



# CGS WORKING PAPER

## Quantity, Quality, and Originality: The Effects of Incentives on Creativity

---

Katharina Laske (University of Cologne)

Marina Schröder (University of Cologne)

Cologne Graduate School  
in Management, Economics  
and Social Sciences  
Albertus-Magnus-Platz  
50923 Köln  
[www.cgs.uni-koeln.de](http://www.cgs.uni-koeln.de)

University of Cologne



# Quantity, Quality, and Originality: The Effects of Incentives on Creativity\*

By KATHARINA LASKE<sup>†</sup> AND MARINA SCHRÖDER<sup>‡</sup>

*We introduce a novel experimental design in which creativity is incentivized and measured along three dimensions: quantity, quality and originality of ideas. We implement piece rate incentives for quantity alone, quantity in combination with quality and quantity in combination with originality and compare the results to a baseline with a fixed wage. We find that incentives significantly affect the quantity and average quality of ideas, but not the average originality. Incentives for both quantity and originality perform best in fostering innovative ideas.*

*Keywords: creativity, multitasking, laboratory experiment, real-effort, incentives,*

*JEL-Classification; C91; J33; M52; O30*

---

<sup>†</sup> University of Cologne, Seminar of Personnel Economics and Human Resource Management, Albertus-Magnus-Platz, 50923 Cologne, Germany, phone: +49 221 470 7955, Email: laske@wiso.uni-koeln.de.

<sup>‡</sup> University of Cologne, Seminar of Personnel Economics and Human Resource Management, Albertus-Magnus-Platz, 50923 Cologne, Germany, phone: +49 221 470 6939, Email: marina.schroeder@uni-koeln.de.

\* We thank Michèle Belot, Uri Gneezy, Patrick Kampkötter, Philipp Kircher, Karim Sadrieh, and Dirk Sliwka for helpful comments and support in conducting this research. We thank participants of the PhD course Behavior and Incentives, the Second Incentives and Behavior Change Workshop, the Morality, Incentives and Unethical Behavior Conference, the Annual Meeting of the German Association for Experimental Economics, the 18<sup>th</sup> Colloquium on Personnel Economics, and research seminar participants in, Berlin, Cologne, Munich, San Diego, and Regensburg for helpful comments and suggestions. Special thanks goes to the student assistants for their help in programming and conducting the experiment. Financial support of the University of Cologne Female Research Grant and the Deutsche Forschungsgemeinschaft [DFG FOR 1371] is gratefully acknowledged.

## I. Introduction

Creative thinking is crucial for any innovation, and as such, it is highly relevant for progress and economic success. Surveys show that fostering creativity and innovation has been one of global top managers' primary concerns in recent years (The Conference Board, 2012, 2013, 2014). While creativity is studied extensively in psychology (see e.g., Amabile 1996 or Kaufman and Sternberg 2010 for overviews of the psychological research), there is only little research into how economic factors, such as monetary incentives, affect creative behavior.<sup>1</sup>

One reason for the scarcity of research in this field is the difficulty of defining and quantifying creativity. Typically, creativity is context-specific and the evaluation of creative performance is subjective. Additionally, creativity is not a unitary construct but a multidimensional phenomenon. Common definitions of creativity in the literature refer to quantity, quality and originality as important dimensions (see e.g., Simon, 1983, Amabile, 1996, Kaufman and Sternberg, 2010), where quantity relates to the number of ideas, quality to the degree to which an idea is suitable for its intended purpose, and originality to the degree of innovativeness of an idea.

In this paper, we introduce a novel experimental design that clearly specifies the creative context and provides objective measures for these three dimensions of creativity. The design enables us to separately assess the effect of piece rate incentives on performance for each of the three dimensions of creativity and thus to study the channel through which incentives affect creative performance. However, to result in successful innovation creative ideas have to be at the same time of high quality and original. We therefore additionally introduce a combined

<sup>1</sup> see Becker et al. 2012 for an overview of the research on the impact of incentives on creativity.

measure of innovation that captures the number of creative ideas that are both of high quality and original. Our research contributes to the debate among academics and practitioners about whether and *how* incentives improve or undermine creativity and innovation.

In the experimental task, participants are asked to illustrate words with the help of a given set of materials. They are directed to create as many illustrations as possible (quantity) that can be recognized by independent raters (quality), and that are statistically infrequent (originality). Participants do not receive a list or any specifications on which words to illustrate or how to use the provided materials. Hence, they have to come up with both the words they want to illustrate and a way of illustrating these words. The design offers the advantage that it allows us to differentiate creativity along its dimensions quantity, quality and originality, with clearly defined objective measures where performance in each of these three dimensions requires creative thinking. In the experiment we implement piece rate incentives for quantity alone, quantity & quality and quantity & originality, and measure the effect of these incentives on all three dimensions of creativity. If creative performance can be influenced by effort, economic theory predicts that introducing piece rate incentives will increase overall effort provision (incentive effect). Additionally, in a multidimensional task such as creativity, economic theory predicts that incentives will result in an increase in effort provision for incentivized dimensions and a decrease in effort provision for non-incentivized dimensions (spillover effect) whenever tasks are substitutes in workers' effort costs (Holmstrom and Milgrom, 1991). Therefore, if creativity is effort-dependent and can be influenced by monetary incentives, we expect to observe positive direct performance effects in the incentivized dimension and negative spillover effects on performance in the non-incentivized dimensions of creativity. However, creative performance may not be determined (only) by effort but (also) by an individual's talent (see e.g. Amabile et. al, 1986). In the most extreme case, it is possible that

effort has no effect on creative performance. A combination of the two forces – effort and talent – is also conceivable: Performance in some dimensions may be rather related to effort, while performance in other dimensions may be independent of effort and merely related to talent. As a result of such differences, the effect of incentives on creativity may depend crucially on the dimension of creativity that is incentivized. Incentives for some dimensions of creativity may have a positive effect on performance, while incentives for other dimensions may not. Since our design allows us to separately assess the effect of incentives on performance for each of the three creativity dimensions it is suitable to measuring differences in the impact of incentives on creative performance depending on the dimension of creativity.

We find that performance in the quantity and quality dimensions is significantly affected by effort. The effects of incentives on quantity and quality are in line with predictions of economic theory (Holmstrom and Milgrom, 1991, Lazear, 2000). Incentivizing either *quantity* or both *quantity & originality* leads to an increase in quantity and a decrease in the average quality of creative ideas compared to a baseline treatment with a fixed wage. When both *quantity & quality* are incentivized, performance in both dimensions increases slightly but insignificantly. We find that average originality is not significantly affected by effort provision. The likelihood that a certain idea is original is not influenced by monetary incentives but instead it seems to be ascribed to an individual's talent. However, the number of original ideas increases with the overall number of ideas. Thus, incentives do not have an impact on the average originality of ideas but they do affect the number of original ideas.

To study the effect of incentives on innovation, we introduce a combined measure for innovation that captures the number of creative ideas that are at the same time of high quality and original. We find that incentives for quantity in combination

with originality result in the highest degree of innovation. We conjecture that this is due to higher experimentation whenever originality is incentivized.

Closely related to our research, there are some recent lab studies (Eckartz et al, 2012, Mohnen and Ostermaier, 2013, Charness and Grieco, 2014, Erat and Gneezy, 2015, Bradler et al., 2016) and field experiments (Azoulay et al., 2011, Gross, 2014, Neckermann et al, 2014) in economics that study the effect of incentives on creativity. These previous studies mainly focus on the performance effect of different incentive schemes. However, they do not investigate spillover effects of incentives that go beyond a direct effect of the incentivized dimensions. The only study we are aware of that explicitly looks at spillover effects when incentivizing different dimensions of creative performances is the study by Kachelmeier et al. (2008), which measures creativity as a two-dimensional construct consisting of quantity and a subjective measure for creativeness. Incentivizing these dimensions separately, they provide evidence for direct and spillover effects of monetary incentives on performance. We add to the research by introducing an experimental design that allows objectively measuring quantity, quality, and originality. We show that this distinction is crucial in order to better understand the direct and spillover effects of incentives on creative behavior. Additionally, since firms look for employees that can further organizational innovation, we expand the view and examine the effects of incentives on a combined measure of innovation. The importance of our finding lies in showing that the effect of incentives on creative performance varies with the dimensions of creativity and that they entail negative spillover effects on non-incentivized dimensions of creativity. Using a combined measure for innovation, we show that the effect of incentives goes beyond the effects on the separate dimensions of creativity. It seems that incentives also have an effect on the type of creative ideas that are generated. One explanation for this finding is that individuals explore more when facing certain incentive schemes

compared to others. Our research is an important step towards understanding the effect of incentives on creative performance.

## **II. Experimental Design and Procedure**

### *A. The Task*

We propose a novel experimental design that allows to objectively assess performance in three dimensions of creativity, i.e. quantity, quality and originality. In this experimental design, we ask participants to illustrate words using several simple materials. The set of materials provided for each participant consists of one string, two O-rings, four wooden sticks, and twelve colored glass pebbles (see Figure 1 left picture). Participants can use some or all of these materials to illustrate words (see Figure 2 for example illustrations). Participants do not receive a list or any specifications on which words to illustrate or how to use the provided materials; hence, in this task, they have to come up with both the words they want to illustrate and a way of illustrating these words. They can illustrate as many words as they want within a period of 20 minutes. For each illustration, participants were instructed to take a picture using a pre-installed camera and to type in the illustrated word<sup>2</sup>. The advantage of the task is that it allows us to objectively measure multiple dimensions of participants' creative performance, i.e., quantity (number of valid illustrations), quality (recognition rate among raters), and originality (statistical infrequency of the illustrated word among all words illustrated by the participants in the experiment).

<sup>2</sup> A detailed description of how lab participants took a picture of an illustration is given in section 2.3 and in the instructions in Appendix A.



FIGURE 1. SET OF MATERIALS AND EXPERIMENTAL SETUP

We measure quantity as the number of different words illustrated. That is, a participant scores high (low) on this dimension if she illustrates a high (low) number of different words. We directed participants to only illustrate single words (e.g., “tree”, “face”), to illustrate each word only once, and that they are not allowed to use or illustrate any symbol found on the keyboard (e.g., “→”, “8”, “b”, “@”, “>” “+”). Illustrations of phrases consisting of more than one word (e.g., “tree in the woods”, “happy face”), multiple illustrations of the same word (e.g., two different illustrations of the word “house”), or illustrations including symbols from the keyboard were not valid (e.g., using “8” to illustrate the word “eight”). We clearly instructed participants about this rule and informed them that illustrations violating this rule would not be considered for payment. See Appendix A for the instructions.

To measure the quality of each valid illustration, we elicit the recognition rate by external raters through an online survey, which was conducted two weeks after the lab experiment. In this online survey, raters are provided with pictures of the illustrations from the lab experiment and have to type in the exact word that is illustrated.<sup>3</sup> We incentivize answers in this online survey such that the online raters

<sup>3</sup> In the assessment of quality, we did not account for synonyms since we explicitly informed participants of both the lab experiment and the online survey in the instructions that only the exact match of the illustrated word by the lab participant

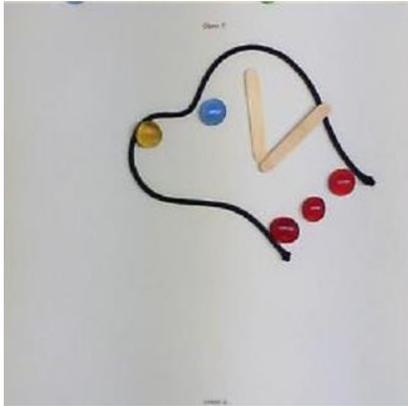
receive €0.10 for each correct answer, which is defined as an exact match of a word illustrated by a lab participant and the answer by the online participant. See Appendix B1 for the instructions of the online survey and Appendix B2 for a screenshot of the online survey. Raters in the online survey did not take part in any previous related experiments and were blind to treatments. At least ten online raters rated each illustration, and each rater in the online survey rated a random sample of 50 illustrations. For each illustration, we derive quality as the fraction of raters who correctly identified the illustrated word. For example, if 10 out of 10 raters recognize an illustrated word, it would receive the highest quality score of 1. An illustration that is only identified by 1 out of 10 raters receives a quality score of 0.1. See Figure 2 (left column) for examples of high- and low-quality illustrations.

We measure originality as the statistical infrequency of an illustrated word within the entire experiment. Specifically, we derive the originality of an illustration as the ratio of 1 and the number of times the same word was illustrated in the sample. For example, a word that is illustrated once in the whole experiment, such as “tennis,” receives the highest originality score of 1. A word that is illustrated many times, such as “house,” which was illustrated 82 times, receives a low score of 0.012. Note that in this measure the marginal effect of one additional illustration of the same word is greater for words that are rarely illustrated compared to those that appear very frequently. See Figure 1 (right column) for examples of illustrations that scored high or low on originality. In addition, we calculate uniqueness within a randomly assigned group of four. An illustration of a word is unique if no other participant in the group illustrated the same word.<sup>4</sup>

and the answer by the online rater will be considered for payment. Spelling errors were not corrected. The special characters ä, ö, ü and ß were standardized to a, o, u and ss, respectively. Capitalization of letters was not taken into account.

<sup>4</sup> While originality is the more precise measure, we used uniqueness to incentivize participants in the incentivize *quantity* & *originality* treatment. We used uniqueness to simplify the procedure of conducting this experiment. Since participants cannot anticipate the words illustrated by group members, this design element should not have an effect on subject’s strategic considerations.

HIGH QUALITY ILLUSTRATION



illustrated word: dog

quality: 1

originality: 0.17

HIGH ORIGINALITY ILLUSTRATION



illustrated word: tennis

quality: 0.9

originality: 1

LOW QUALITY ILLUSTRATION



Illustrated word: pig

Quality: 0

Originality: 0.33

LOW ORIGINALITY ILLUSTRATION



Illustrated word: house

Quality: 1

Originality: 0.01

FIGURE 2. EXAMPLES OF ILLUSTRATIONS

## B. Treatments

In three experimental treatments and a baseline treatment, we vary the incentivized dimensions of creativity and estimate the direct and spillover effects of these incentives on the incentivized and non-incentivized dimensions of creative performance. Table 1 summarizes the treatments in our experiment. Overall, 126 participants took part in the lab experiment, 31 each in the *incentivize quantity* and the *incentivize quantity & quality* treatments and 32 each in the *baseline* and the *incentivize quantity & originality* treatments.<sup>5</sup>

Except the section explaining the monetary incentives, all participants receive the same instructions, in which they are asked to illustrate as many words as possible that can be recognized by external raters and are unique. Thus, all participants are informed about the relevant dimensions of creativity and receive information on the scoring procedure for measuring each of the dimensions.

In the *baseline* treatment, all participants receive a €10 fixed payment, independent of their performance. After conducting the *baseline* treatment, we calibrated the size of the piece rate incentives for the three treatments according to the performance in the *baseline* experiment. That is, we ensured that if participants in the three piece rate treatments perform at the same level for each of the creativity dimensions as participants in the *baseline* treatment, the expected payoff is equal to that in the *baseline* treatment.

In the *incentivize quantity* treatment, participants are paid according to the number of words illustrated. For each illustration, they receive €0.60. In this treatment, a participant's payment only depends on the quantity of illustrations, while quality and originality are not relevant for the payoff. Thus, if effort is costly,

<sup>5</sup> For our analysis, we exclude one observation from the *incentivize quantity* treatment, because this participant only generated invalid illustrations.

payoff-maximizing participants should provide effort only to generate quantity but not to generate quality or originality.

In the *incentivize quantity & quality* treatment, each illustration is rated by at least 10 raters who are incentivized to recognize the word that is illustrated. For each illustration, participants receive €0.10 per rater who correctly identifies the illustrated word.<sup>6</sup> In this treatment, payment depends on both the quantity and the quality of illustrations. Payoff-maximizing participants should choose an optimal division of effort between these two dimensions, while originality is not payoff-relevant. Previous experimental studies on non-creative real effort tasks reveal a tradeoff between quantity and quality, which is in line with the assumption of substitutable effort costs for quantity and quality (see e.g., Hong et al., 2013, Al-Ubaydli et al., 2015). In our design, creating high quality illustrations is likely to be more time-consuming, which limits the number of illustrations a participant can create in the given time. As a result of this tradeoff, the direct effects of incentives in each of the two dimensions may be lower compared to situations where only single dimensions of creativity, either quantity or quality, are incentivized.

Finally, in the *incentivize quantity & originality* treatment, participants are paid according to the number of illustrations that are unique within a group of four participants. For each illustration of a word that no other participant in a randomly assigned group of four participants has illustrated, participants receive €0.85. Payment in this treatment depends on the number of original illustrations. If quantity and originality depend on effort provision (i.e. time allocated), participants in this treatment will again face a tradeoff between generating a high number of illustrations, where for each illustration the probability that it is unique is low, and illustrating a low number of illustrations, where for each illustration the probability

<sup>6</sup> Note that some illustrations were rated by more than ten raters. In these cases we only consider the first ten raters to derive the payoff of the participants in the *incentivize quality* treatment. We consider the average recognition rate for our data analysis.

that an illustration is unique is high. However, some previous studies do not find evidence for such a tradeoff between quantity and originality (Diehl and Stroebe 1987, Dennis et al. 1996). It may be that the originality of a certain idea is independent of effort provision, but solely depends on talent, luck, or on the overall number of ideas generated. For example, it is conceivable that the likelihood of creating original illustrations increases with the overall number of illustrations or over time. If originality is independent of effort provided for originality itself, participants in the *incentivize quantity & originality* treatment cannot influence the originality of each single illustration through increasing effort provided for originality but can only affect their payout through effort provided to generate quantity. In this case, the effect of incentives for both *quantity & originality* should be similar but weaker (due to the introduction of uncertainty) compared to the effect of incentives for *quantity* only.

TABLE 1 - TREATMENTS

Treatment	Payment	Amount	N
baseline	fixed payment	€10	32
incentivize quantity	number of illustrations	€0.60 per illustration	31
incentivize quantity & quality	number of raters who correctly identify each illustration	€0.10 per correct identification of each illustration per rater	31
incentivize quantity & originality	number of unique illustrations (in a group of four)	€0.85 per unique illustration	32

### *C. Procedural Details*

The experiment was conducted at the Cologne Laboratory for Economic Research at the University of Cologne. Participants were recruited with the online recruiting system ORSEE (Greiner 2004). We ran eight sessions in May 2014, with two sessions for each treatment condition. Participants were randomly seated in separated cubicles in the lab. To inform participants about the task, they received written instructions, which were read aloud by the experimenter. After the experimenter had answered all questions individually, the set of materials was handed to the participants. All illustrations of words had to be placed within a designated area on the desks. We told participants to place all materials that were not used for the illustration outside this area. Additionally, participants were instructed not to use any materials other than those provided by the experimenter. Once a participant made an illustration, she pressed a button on the screen of the computer so that the software would automatically take a picture of the designated area including the illustration. If participants were satisfied with the picture, they had to type in the word that they had illustrated and could then proceed with their next illustration. If not satisfied they could take another picture before proceeding. Figure 1 (right side) illustrates the cubicle in the laboratory, including the designated area in which participants provided illustrations and the web cam taking the pictures. As soon as the working time of 20 minutes was over, the experimental software automatically stopped and then initiated a questionnaire. In this questionnaire, we asked some general demographic questions. Additionally, we elicited how much the participants enjoyed performing the task and how well they

think they performed in this task. We will later use this measure of enjoyment as a proxy for intrinsic motivation.<sup>7</sup>

On average, each session lasted 40 minutes, and the average payoff was €14.43. The final payoff for each participant consisted of the money earned during the experiment and a standard show-up fee of €2.50. The money was paid out two weeks after the experiment, and participants could choose whether they wanted to collect the money in cash at the university or whether they preferred to have it transferred directly to their bank account.

For the online survey, we recruited 540 participants from the same subject pool via ORSEE (Greiner 2004) and excluded participants who had previously taken part in the lab experiment. The online survey lasted about 20 minutes, and the average earnings were €4.50, including a €2.00 show-up fee. As in the lab experiment, participants had a choice between collecting the money in cash and a bank transfer.

### **III. Results**

#### *A. Separate Dimensions of Creativity*

Table 2 provides summary statistics on performance in each treatment separately and overall. On average, each participant in this experiment created 21 valid illustrations, resulting in 2,648 illustrations that are considered for the analysis. We find that participants in the *incentivize quantity* and the *incentivize quantity & originality* treatments create significantly more illustrations than those in the *baseline* or the *incentivize quantity & quality* treatments (pairwise U-test,  $p < 0.01$ ,

<sup>7</sup> Note that this measure of intrinsic motivation is only suitable to derive immediate effect of incentive schemes on intrinsic motivation but not on the effect of monetary incentives per se or on the long-term effects of incentives on creativity.

two-tailed). We do not observe significant differences in the number of illustrations between the other treatments (pairwise U-test,  $p > 0.20$ , two-tailed).

TABLE 2: SUMMARY STATISTICS

	<i>baseline</i>	<i>incentivize quantity</i>	<i>incentivize quantity &amp; quality</i>	<i>incentivize quantity &amp; originality</i>	<i>overall</i>
Avg. no. of illustrations per subject	16.84 (8.35)	25.61 (9.79)	18.20 (5.31)	24.66 (9.70)	21.34 (9.26)
Avg. quality of illustrations	0.50 (0.18)	0.46 (0.15)	0.55 (0.13)	0.49 (0.14)	0.50 (0.15)
Avg. originality of illustrations	0.27 (0.21)	0.26 (0.16)	0.20 (0.11)	0.25 (0.11)	0.24 (0.15)
Intrinsic motivation	3.47 (1.11)	3.25 (1.26)	3.50 (1.20)	3.75 (1.19)	3.50 (1.19)
Earnings per subject (excl. show-up fee)	10 (0.00)	15.37 (5.87)	10.44 (3.13)	12.35 (6.44)	12.03 (5.05)

*Notes:* Standard deviations in parentheses. (We report standard deviations between participants, i.e., how much participants quantity, mean quality and mean originality of their illustrations, as the non-parametric analysis is also based on a comparison values per participant.)

The average quality of illustrations is 0.50, indicating that, on average, half of the raters correctly identified an illustrated word. Average quality is highest in the *incentivize quantity & quality* treatment. The difference in quality is significant when comparing the *incentivize quantity & quality* treatment and the *incentivize quantity* treatment (U-test,  $p = 0.01$ , two-tailed) and insignificant for all other pairwise comparisons (pairwise U-test,  $p > 0.15$ , two-tailed).

As discussed above, we measure originality as the ratio of 1 and the number of participants who illustrated the same word in the entire experiment. Thus, the maximum originality is one (no other participant illustrated the same word), and the minimum in our dataset is equal to 0.012. This minimum value is appointed to illustrations of the word ‘house,’ which was illustrated by 66 percent of the participants in our experiment. We observe a slight tendency that illustrations are less original in the *incentivize quantity & quality* treatment compared to the other treatments. This difference is not significant compared to the *baseline*, but it is

compared to the *incentivize quantity* treatment (U-test,  $p=0.05$ , two-tailed) and to the *incentivize quantity & originality* treatment (U-test,  $p=0.09$ , two-tailed). Thus, *incentivizing quantity & quality* has a significant negative effect on originality even though this dimension of creativity is not targeted by the incentive scheme. One reason for this negative effect is the negative correlation between quality and originality (Spearman,  $\rho=-0.7$ ,  $p=0.00$ ). To be better able to analyze interaction and spillover effects between the different dimensions of creativity, we conduct regression analyses controlling for the different dimensions of creativity.

Table 3 provides regression results separately for the three dimensions of creativity. Starting with quantity, col (1 to 3) show the results of OLS regressions on the overall number of valid illustrations per participant. In the model in col (1), we control for treatment dummies (reference group is the baseline treatment). As in the non-parametric analysis above, this model reveals that participants in the *incentivize quantity* and the *incentivize quantity & originality* treatments create significantly more illustrations compared to the *baseline*. We find no significant effect of gender on the number of illustrations. As discussed above, changes in performance in one dimension could be due to an incentive or a spillover effect. To control for such spillovers, we add controls for productivity in the other dimensions of creativity in col (2) (i.e., controls for the average quality and the average originality of illustrations). The coefficients of the *incentivize quantity* treatment and of the *incentivize quantity & originality* treatment remain large and significant even when adding these controls. However, the coefficient in the *incentivize quantity* treatment is significantly lower ( $p=0.056$ ) compared to the model in col (1). Thus, we find evidence for incentive and spillover effects of piece rate incentives on the quantity of illustrations. Further, we find a significant negative relationship between quantity and average quality, while the coefficient for average originality is not significant. As a proxy for intrinsic motivation, we asked participants in the post-experimental survey to indicate how much they enjoyed

working on the task on a scale from 1 ‘not at all’ to 5 ‘very much’.<sup>8</sup> In the model presented in col (3), we control for intrinsic motivation. We observe that the number of illustrations increases significantly with this measure for intrinsic motivation, while adding this control does not substantially change the other coefficients and importantly does not seem to have a significant effect on the treatment differences that we find.

Col (4 to 6) display the results of a random effects regression model, with quality of an illustration as the dependent variable. In these models related to quality and the following models related to originality (col 7 to 9), we focus on results on an illustration level. Col (4) reveals a significant decrease in average quality of illustrations when *quantity* is incentivized compared to the *baseline*. Additionally, we observe a slight but insignificant increase in average quality when *quantity & quality* are incentivized and a slight but insignificant decrease in average quality when *quantity & originality* are incentivized. We find that illustrations by female participants are of significantly higher quality compared to those of male participants. In col (5), we additionally control for performance in the other dimensions of creativity to test for the relevance of spillover effects. Adding these controls, we no longer observe a significant effect on the quality in the *incentivize quantity* treatment and also the coefficient for the *incentivize quantity & originality* treatment is close to zero. Our results suggest that the observed negative treatment effects on average quality are due to spillover effects. In col (6), we additionally control for intrinsic motivation, the work minute in which an illustration was created and the number of characters of the word that is illustrated. We find that intrinsic motivation has a positive effect on the quality of illustrations.

<sup>8</sup> We find that the treatments do not have a major impact on this measure of intrinsic motivation. Participants in the *incentivize quantity & originality* treatment seemed to have liked the task slightly more than participants in the *incentivize quantity* treatment (U-test,  $p=0.09$ , two-tailed). All other treatment comparisons are not significant (pairwise U-test,  $p>0.20$ , two-tailed).

Additionally, we find that the quality of illustrations decreases over time and that illustrations of longer words are on average of significantly lower quality compared to illustrations of shorter words.<sup>9</sup> Again, adding these controls does not have a significant impact on our treatment effects. Col (7 to 9) display the results of a random effects regression model, with originality of an illustration as the dependent variable. In this analysis, we focus on the likelihood that each single illustration is original. In the model presented in col (7), we control for treatments and gender of the participants. Compared to the baseline, we find no significant effects of piece rate incentives on average originality. We observe that female participants perform less well in creating original illustrations. In col (8), we introduce controls for quantity and average quality of illustrations. As discussed above, there is a negative correlation between quality and originality of separate ideas. We find no significant correlation between quantity and originality. In col (9) we additionally control for intrinsic motivation, the work minute at which an illustration was created and the number of characters of the illustrated word. We find no significant effect of intrinsic motivation, while the average originality increases with the work minute and the number of characters of the illustrated word. Hence, we find that neither incentives nor intrinsic motivation have a significant effect on the average originality of creative ideas while the timing and the length of illustrated words do matter.

<sup>9</sup> Note that in the models in col (1 to 3) we analyze data on an individual level and thus could not include the work minute and the no. of characters of a word since these variables are measured on illustration level.

TABLE 3: REGRESSION RESULTS FOR SEPARATE DIMENSIONS OF CREATIVITY (1)

	(1) quantity (OLS)	(2) quantity (OLS)	(3) quantity (OLS)	(4) (average) quality (re)	(5) (average) quality (re)	(6) (average) quality (re)	(7) (average) originality (re)	(8) (average) originality (re)	(9) (average) originality (re)
incentivize quantity	9.467*** (2.427)	7.505*** (2.283)	7.949*** (2.184)	-0.080** (0.038)	-0.024 (0.022)	-0.018 (0.022)	0.036 (0.040)	-0.018 (0.025)	-0.017 (0.022)
incentivize quantity & quality	1.436 (1.757)	2.080 (1.534)	2.040 (1.592)	0.034 (0.033)	0.006 (0.020)	0.004 (0.020)	-0.049 (0.034)	-0.028 (0.023)	-0.018 (0.020)
incentivize quantity & originality	8.212*** (2.230)	7.086*** (2.084)	6.703*** (2.034)	-0.042 (0.033)	-0.005 (0.023)	-0.010 (0.023)	0.009 (0.031)	-0.023 (0.025)	-0.017 (0.023)
female	-2.556 (1.771)	0.281 (1.738)	-0.323 (1.675)	0.131*** (0.026)	0.064*** (0.016)	0.056*** (0.016)	-0.090*** (0.025)	-0.022 (0.015)	-0.022 (0.014)
quantity	-	-	-	-	-0.003*** (0.001)	-0.004*** (0.001)	-	0.002 (0.001)	0.002* (0.001)
(average) quality	-	-30.919*** (7.135)	-32.202*** (6.695)	-	-	-	-	-0.541*** (0.019)	-0.464*** (0.019)
(average) originality	-	-10.266 (11.665)	-11.316 (11.182)	-	-0.608*** (0.018)	-0.533*** (0.018)	-	-	-
intrinsic motivation	-	-	1.557** (0.605)	-	-	0.012* (0.006)	-	-	0.003 (0.006)
work minute	-	-	-	-	-	-0.007*** (0.001)	-	-	0.005*** (0.001)
no. of characters of word	-	-	-	-	-	-0.016*** (0.003)	-	-	0.028*** (0.003)
Constant	18.042*** (1.693)	34.904*** (5.947)	30.704*** (5.820)	0.440*** (0.027)	0.689*** (0.025)	0.825*** (0.039)	0.298*** (0.031)	0.506*** (0.037)	0.206*** (0.045)
Observations	125	125	125	2,648	2,648	2,647	2,648	2,648	2,647
R-squared	0.193	0.316	0.354	-	-	-	-	-	-
Number of id_lab	-	-	-	125	125	125	125	125	125

Notes: Robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Dependent variables: Col. (1-3): overall quantity; Col. (4-6): quality of an illustration; Col. (6-9): originality of an illustration. Col (4-9) display panel regression models controlling for illustrator id. In Col (1 to 3), we control for average originality and average quality on participant level, while we control for originality and quality on the illustration level in Col (4-9).

Summarizing, we find that the effects of incentives on quantity and average quality are in line with the predictions of multitasking theory, while average originality is not significantly affected by piece rate incentives. *Incentivizing quantity* results in an increase in quantity and a decrease in the average quality of creative ideas compared to the *baseline*. These results are partially due to an increase in overall effort provision and partially due to spillover effects. *Incentivizing quantity & quality* does not have a significant effect on any of the dimensions of creativity. While both the quantity and the average quality of illustrations slightly increase, these differences are not significant compared to the baseline. This can be ascribed to the fact that the two incentivized dimensions quantity and average quality are substitutes: An increase in one of these dimensions comes with a decrease in performance in the other dimension. *Incentivizing quantity & originality* results in an increase in the number of illustrations and a slight decrease in the average quality of illustrations. The effects of *incentivizing quantity & originality* on quantity and quality are similar but slightly weaker compared to those of *incentivizing quantity*. Additionally, intrinsic motivation results in an increase in performance in quantity and average quality while average originality is independent of it.

Average originality does not increase with any of the piece rate incentives that we introduce. However, we observe that the number of illustrations of a word that no other participant in the experiment came up with (frequency=1) increases with the overall number of illustrations that an individual illustrates (Spearman,  $\rho=0.6$ ,  $p=0.00$ ). Thus, we have a greater number of original ideas (frequency=1) in the *incentivize quantity* and the *incentivize quantity & originality* treatments compared to the *baseline* (U-test,  $p=0.11$  and  $p=0.15$ , correspondingly) and the *incentivize quantity & quality treatment* (U-test,  $p=0.02$  and  $p=0.03$ , correspondingly). Hence, we observe that average originality is unaffected by incentives and intrinsic

motivation but that the number of original ideas increases with the overall number of ideas an individual has.

We additionally elicited two additional measures of creativity after the experiment: a subjective measure of quality and a subjective measure of flexibility to indicate how different the approaches to generating the separate illustrations are. See Appendix C1 (C3) for instructions of the questionnaire used to assess subjective quality (subjective flexibility). We conducted the same regression analysis as illustrated in Table 3 for this two additional measures and do not find any significant effect of our treatments on any of these measures. The detailed results can be found in Table 4 in Appendix C3.

### *B. Effect on Innovation*

In the previous section, we have shown that incentives have positive direct effects and negative spillover effects on the quantity and average quality of creative ideas, while incentives do not have an effect on the average quality. The open question is how performance in the separate dimensions of creativity actually translates into changes in the innovative success. To result in successful innovation, creative ideas have to be both of high quality and original, and organizations clearly want many ideas of this type. In this section, we analyze the effect of incentives on a combined measure of creativity, i.e., the number of illustrations that are at the same time of high quality and original. In the following, we will refer to this measure as a measure for innovation.

Figure 3 shows the number of illustrations created within the experiment for each quality level. Roughly, one fifth of the illustrations have a quality measure of zero, meaning that none of the raters were able to identify the illustrated word. About one fourth of the illustrations are of quality  $\geq 0.90$ , meaning that at least 9 out of the

10 raters were able to correctly identify the illustrated word. For each quality level, we indicate the number of illustrations of words that were only illustrated once throughout the entire experiment (frequency =1) in black, the number of illustrations of words that were illustrated more than once but at most 15 times in dark grey and the number of illustrations of words that were illustrated more than 15 times in light grey. We find that the vast majority of illustrations that were depicted only once (70 percent) have a quality measure of zero, while 90 percent of the illustrations of quality  $\geq 0.90$  occur more than 15 times. Overall, we only observe 2 out of 2,648 illustrations that are of quality=1 and were illustrated only once in the experiment. We observe 57 illustrations of quality  $\geq 0.90$  that were illustrated fewer than 15 times; this corresponds to 2 percent of the total illustrations generated in our experiment.

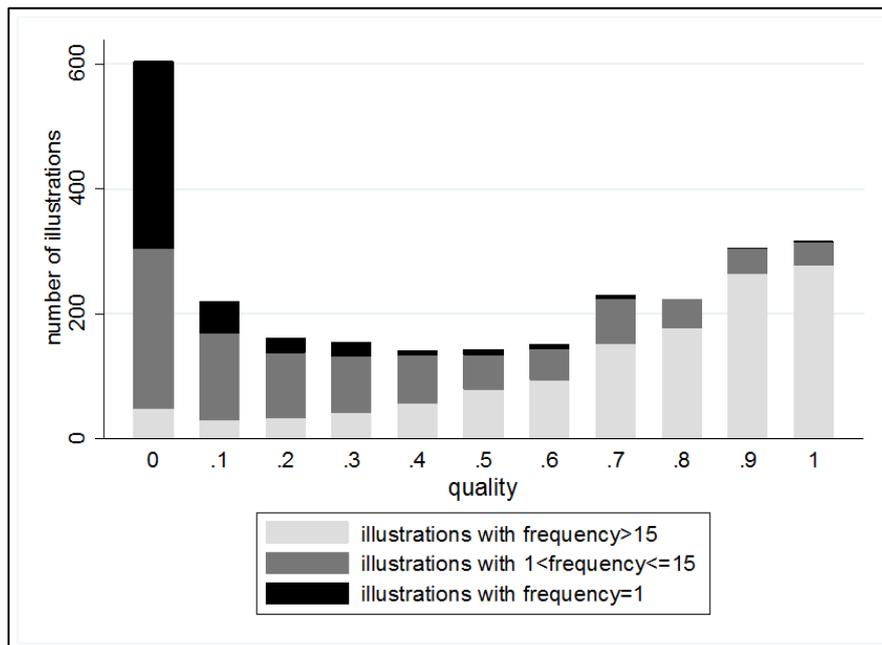


FIGURE 3. NUMBER OF ILLUSTRATIONS BY QUALITY AND ORIGINALITY

Illustrations that are at the same time of high quality and original are rare in our dataset. This is not only true for our specific design but can be considered as a

general property of innovation since good creative ideas are typically rare. Due to this fact, it is difficult to study such innovative ideas, and we can only provide proxies for the number of high quality and original ideas by determining certain thresholds from which point on an idea is classified as high quality and original. In the following we will consider all illustrations of words that occur 15 or fewer times as original and will focus on three different quality measures, i.e.,  $\text{quality} \geq 0.8$ ,  $\text{quality} \geq 0.9$  and  $\text{quality} = 1$ . Table 4 summarizes the number of illustrations that meet these requirements.

TABLE 4 – NO OF ORIGINAL ILLUSTRATIONS (FREQUENCY  $\leq 15$ ) BY TREATMENT AND QUALITY

Treatment	quality $\geq 0.80$	quality $\geq 0.90$	quality = 1.00
<i>baseline</i>	14	11	6
<i>incentivize quantity</i>	23	14	5
<i>incentivize quantity &amp; quality</i>	25	17	9
<i>incentivize quantity &amp; originality</i>	32	19	8

In Table 5, we show regression results for the number of high quality and original illustrations by each participant in our experiment. In Appendix D1 and D2, we provide robustness checks with varying originality requirements. In col (1), we consider the number of illustrations that at least 8 out of 10 raters correctly identified, controlling for treatment and gender. It seems that piece rate incentives, generally, result in an increase in the number of original illustrations with  $\text{quality} \geq 0.8$  compared to the *baseline*. This increase is significant in the *incentivize quantity & quality* treatment and in the *incentivize quantity & originality* treatment. In col (2), we include controls for quantity, average quality, and average originality. We find that the coefficients for all dimensions of creativity are highly significant, which suggests that performance in this combined measure is indeed related to performance in the separate dimensions of creativity. In addition, we observe that the coefficients for the treatment effects are reduced by the introduction of these additional control variables. Thus, our separate dimensions explain part of the

treatment effects on the combined measure of creativity. However, we observe significant and strong treatment effects, even after controlling for performance in the separate dimensions. This suggests that the effect of incentives on creative performance goes beyond the effect of incentives on the separate dimensions of creativity. In col (3) we include control variables for intrinsic motivation, subjective quality and flexibility. We observe a positive significant but small correlation between the subjective quality measure and the objective quality  $\geq 0.8$ .

Col (4-6) show the results for illustrations with quality  $\geq 0.90$ . While coefficients for all treatments are still positive, we only find a significant increase in the number of original high quality illustrations for the *incentivize quantity & originality* treatment. Again, this result remains significant even after controlling for performance in the separate dimensions of creativity. In col. (7-9) we consider original illustrations with maximum possible quality. For this quality requirement, we do not observe a significant effect of incentives on creativity. Remember that the number of illustrations that meet these quality requirements is very small. The regression analysis reveals that *incentivizing quantity & originality* performs best in generating innovative ideas. The open question is why the *incentivize quantity & originality* treatment performs significantly better than the *baseline*, even after controlling for average performance in the three separate dimensions of creativity. It may be that, in addition to affecting average performance in the separate dimensions, incentives also have an impact on the degree of experimentation that participants engage in. If this is the case, we should observe differences in the number of outlier ideas. A higher degree of experimentation is likely to result in a greater number of both very good and very bad illustrations.

TABLE 5 – REGRESSION ANALYSIS FOR ILLUSTRATIONS WITH FREQUENCY  $\leq 15$  (1)

	Quality $\geq 0.8$			Quality $\geq 0.9$			Quality=1		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
incentivize	0.338	0.281	0.322	0.0632	0.0188	0.0177	-0.0410	-0.0194	0.0314
quantity	(0.240)	(0.244)	(0.244)	(0.191)	(0.180)	(0.190)	(0.120)	(0.124)	(0.117)
incentivize	0.365*	0.309*	0.331*	0.187	0.161	0.190	0.0777	0.0653	0.0906
quantity & quality	(0.213)	(0.183)	(0.179)	(0.170)	(0.151)	(0.149)	(0.141)	(0.118)	(0.118)
incentivize	0.836***	0.721***	0.636***	0.423*	0.344*	0.298	0.128	0.117	0.0777
quantity & originality	(0.306)	(0.244)	(0.234)	(0.226)	(0.207)	(0.202)	(0.139)	(0.133)	(0.129)
female	0.250	-0.0670	-0.132	0.293**	0.0894	0.0535	0.180*	0.0168	-0.0460
	(0.197)	(0.170)	(0.177)	(0.146)	(0.128)	(0.138)	(0.0965)	(0.0888)	(0.0912)
quantity		0.0451***	-0.0636*		0.0312***	-0.0258		0.0169***	-0.0357**
		(0.0124)	(0.0382)		(0.00789)	(0.0253)		(0.00556)	(0.0154)
(average) quality		6.578***	5.067***		4.535***	3.747***		3.267***	2.417***
		(0.972)	(1.003)		(0.821)	(0.874)		(0.643)	(0.615)
(average) originality		4.505***	4.036***		3.306***	3.085***		2.355***	2.069***
		(0.865)	(0.803)		(0.682)	(0.667)		(0.558)	(0.514)
intrinsic motivation			0.0135			-0.041			0.060*
			(0.065)			(0.060)			(0.034)
subjective quality evaluation			0.002***			0.001**			0.001***
			(0.00)			(0.000)			(0.000)
flexibility			0.513			0.589			0.441
			(0.721)			(0.567)			(0.413)
Constant	0.477***	-4.614***	-4.282***	0.300**	-3.271***	-3.273***	0.166	-2.298***	-2.384***
	(0.163)	(0.796)	(0.939)	(0.134)	(0.608)	(0.790)	(0.110)	(0.462)	(0.577)
Observations	125	125	125	125	125	125	125	125	125
R-squared	0.089	0.332	0.391	0.069	0.279	0.316	0.039	0.277	0.359

Notes: Robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Dependent variable: no. of original illustrations per subject. Our main results from this regression analysis are unchanged if we add a control for the self-evaluation of how well participants think they performed in this task.

Table 6 provides summary statistics on the number of high and low quality and the number of high and low originality illustrations for each treatment and overall. We find that the number of illustrations with high quality (quality=1; quality $\geq$ 0.90; quality $\geq$ 0.80) is largest in the incentivize *quantity & originality* treatment, and the difference to the *baseline* is significant (pairwise U-test,  $p\leq 0.02$ , two-tailed). At the same time, the number of illustrations that no rater identified (quality =0) is also (insignificantly) larger in the incentivize *quantity & originality* treatment compared to the *baseline* (U-test,  $p=0.14$ , two-tailed). Remember that the average quality is not different in the incentivize *quantity & originality* treatment compared to the *baseline* (see Table 2 and analysis above). We observe a similar pattern for the dimension originality. We see a higher number of both high and low originality illustrations in the *incentivize quantity & originality* treatment compared to the *baseline*. For illustrations of words with a frequency of 1, this difference is not significant (U-test,  $p=0.11$ , two-tailed), whereas it is significant for illustrations of words that occur more than once but 15 times or fewer (U-test,  $p<0.01$ , two-tailed), for word illustrations that occur more than 15 times in our dataset (U-test,  $p=0.01$ , two-tailed) as well as for word illustrations that occur more than 50 times in our dataset (U-Test,  $p<0.1$ , two-tailed). Recall that the average originality in the *incentivize quantity & originality* treatment is lower but not significantly different to the *baseline* (see Table 2 and analysis above). Thus, it seems that participants in the *incentivize quantity & originality* treatment create more illustrations with extreme values for quality and originality, i.e., participants in these treatments explore more. This may be the reason why the probability that one specific illustration is at the same time of high quality and original increases in this treatment compared to the *baseline*.

TABLE 6 – NO. OF HIGH AND LOW QUALITY AND ORIGINALITY ILLUSTRATIONS

	<i>baseline</i>	<i>incentivize quantity</i>	<i>incentivize quantity &amp; quality</i>	<i>incentivize quantity &amp; originality</i>	<i>Overall</i>
Avg. no. of illustrations with quality=1	2 (2.86)	2.16 (1.46)	2.9 (1.99)	3 (1.64)	2.50 (1.69)
Avg. no. of illustrations with quality $\geq 0.90$	4.09 (2.01)	4.81 (2.65)	5.03 (2.33)	5.94 (3.10)	4.98 (2.61)
Avg. no. of illustrations with quality $\geq 0.80$	5.38 (2.77)	7.10 (3.59)	6.7 (2.74)	7.78 (3.76)	6.74 (3.33)
Avg. no. of illustrations with quality =0	4 (4.18)	6.42 (8.64)	3.1 (3.58)	5.75 (5.92)	4.83 (5.60)
Avg. no. of illustrations with frequency=1	2.91 (3.59)	4.71 (7.49)	2.07 (2.16)	3.94 (3.71)	3.42 (4.72)
Avg. no. of illustrations with $1 < \text{frequency} \leq 15$	5.66 (3.96)	9.26 (3.79)	6.37 (2.98)	9.5 (4.45)	7.70 (4.17)
Avg. no of illustrations with frequency >15	7.88 (3.84)	11.48 (5.12)	9.77 (3.20)	11.16 (4.43)	10.06 (4.40)
Avg. no. of illustrations with frequency >50	4.50 (2.29)	5.6 (2.45)	5.20 (1.90)	5.20 (2.06)	5.20 (2.20)

*Notes:* Robust standard errors in parentheses.

#### IV. Conclusion

The relevance of creativity as a driving force of economic growth raises the question of which factors influence creative performance. In this paper, we study which dimensions of creativity are affected by incentives and which dimensions are merely driven by an individual's talent. We find that the quantity and the quality of ideas are significantly affected by incentives. The observed effect of incentives on performance in the quantity and quality dimensions is in line with predictions of economic theory (Holmstrom and Milgrom, 1991, Lazear, 2000). That is, participants respond to incentives in these dimensions of creativity similarly as they do to simple routine tasks (see for example Hong et al. 2013). In contrast, we observe that individuals' performance in coming up with original ideas is not affected by incentives but instead seems to be determined by talent.

To further understand the effect of motivation on creative performance, we also use a combined measure of innovation that captures the number of high quality and original illustrations. We find that performance in this combined measure is best in the *incentivize quantity & originality* treatment. We also observe that the effect of incentives is partially driven by differences in the performances in the separate dimensions of creativity. However, the treatment effect in this combined measure remains large and significant even after controlling for performance in the separate dimensions of creativity. Interestingly, we observe a higher number of outliers both in quality and in the number of original ideas in the *incentivize quantity & originality* treatment compared to the *baseline*. This indicates that incentives can foster a more exploratory approach to generating ideas. One interpretation for this result is that incentives can shift the mode of creative thinking from so called “in the box” thinking towards a more exploratory “out of the box” approach. Future research could address how incentives change participants’ thinking mode. Further, since creativity is highly context-specific, different types of creative tasks may require different degrees of each of the dimensions of performance. It is left to future research to show how the effect of incentives varies with the type of creative task.

To the best of our knowledge, we are the first to show that the effect of incentives on creativity depends crucially on the dimension of creativity that is incentivized. Our results have implications for organizations seeking to foster creativity. We find that incentives do not have an effect on average originality and that the number of original ideas can only be raised by increasing the quantity of ideas. However, we also show that incentives help increasing quantity results in a decrease in the quality. Thus, organizations face a tradeoff when introducing incentives for creative performance.

## REFERENCES

- Al-Ubaydli, O., Anderson, S., Gneezy, U., and J. A. List. 2015. "Carrots That Look Like Sticks: Towards and Understanding of Multitasking Incentive Schemes." *Southern Economic Journal* 81(3). 538-561.
- Amabile, T., 1996. "Creativity in context: Update to the social Psychology of Creativity." New York: Westview Press.
- Amabile, T., 1993. "Motivational synergy: Toward a new conceptualization of intrinsic and extrinsic motivation in the workplace." *Human Resource Management Review* 3. 185-201.
- Ariely, D., Gneezy, U., Loewenstein, G., and N. Mazar. 2009. "Large Stakes and Big Mistakes." *The Review of Economic Studies* 76(2). 451-469.
- Azoulay, P., Graff Zivin, J.S., and G. Manso. 2011. "Incentives and creativity: evidence from the academic life sciences." *The RAND Journal of Economics*, 42(3). 527-554.
- Bénabou, R., and J. Tirole. 2003. "Intrinsic and extrinsic motivation." *Review of Economic Studies* 70. 489–520.
- Bradler, C., Neckermann, S., and A. Warnke. 2016. "Incentivizing Creativity: a Large-Scale Experiment with Tournaments and Gifts." Mimeo.
- Charness, G., and D. Grieco. 2014. "Creativity and Financial Incentives." Mimeo.
- Deci, E.L., Ryan, R.M., and R. Koestner. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125 (6). 627-668.
- Dennis, A.R., Valacich, J.S., Connolly, T., and B. Wynne. 1996. "Process structuring in group brainstorming." *Information Systems Research*, 7. 268–277.
- Diehl, M., and W. Stroebe. 1987. "Productivity loss in brainstorming groups: Toward the solution of a riddle." *Journal of Personality and Social Psychology*, 53. 497–509.

- Eckartz, K., Kirchkamp O., and D. Schunk. 2012. "How do Incentives Affect Creativity?" Working Paper No. 4049. Munich, Germany: CESifo.
- Erat, S., and U. Gneezy. 2015. „Incentives for Creativity.” *Experimental Economics*. 1-12.
- Gneezy, U., 2000. "Pay enough or don't pay at all." *The Quarterly Journal of Economics* 115(3). 791-810.
- Greiner, B., 2004. "An online recruiting system for economic experiments" *Forschung und wissenschaftliches Rechnen* 2003. Vol 63. GWDG Bericht. 79-93.
- Gross, D., 2014. "Creativity Under Fire: The Effects of Competition on Innovation and the Creative Process." Mimeo.
- Harhoff, D., and K. Hoisl. 2007. "Institutionalized incentives for ingenuity – Patent value and the German Employees" *Invention Act. Research Policy* 36. 1143-1162.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership and job design." *Journal of Law, Economics, and Organization* 7 (Special Issue). 24–52.
- Hong, F., Hossain, T., List, J. A., and M. Tanaka. 2013. "Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories." NBER Working Paper, No. 19660.
- Kachelmeier, S.J., Reichert, B.E., and M.G. Williamson. 2008. "Measuring and Motivating Quantity, Creativity, or Both." *Journal of Accounting Research* 46. 341-374.
- Kaufmann, J.C., and R.J. Sternberg. 2010. „The Cambridge Handbook of Creativity." Cambridge University Press.
- Lazear, E.P., 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5). 1346-1361.
- Mohnen, A., and A. Ostermaier. 2013. „Incentives for Creativity: Limits of

Objective Performance Evaluation” Mimeo.

Simon, H.A., 1983. “Discovery, invention, and development: Human creative thinking.” *Proceedings of the National Academy of Sciences* 80. 4569-4571.

Sliwka, D., 2007. “Trust as a signal of a social norm and the hidden costs of incentive schemes” *American Economic Review* 97(3). 999–1012.

The Conference Board, 2012. *CEO Challenge 2012*, Research Report R-1491-12-RR.

The Conference Board, 2013. *CEO Challenge 2013*, Research Report R-1511-13-ES.

The Conference Board, 2014. *CEO Challenge 2014*, Research Report R-1537-14-RR.

## For Online Publication

### Appendix A: Instructions lab experiment (translation from German)

#### *Instructions*

Welcome to this experiment!

Please carefully read the following instructions. If you have any questions, raise your hand. We will come to you and answer your question. Please do not **begin** the experiment until we ask you to do so. None of the other participants will receive information about your payoff. Communication with other participants is forbidden throughout the entire experiment. We also request that you switch off your mobile phone and remove it from the desk.

*Task.* - Immediately before the start of the task, you will receive various materials. The task consists of illustrating words with the provided set of materials. The goal is:

- To illustrate as many different words as possible,
- Which can be identified by others,
- And that the illustrated words are unique, meaning that they were not illustrated by any of the participants in the randomly selected four-person group.

After the experiment, we will evaluate how well you achieved this goal.

Please proceed with the illustration of each word in the following manner:

- i. Illustrate the word in the designated area using the provided materials.

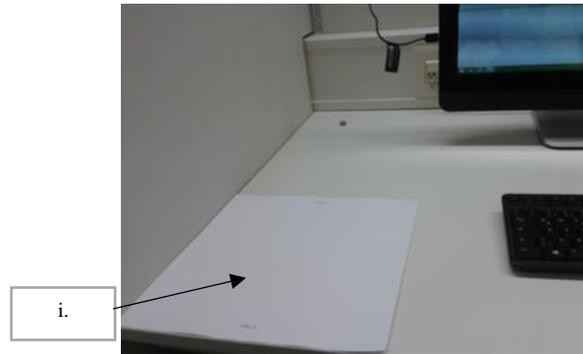


FIGURE A1

- ii. Take a picture of the illustrated word.
- iii. Enter the word that you illustrated in the field “illustrated word”.
- iv. Save the picture by clicking on the “save” button.

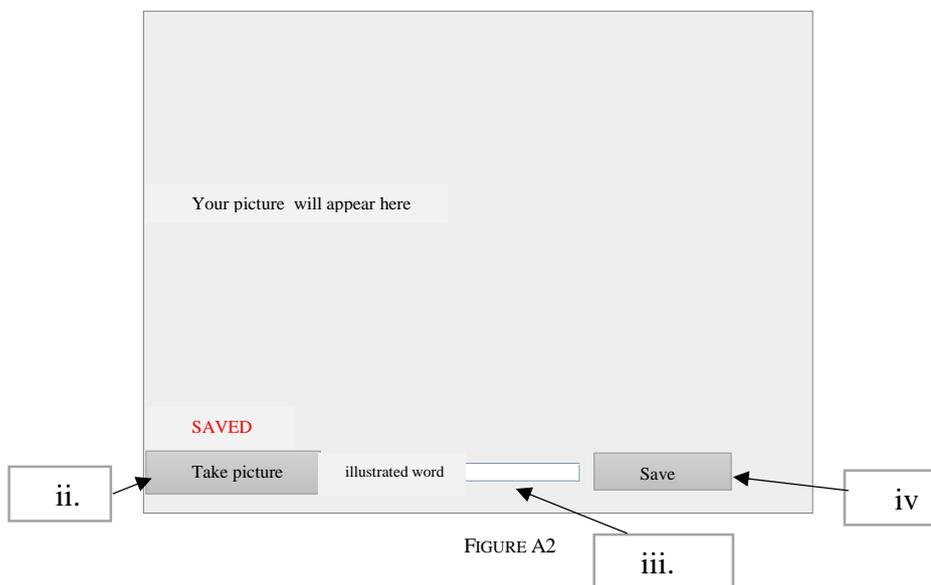


FIGURE A2

Please keep the following in mind:

- Use **only** the materials provided.
- **For each** illustrated word, you can use all of the materials or a selection of them.
- The illustration of the word should only be placed **within** the designated area on the sheet of paper (only this area will be captured by the camera).
- Make sure that your illustration is made in the correct **direction** (the sheet is marked “top” and “bottom”).
- Make sure that your **hands are not visible** in the designated area.
- Keep any **unused materials outside** of the designated area.
- Illustrate only **one** word at a time. This means that the name of the picture should only consist of **one word**. Terms that consist of multiple words are not permitted and will not be evaluated.
- You may only illustrate each word **once**.
- Your illustrations may not include any symbol that is depicted on the keyboard (for example, illustrations that include “→”, “8”, “b”, “@”, “>” or “+” are not permitted).

*Time.* - You have a total of 20 minutes for this task. After this time has expired, we ask you to answer questionnaire before the end of the experiment.

*Payment.* - [This part is different with regard to the four treatments of the experiment]

Baseline: You are paid €10 for this task. In addition, you receive a show-up payment of €2.50. You will receive your payment two weeks after the experiment takes place. You can choose whether you would like to receive an electronic transfer or pick up the payment in cash.

Incentivize Quantity: You are paid €0.60 for each admissible word that you illustrate. You also receive a show-up fee of €2.50. You can choose whether you would like to receive an electronic transfer or pick up the payment in cash.

Incentivize Quantity & Quality: After this experiment, we will show the pictures of all of the admissible words you illustrated to other people. These other persons have not participated in this experiment or similar experiments. The task assigned to them is to identify the illustrated words using the pictures taken in the experiment. These other persons only receive a positive payout if they enter exactly the word that you saved along with the respective picture.

Each word will be presented to ten other people. We measure how many of these ten people correctly identify the respective word. For each illustrated word, you are paid **€0.10** for each person who correctly identifies it. That means you can earn up to **€1** for each illustrated word, assuming it is correctly identified by each of the ten people. In addition, you receive a show-up payment of €2.50. You will receive your payment two weeks after the experiment takes place. You can choose whether you would like to receive an electronic transfer or pick up the payment in cash.

Incentivize Quantity & Originality: After this experiment, you will be randomly assigned to a group of four people who participated in the same experiment. For each admissible word that you alone in the group illustrated, you are paid €0.85. If at least one other person in the group illustrated the same word, then you receive €0 for illustrating this word. In addition, you receive a show-up payment of €2.50. You will receive your payment two weeks after the experiment takes place. You can choose whether you would like to receive an electronic transfer or pick up the payment in cash.

## **Appendix B: Online survey to assess (objective) quality**

### **B1: Instructions (translation from German)**

#### *Instructions*

Please carefully read the following instructions. If you have any questions about these instructions or if you have any trouble with the experiment, please contact us by e-mail at [internetexperimente@wiso.uni-koeln.de](mailto:internetexperimente@wiso.uni-koeln.de). Please note that you are not allowed to go back to a previous page at any time during the experiment. Next, you will see 50 consecutive pictures on your screen. These pictures were taken by participants in a prior experiment. These participants' task was to illustrate words using the materials provided. The words could be chosen freely and had to consist of only one word.

*Your Task.* - **Your task is to identify the illustrated words.** In order to receive payment for a picture, you must enter the **exact** word that the other participant assigned to that picture. If you do not make an entry for a picture, or if the word you enter does not exactly correspond to the respective word assigned by the other participant, then you do not receive any payment for this picture. Please take note of the fact that each of the illustrated terms consists of only **one word**. Your entries may also only consist of one word each. If you enter more than one word for a picture, it will be classified as 'not identified.'

Please also note that the words were illustrated by different participants. This means that it is possible to see more than one illustration of the same term.

*Payment.* - You will receive your payment only if you complete the entire experiment. You receive €2.00 for participating in the experiment. In addition, you receive €0.10 for each picture that you correctly identify. At the end of the

experiment, you can choose whether you would like to receive an electronic transfer or pick up the payment in cash.





Welcher Begriff wird oben dargestellt?  
(Sie dürfen nur ein Wort eingeben)

What word is illustrated above? (You may enter one word)

NECKLACE

Weiter

Universität zu Köln



FIGURE B1. SCREEN OF QUESTIONNAIRE (EXAMPLE)

## Appendix C: Additional Measures of Creativity

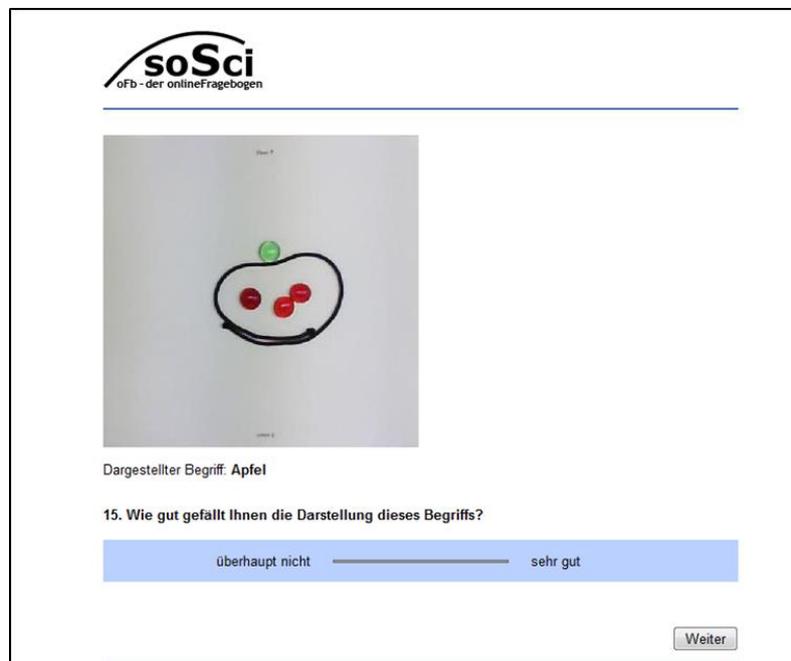
### C1: Instructions to assess subjective quality (translation from German)

#### *Instructions*

Please carefully read the following instructions. Please note that you are not allowed to go back to a previous page at any time during the experiment.

Next, you will see consecutive photos on your screen. These photos were taken by participants in a prior experiment. These participants' task was to illustrate words using the materials provided. The words could be chosen freely and had to consist of only one word. Which word is illustrated in the photo is written below the photo.

*Your Task.* - Your task is to evaluate on a scale between 0 (not at all) to 100 (very much) how much you like the picture.



The screenshot shows a questionnaire interface. At the top left is the logo for 'soSci ofB - der onlineFragebogen'. Below the logo is a photograph of a drawing on a white surface. The drawing depicts an apple with a green stem and leaf, and three red dots representing seeds. Below the photograph, the text reads 'Dargestellter Begriff: Apfel'. Underneath this, the question is '15. Wie gut gefällt Ihnen die Darstellung dieses Begriffs?'. A horizontal rating scale is shown with 'überhaupt nicht' on the left and 'sehr gut' on the right. A 'Weiter' button is located at the bottom right of the screen.

FIGURE C1. SCREEN OF QUESTIONNAIRE (EXAMPLE)

## **C2: Instructions to assess subjective flexibility (translation from German)**

### *Instructions*

You have in front of you a USB drive containing several folders. Each folder contains images that were created in a laboratory experiment. The instructions for this experiment are attached. Please carefully read through these before continuing. If you have any questions, let us know.

The images found in the folders were created by one participant in the experiment (per folder), respectively, and are organized in chronological order. Please count the number of distinct thought categories and enter this number into the Excel file saved on the USB drive.

A new unique thought category should be counted as such if the participant's train of thought cannot be clearly recognized when going from one image to the next. In other words, as long as you can imagine how the participant got the idea to illustrate a particular idea, then those images belong to the same thought category.

### *Examples:-*

- i. A participant illustrates: 1. *stroller*, 2. *car*, 3. *bicycle* and 4. *Tractor*  
It is evident here that the participant was thinking about vehicles and illustrated all of the types of vehicles that occurred to him. The images thus fall under the same thought category.
  
- ii. A participant illustrates different words using one piece of string: 1. *pretzel*, 2. *infinity sign* and 3. *Worm*  
The participant found a material (or a piece of the materials) and illustrated all possible objects using the same material. Also, in this case, the images fall under the same thought category.

- iii. A participant illustrates: 1. *stroller*, 2. *infinity sign* and 3. *snail*, each using different materials

Here there is no recognizable connection between the individual images. Therefore, the images each fall under different thought categories (that is, you would enter “3” for the number of thought categories).

When counting, please **exactly** follow the order in which the participant illustrated the images. This order is the same order by which the images are arranged in the folders. Pay attention that you do not rearrange the images.

*Examples.-*

- i. A participant illustrates: 1. *tractor*, 2. *worm* and 3. *car*, each using different materials.

The number of unique thought categories for this participant is 3.

- ii. A participant illustrates: 1. *tractor*, 2. *car* and 3. *worm*, each using different materials.

The number of unique thought categories for this participant is 2.

### C3: Results for subjective quality and subjective flexibility

TABLE 4 – REGRESSION RESULTS FOR SUBJECTIVE QUALITY AND SUBJECTIVE FLEXIBILITY

	subjective quality (re)	subjective quality (re)	subjective quality (re)	flexibility (OLS)	flexibility (OLS)	flexibility (OLS)
incentivize quantity	-1.598 (1.533)	-0.133 (1.467)	0.411 (1.387)	4.849 *** (1.239)	0.548 (0.583)	0.532 (0.589)
incentivize quantity & quality	-0.546 (1.513)	-0.915 (1.407)	-0.661 (1.333)	0.265 (1.017)	-0.646 (0.553)	-0.648 (0.555)
incentivize quantity & originality	0.810 (1.386)	1.809 (1.315)	2.019 (1.260)	4.081 *** (1.254)	0.253 (0.559)	0.253 (0.561)
female	2.313** (1.063)	0.866 (1.121)	0.363 (1.115)	0.584 (0.859)	1.474 *** (0.432)	1.485 *** (0.422)
quantity		-0.099 (0.060)	-0.094* (0.056)		0.463 *** (0.042)	0.464 *** (0.045)
(average) quality		5.815*** (0.890)	8.772*** (0.907)		0.062 (2.329)	0.116 (2.425)
(average) originality		-4.000 (5.191)	-6.600 (4.681)		-3.864 (2.517)	-3.835 (2.623)
intrinsic motivation			0.761** (0.382)			-0.028 (0.188)
work minute			0.215*** (0.063)			
no. of characters of word			1.040*** (0.151)			
Constant	55.960** * (1.014)	56.548*** (1.683)	44.230*** (2.378)	11.976*** (0.915)	4.743 ** (2.131)	4.784 ** (2.083)
Observations	2,416	2,416	2,415	125	125	125
R-squared				0.211	0.788	0.788
Number of id_lab	125	125	125			

Notes: Robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Dependent variables: Col. (1-3): overall flexibility; Col. (4-6): subjective quality of an illustration. In Col (1 to 3), we control for average originality and average quality on participant level, while we control for originality and quality on the illustration level in Col (4-6)

## Appendix D: Robustness tests for analysis of the number of high quality and original ideas

TABLE D1 – REGRESSION ANALYSIS FOR ILLUSTRATIONS WITH FREQUENCY  $\leq 20$

	Quality $\geq 0.8$		Quality $\geq 0.9$		Quality=1	
	(1)	(2)	(3)	(4)	(5)	(6)
incentivize quantity	0.470 (0.294)	0.329 (0.278)	0.0981 (0.226)	0.00149 (0.205)	-0.0375 (0.123)	0.00605 (0.124)
incentivize quantity & quality	0.436* (0.238)	0.326 (0.203)	0.222 (0.197)	0.156 (0.178)	0.0458 (0.155)	0.0304 (0.127)
incentivize quantity & originality	0.938*** (0.326)	0.782*** (0.267)	0.549** (0.253)	0.445** (0.223)	0.0997 (0.150)	0.121 (0.134)
female	0.275 (0.218)	-0.0671 (0.197)	0.351** (0.168)	0.124 (0.161)	0.148 (0.103)	-0.0374 (0.0923)
intrinsic motivation	0.178** (0.0849)	0.0657 (0.0784)	0.0741 (0.0699)	-0.00320 (0.0677)	0.119*** (0.0386)	0.0795** (0.0375)
quantity		0.0565*** (0.0135)		0.0387*** (0.00899)		0.0162** (0.00631)
average quality		7.415*** (1.128)		5.142*** (0.892)		3.744*** (0.672)
average originality		4.418*** (0.979)		3.222*** (0.686)		2.592*** (0.585)
constant	0.00400 (0.317)	-5.274*** (0.868)	0.110 (0.253)	-3.591*** (0.664)	-0.169 (0.157)	-2.777*** (0.533)
observations	125	125	125	125	125	125
R-squared	0.124	0.363	0.100	0.300	0.091	0.336

*Notes:* Standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Dependent variable: no of original illustrations per subject. Our main results from this regression analysis are unchanged if we add a control for the self-evaluation of how well participants think they performed in this task.

TABLE D2 – REGRESSION ANALYSIS FOR ILLUSTRATIONS WITH FREQUENCY ≤ 10

	Quality≥0.8		Quality≥0.9		Quality=1	
	(1)	(2)	(3)	(4)	(5)	(6)
incentivize	0.286	0.210	0.0743	0.0234	-0.0248	0.0253
quantity	(0.208)	(0.212)	(0.179)	(0.177)	(0.103)	(0.108)
incentivize	0.388*	0.359**	0.217	0.195	0.106	0.111
quantity & quality	(0.202)	(0.176)	(0.164)	(0.147)	(0.128)	(0.109)
incentivize	0.507**	0.431**	0.213	0.161	0.0157	0.0504
quantity & originality	(0.253)	(0.214)	(0.178)	(0.167)	(0.111)	(0.110)
female	0.155	-0.0604	0.157	-0.00114	0.0862	-0.0357
	(0.170)	(0.152)	(0.134)	(0.126)	(0.0837)	(0.0742)
intrinsic	0.111	0.0331	0.0438	-0.0119	0.119***	0.0944***
motivation	(0.0697)	(0.0634)	(0.0591)	(0.0582)	(0.0284)	(0.0271)
quantity		0.0377***		0.0269***		0.00888*
		(0.0103)		(0.00665)		(0.00484)
average quality		5.551***		3.997***		2.601***
		(0.902)		(0.695)		(0.550)
average		4.054***		2.895***		1.993***
originality		(0.775)		(0.570)		(0.507)
constant	-0.0209	-4.132***	0.118	-2.830***	-0.264**	-2.100***
	(0.256)	(0.735)	(0.203)	(0.540)	(0.108)	(0.435)
observations	125	125	125	125	125	125
R-squared	0.073	0.301	0.039	0.256	0.117	0.309

Notes: Standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Dependent variable: no of original illustrations per subject. Our main results from this regression analysis are unchanged if we add a control for the self-evaluation of how well participants think they performed in this task.