantity-only	8	15	36.25	35.53
	Creativity-weighted-quantity	10	12	21.60 [4]	22.50 [4]
Number of highly creative puzzles	Quantity-only	8	15	8.50	8.73
	Creativity-weighted-quantity	10	12	8.40	5.75 [5]
Overall performance	Quantity-only	8	15	165.29	166.53
	Creativity-weighted-quantity	10	12	109.93 [6]	103.51 [6]

This table presents descriptive statistics and tests of differences for individuals with high and low creative capacity. Panel A presents the descriptive statistics for participants that performed an open task and Panel B presents the descriptive statistics for participants that performed a closed task. See Appendix A for key variable definitions and descriptions on experiment conditions. *Creative capacity* is measured using a 3-item, 7-point Likert scale by Tierney and Farmer (2002) leading to one factor score. Participants with self-reported creative capacity above the median for the sample are designated as the *high creative capacity* (*HCC*) group. Participants with self-reported creative capacity below the median for the sample are designated as the *low creative capacity* (*LCC*) group.

[1] From ANOVA, the number of puzzles is significantly lower for creativity-weighted-quantity incentives, compared to quantity-only incentives (p-value = 0.06, one-tailed), and significantly higher for the HCC group, compared to the LCC group (p-value = 0.01, one-tailed).

[2] From ANOVA, the number of highly creative puzzles is significantly higher for creativity-weighted-quantity incentives, compared to quantity-only incentives, for the HCC group (p-value < 0.10 one-tailed), but not for the LCC group.

[3] From ANOVA, overall performance is significantly higher for the HCC group, compared to the LCC group, regardless of incentive type (p-value = 0.01, one-tailed).

[4] From ANOVA, the number of puzzles is significantly lower for creativity-weighted-quantity incentives, compared to quantity-only incentives (p-value < 0.01, one-tailed) regardless of individual creative capacity.

[5] From ANOVA, the number of highly creative puzzles is significantly lower for creativity-weighted-quantity incentives, compared to quantity-only incentives, for the LCC group (p-value < 0.10 one-tailed), but not for the HCC group.

[6] From ANOVA, overall performance is significantly lower for creativity-weighted-quantity incentives, compared to quantity-only incentives (p-value < 0.01, one-tailed) regardless of individual creative capacity.

Table 7: Descriptive statistics for Experiment 2

Panel A: Comparison of key performance measures from Experiment 2 to comparable cells in Experiment 1

		Experiment 1 + Quantity-only incentives [Strong Incentives]	Experiment 2 + Quantity-only incentives + Strong Incentives	Difference *p*-value	Experiment 1 + Creativity-weighted-quantity incentives [Strong Incentives]	Experiment 2 + Creativity-weighted-quantity incentives + Strong Incentives	Difference *p*-value
		A	B	A vs B	C	D	C vs D
Number of participants	N	23	10		22	8	
No of puzzles per participant	Mean	35.78	36.30	0.90	22.09	20.88	0.70
Average creativity per participant	Mean	4.66	4.78	0.25	4.86	5.03	0.42
The number of highly creative puzzles per participant	Mean	8.65	8.70	0.97	6.95	6.63	0.79
Overall performance per participant	Mean	166.10	172.97	0.71	106.43	104.28	0.88

Panel B: Descriptive statistics for the pooled sample from Experiments 1 and 2

		Weak incentives + Quantity-only incentives	Strong incentives + Quantity-only incentives	Weak incentives + Creativity-weighted-quantity incentives	Strong incentives + Creativity-weighted-quantity incentives	Differences Weak versus Strong incentives (*p*-value)	Differences Quantity-only versus Creativity-weighted-quantity (*p*-value)
		A	B	C	D	(A+C) vs (B+D)	(A+B) vs (C+D)
Number of participants	N	19	33	16	30		
No of puzzles per participant	Mean	34.37	35.94	27.75	21.77	0.38	**<0.01**
	Min	13.00	20.00	14.00	9.00	.	.
	Max	60.00	58.00	54.00	33.00	.	.
	SDev	11.31	10.33	11.65	7.46	.	.
Average creativity per participant	Mean	4.69	4.69	4.90	4.91	0.91	<**0.01**
	Min	4.16	4.02	4.16	4.06	.	.
	Max	5.34	5.14	5.69	5.99	.	.
	SDev	0.35	0.26	0.44	0.48	.	.
The number of highly creative puzzles per participant	Mean	7.79	8.67	6.88	6.87	0.51	**<0.01**
	Min	2.00	3.00	4.00	1.00	.	.
	Max	15.00	16.00	12.00	13.00	.	.
	SDev	3.43	3.30	2.63	2.87	.	.
Overall performance per participant	Mean	159.64	168.18	133.40	105.86	0.40	**<0.01**
	Min	69.36	101.61	72.64	44.36	.	.
	Max	258.09	275.47	246.89	155.79	.	.
	SDev	49.29	47.29	49.22	34.67	.	.

This table presents descriptive statistics for Experiment 2. Panel A compares descriptive statistics for cells with a closed task and strong incentives in Experiment 2 to the comparable cells from Experiment 1 to verify whether pooling data from Experiment 2 with comparable conditions in Experiment 2 is justifiable. Panel B presents descriptive statistics for the pooled sample across the experiment conditions in Experiment 2. See Appendix A for variable definitions and descriptions on all experiment conditions. The *p*-values for differences are two-tailed. All participants included in the above analysis performed a closed task (i.e. constructed rebus puzzles for word-phrases provided by the experimenters).

Table 8: Results from ANOVA for Experiment 2

Panel A: Effect of incentive strength and type on the number of highly creative puzzles

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	41.22	4.29	**0.02**
Weak_Pay	1	4.22	0.44	0.25
Creativity_Pay Weak_Pay*	1	4.39	0.46	0.25

Panel B: Effect of incentive strength and type on average creativity

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	1.03	6.92	**<0.01**
Weak_Pay	1	0.00	0.00	0.47
Creativity_Pay Weak_Pay*	1	0.00	0.00	0.49

Panel C: Effect of incentive strength and type on overall performance

Factor	df	Sum of squares	*F*	*p*-value
Creativity_Pay	1	43,879.43	22.12	**0.00**
Weak_Pay	1	2,019.39	1.02	0.16
Creativity_Pay Weak_Pay*	1	7,281.92	3.67	**0.03**
Simple effects				
Effect of *Creativity_Pay* given *Weak_Pay* =1	1			**0.04**
Effect of *Creativity_Pay* given *Weak_Pay* =0	1			**<0.01**

This table presents ANOVA results and simple effects for creative and overall performance for Experiment 2. See Appendix A for variable definitions and descriptions on experiment conditions. All participants included in the above analysis performed a closed task (i.e. constructed rebus puzzles for word-phrases provided by the experimenters). The *p*-values are one-tailed because results from Experiment 1 allows directional predictions.

Table 9: Spread of creativity across participants

Panel A: Spread of highly creative puzzles and mediocre puzzles

Criteria per participant	Number of participants	
	Experiment 1	Experiment 2
At least 1 *minimum* rated puzzle	33	38
At least 1 *maximum* rated puzzle	37	43
At least 1 *minimum* rated puzzle and 1 *maximum* rated puzzle	26	28
At least 1 puzzle *in the bottom 25%*	44	53
At least 1 puzzle *in the top 25%*	45	53
At least 1 puzzle *in the top 25% and 1 in the bottom 25%*	44	53

Panel B: Minimum versus maximum rated puzzles

Criteria per participant	Average number of puzzles per participant	
	Experiment 1	Experiment 2
Minimum rated puzzles *given no* maximum rated puzzles	3.00	2.50
Minimum rated puzzles *given 1* maximum rated puzzle	2.71	1.74
Minimum rated puzzles *given 2* maximum rated puzzles	1.43	2.30
Minimum rated puzzles *given 3* maximum rated puzzles	2.86	2.20
Minimum rated puzzles *given 4* maximum rated puzzles	1.17	0.67
Minimum rated puzzles *given 5* maximum rated puzzles	2.50	0.00
Minimum rated puzzles *given 6* maximum rated puzzles	1.00	2.00
Minimum rated puzzles *given 7* maximum rated puzzles	N/A	1.00

This table presents the spread of highly creative and mediocre puzzles among participants for both Experiments 1 and 2. Panel A shows the number of participants that meet different criteria of creative performance. Panel B provides descriptive statistics of the extremes (minimum and maximum) in creative performance.

Chapter 4
Conclusion

4.1. Summary

Creativity has become an essential part of doing business (Amabile & Khaire, 2008; Unsworth, 2018), and recent research examines the effects of including a creativity measure as part of a multidimensional incentive contract (Kachelmeier et al., 2008; Kachelmeier et al., 2019; Kachelmeier & Williamson, 2010). This thesis set out to explore whether including incentives for creativity is always ineffective as found in the prior literature, or whether other important attributes relating to individuals, tasks and incentives will lead to different conclusions. Collectively, Studies 1 and 2 examine three such previously unexplored, yet important, attributes relating to incentivizing creativity.

Study 1 examines the effectiveness of creativity-weighted-quantity incentives for individuals with different creative capacities. Drawing on psychology and economics literatures, I predict and find, through experimental evidence, that creativity-weighted-quantity incentives lead to an improvement in creative performance for individuals with high creative capacity, but not for individuals with low creative capacity. The individuals with high creative capacity sustain highly creative performance under creativity-weighted-quantity incentives compared to quantity-only incentives. The individuals with low creative capacity under creativity-weighted-quantity incentives produce a burst of highly creative performance at the start but are not able to sustain such performance. These results suggest that creativity-

weighted-quantity incentives are effective for the types of individuals that are most likely to be employed for creative work (Hunter et al., 2012).

Study 2 examines the effectiveness of creativity-weighted-quantity incentives for closed creative tasks, at varying levels of incentive strength. Using a laboratory experiment I find that, contrary to prior theory, creativity-weighted-quantity incentives negatively affect creative and overall performance for closed tasks. A follow-up experiment shows that the negative effect on overall performance is predictably lower when incentives are weaker than when they are stronger. These results illustrate the differences in the efficacy of incentives for different types of creative tasks and levels of incentive strength.

The overarching conclusion from this thesis is that the effectiveness of including creativity as a performance dimension in a multidimensional incentive structure crucially depends on characteristics of the individual, the task and the incentives. Notably, the creative performance of individuals with high creative capacity can be increased by including incentives for creativity, while the same is not true for individuals with low creative capacity. Furthermore, including creativity incentives is detrimental for closed creative tasks without a unique solution, especially when the incentives are high compared to base pay.

4.2. Contributions

This thesis contributes to academic research, as well as informs practice, on several different fronts. Firstly, it advances the existing knowledge on using measures of creativity in multidimensional incentive systems. While the dominant prior belief was that creativity-weighted-quantity incentives have no significant effect on highly creative output, in this thesis I show that (1) for individuals with high creative capacity, creativity-weighted-quantity incentives lead to sustained higher creative performance, and (2) for creative tasks with pre-specified problems (i.e. closed tasks) creativity-weighted-quantity incentives are detrimental

to creative performance. These findings add to our current understanding on the efficacy of creativity incentives.

Secondly, it contributes to the large literature on creativity in management and psychology disciplines by (1) providing insights on the interactions between individual and contextual factors that influence creative performance, and (2) unpacking characteristics of the umbrella term "creativity". I provide insight on the interactions of creativity incentives with individual characteristics (Study 1), and task and incentive characteristics (Study 2). Shalley and Gilson (2004) identifies the need for research to consider such interactions. By specifically studying two different types of creative tasks (Study 2), I also directly respond to the recent calls in the creativity literature to distinguish between different task types historically pooled under creativity (Gilson & Madjar, 2011; Malik & Butt, 2017; Unsworth & Luksyte, 2015).

Thirdly, my thesis also contributes to the more general incentive design literature on incorporating soft measures (Studies 1 and 2), and literatures on individual traits (Study 1) and stimulation strength (Study 2).

It contributes to practice by informing the incentive design process for employees performing creative work. I highlight the importance of tailoring incentives to match employees' creative capacities and the characteristics of the task itself.

4.3. Future directions

While my thesis provides a more nuanced understanding of attributes relating to incentivizing creativity, much is yet to be studied on how organizations can encourage efficient creativity. There is a dearth of research on interactions between incentives and individual/contextual characteristics (Shalley et al., 2004). Although my thesis is limited to monetary incentives, certain contextual characteristics may even render monetary incentives undesirable. For instance, if the initial level of intrinsic interest in a task is moderately high, but not high enough

to positively incorporate extrinsic incentives, monetary incentives can be harmful to performance (Amabile, 1988). Thus, it is important to examine other forms of rewards (e.g. non-monetary, non-financial) and management control mechanisms (e.g. selection, controls, feedback), that will also interact with each other and with contextual characteristics. While some research is beginning to look at these relationships (e.g. Cools et al., 2017; Liu et al., 2019), more input from management accounting researchers will advance our understanding on how to design comprehensive management control systems around creative work.

Limitations in the participant pool and practicalities prevented me from studying all types of incentive options (e.g. fixed, quantity-only, creativity-only, and creativity-weighted-quantity), for all task types based on problem specification (e.g. open, closed with unique solution, closed with multiple possible solutions), at different levels incentive strengths (strong, moderate, weak). A large experimental design approach (Ochs & Roth, 1989) would be ideal for future research on this area to synthesize the findings.

Furthermore, my findings in Study 2 only partially explain the mixed results on the effectiveness of creativity incentives for closed tasks. Given the prevalence of closed creative tasks in practice (Unsworth, 2001), there is opportunity for future studies to explore other critical characteristics that explain the difference.

The means to manage mediocre ideas generated in the creativity process is also an interesting direction that is outside the scope of this thesis. If an organization desires all the implemented ideas to be highly creative, mediocre ideas from employees will have to be screened out or prevented. Examination of cost-benefit trade-offs between prevention and screening is an opportunity for future research. Additionally, for closed tasks that are judged to be mediocre, alternative means to solicit more creative solutions are required. Potential

methods such as rotating tasks among individuals, or assignment of tasks to multiple individuals (e.g. tournaments) are avenues for future research.

Lastly, experiments conducted in a controlled laboratory setting with university student participants can raise concerns about external validity. Although my experiments are intentionally stripped-down versions of the real world, it is possible that other complex business environmental characteristics could lead to more differences within the findings. More field work on important attributes relating to creativity incentives (e.g. Cools et al., 2017) would be welcome complements to this thesis.

In sum, my thesis shows the importance of having a more granular understanding of incentive design in the creativity sphere. This opens the door for further work on performance management and rewards for enhancing creative performance.

References

Abernethy, M. A., & Brownell, P. (1997). Management control systems in research and development organizations: The role of accounting, behavior and personnel controls. *Accounting, Organizations and Society, 22*(3/4), 233-248.

Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews, 8*(1), 21-47.

Alphabet Inc. (2019). *Annual Report 2018*. Retrieved from https://abc.xyz/investor/

Amabile, T. M. (1982). Children's artistic creativity: Detrimental effects of competition in a field setting. *Personality and Social Psychology Bulletin, 8*(3), 573-578.

Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology, 48*(2), 393-399.

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior, 10*(1), 123-167.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal, 39*(5), 1154-1184.

Amabile, T. M., & Khaire, M. (2008). Your organization could use a bigger dose of creativity. *Harvard Business Review, 86*(10), 101-109.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies, 76*(2), 451-469.

Baer, M., Oldham, G. R., & Cummings, A. (2003). Rewarding creativity: When does it really matter? *The Leadership Quarterly, 14*(4-5), 569-586.

Bailyn, L. (1985). Autonomy in the industrial R&D lab. *Human Resource Management, 24*(2), 129-146.

Balkin, D. B., Markman, G. D., & Gomez-Mejia, L. R. (2000). Is CEO pay in high-technology firms related to innovation? *Academy of management journal, 43*(6), 1118-1129.

Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology, 32*(1), 439-476.

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of personality and social psychology, 46*(3), 610.

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining competing theoretical models. *Journal of Applied Psychology, 95*(1), 201-212.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*(1), 40-52.

Chang, L., & Birkett, B. (2004). Managing intellectual capital in a professional service firm: Exploring the creativity–productivity paradox. *Management Accounting Research, 15*(1), 7-31.

Charness, G., & Grieco, D. (2018). Creativity and incentives. *Journal of the European Economic Association, 17*(2), 454-496.

Cools, M., Stouthuysen, K., & Van den Abbeele, A. G. (2017). Management control for stimulating different types of creativity: The role of budgets. *Journal of Management Accounting Research, 29*(3), 1-21.

Costa Jr, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 2 Personality measurement and testing, pp. 179-198). Thousand Oaks: Sage Publications, Inc.

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

Deci, E. L. (1975). *Intrinsic motivation. [electronic resource]*: Springer US.

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

Diekmann, J., & König, C. J. (2015). Personality testing in personnel selection: Love it? Leave it? Understand it! In I. Nikolaou & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 129-147): Psychology Press.

Earley, P. C. (1985). Influence of information, choice and task complexity upon goal acceptance, performance, and personal goals. *Journal of Applied Psychology, 70*(3), 481-491.

Eisenberger, R., & Armeli, S. (1997). Can salient reward increase creative performance without reducing intrinsic creative interest? *Journal of Personality and Social Psychology, 72*(3), 652-663.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2*(4), 290-309.

Feltham, G. A., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review*, 429-453.

Ferreira, A., & Otley, D. (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research, 20*(4), 263-282.

Gilson, L. L., Lim, H. S., D'Innocenzo, L., & Moye, N. (2012). One size does not fit all: Managing radical and incremental creativity. *The Journal of Creative Behavior, 46*(3), 168-191.

Gilson, L. L., & Madjar, N. (2011). Radical and incremental creativity: Antecedents and processes. *Psychology of Aesthetics, Creativity, and the Arts, 5*(1), 21-28.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica, 74*(5), 1365-1384.

Gong, Y., Huang, J.-C., & Farh, J.-L. (2009). Employee learning orientation, transformational leadership, and employee creativity: The mediating role of employee creative self-efficacy. *Academy of Management Journal, 52*(4), 765-778.

Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology, 37*(8), 1398-1405.

Grabner, I. (2014). Incentive system design in creativity-dependent firms. *The Accounting Review, 89*(5), 1729-1750.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization, 7*, 24-52.

Hunter, S. T., Cushenbery, L., & Friedrich, T. (2012). Hiring an innovative workforce: A necessary yet uniquely challenging endeavor. *Human Resource Management Review, 22*(4), 303-322.

Ittner, C. D., Larcker, D. F., & Rajan, M. V. (1997). The choice of performance measures in annual bonus contracts. *The Accounting Review*, 231-255.

Kachelmeier, S. J., Reichert, B. E., & Williamson, M. G. (2008). Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research, 46*(2), 341-373.

Kachelmeier, S. J., Wang, L. W., & Williamson, M. G. (2019). Incentivizing the creative process: from initial quantity to eventual creativity. *The Accounting Review, 94*(2), 249-266.

Kachelmeier, S. J., & Williamson, M. G. (2010). Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review, 85*(5), 1669-1691.

Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*(5), 622-629.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4 ed.): New York, NY: Guilford Press.

Liu, S., Ditillo, A., & Van den Abbeele, A. G. (2019). *When should we reward creativity: A new perspective on creativity types*. Retrieved from https://ssrn.com/abstract=3233126

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*(2), 157-189.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*(1), 125-152.

Madjar, N., Greenberg, E., & Chen, Z. (2011). Factors for radical creativity, incremental creativity, and routine, noncreative performance. *Journal of Applied Psychology, 96*(4), 730-743.

Malik, M. A. R., & Butt, A. N. (2017). Rewards and creativity: Past, present, and future. *Applied Psychology*, *66*(2), 290-325.

Malik, M. A. R., Butt, A. N., & Choi, J. N. (2015). Rewards and employee creative performance: Moderating effects of creative self-efficacy, reward importance, and locus of control. *Journal of Organizational Behavior, 36*(1), 59-74.

McGraw, K. O., & McCullers, J. C. (1979). Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. *Journal of Experimental Social Psychology*, *15*(3), 285-294.

McKinsey & Company. (2012). *Making innovation structures work: McKinsey Global Survey results*. Retrieved from http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/making-innovation-structures-work-mckinsey-global-survey-results

Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *The American Economic Review, 79*(3), 355-384.

Perrow, C. (1970). *Organizational analysis: A sociological view*. London: Tavistock Publications.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of consumer research, 21*(2), 381-391.

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

Richter, A. W., Hirst, G., Van Knippenberg, D., & Baer, M. (2012). Creative self-efficacy and individual creativity in team contexts: Cross-level interactions with team informational resources. *Journal of Applied Psychology*, *97*(6), 1282-1290.

Robbins, S. P., & Coulter, M. K. (2012). *Management* (11 ed.): Boston: Prentice Hall, c2012.

Roberts, J. (2010). Designing incentives in organizations. *Journal of Institutional Economics, 6*(01), 125-132.

Rockness, H. O., & Shields, M. D. (1984). Organizational control systems in research and development. *Accounting, Organizations and Society, 9*(2), 165-177.

Shalley, C. E., & Gilson, L. L. (2004). What leaders need to know: A review of social and contextual factors that can foster or hinder creativity. *The Leadership Quarterly, 15*(1), 33-53.

Shalley, C. E., Zhou, J., & Oldham, G. R. (2004). The effects of personal and contextual characteristics on creativity: Where should we go from here? *Journal of Management, 30*(6), 933-958.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*(4), 525-547.

Terwiesch, C., & Ulrich, K. (2011). Innovation tournaments: More and better ideas. *Wharton@Work*. Retrieved from https://executiveeducation.wharton.upenn.edu/thought-leadership/wharton-at-work/2011/03/innovation-tournaments

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*(6), 1137-1148.

Tierney, P., & Farmer, S. M. (2004). The Pygmalion process and employee creativity. *Journal of Management, 30*(3), 413-432.

Toyota Motor Corporation. (2019). Toyota way. Retrieved from http://www.toyota.com/usa/our-story/index.html#!/toyota-way\

Unsworth, K. (2001). Unpacking creativity. *Academy of Management Review, 26*(2), 289-297.

Unsworth, K. (2018). Creativity in the context of multiple goals. In *The Palgrave handbook of creativity at work* (pp. 23-41): Springer.

Unsworth, K., & Luksyte, A. (2015). Is all creativity created equal?: Exploring differences in the creativity processes across the creativity types. In C. Shalley, M. A. Hitt, & J. Zhou (Eds.), *The Oxford handbook of creativity, innovation, and entrepreneurship*. New York: Oxford University Press.

Vaughn, L. A. (1999). Effects of uncertainty on use of the availability of heuristic for self-efficacy judgments. *European Journal of Social Psychology, 29*(2-3), 407-410.

Vernon, P. E. (1989). The nature-nurture problem in creativity. In *Handbook of creativity* (pp. 93-110): Springer.

Walmart Inc. (2018). Values and behaviors. Retrieved from https://corporate.walmart.com/our-story/working-at-walmart

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482.