Piece-rate incentives and idea generation — An experimental analysis*

KATHARINA LASKE,† NATHALIE RÖMER‡ AND MARINA SCHRÖDER§

December 2024

Abstract

We introduce the word illustration task (WIT), a novel experimental task to quantify performance in an idea generation context. Between treatments, we vary whether or not piece-rate (PR) incentives are implemented and the degree to which these incentives are aligned with the desirable outcome. We show that PR incentives have a positive impact on the number of innovative ideas, i.e., the number of ideas that are of high quality and original. We find that unweighted PR incentives (PR provided for any idea) perform at least as well as more aligned weighted PR incentives that are additionally contingent on the quality and/or originality of ideas. Our results suggest that when it comes to fostering idea generation, it is sufficient to incentivize trying instead of incentivizing succeeding.

JEL-Classification: C90, J33, M52, O31.

Keywords: Innovation, Idea Generation, Real-effort Experiment, Incentives, Creativity

---

*We circulated a previous version of this paper including only the data from the first experiment under the title: "Quantity, quality and originality: The effects of incentives on creativity." Special thanks go to the student assistants for their help in programming and conducting the lab experiment. Financial support of the University of Cologne Female Research Grant and the Deutsche Forschungsgemeinschaft [DFG FOR 1371, DFG 456592273] is gratefully acknowledged.

†University of Cologne, Faculty of Management, Economics and Social Sciences, Albertus-Magnus-Platz, 50923 Cologne, Germany ORCID: 0000-0003-2865-6127, Email: laske@wiso.uni-koeln.de.

‡Leibniz University Hanover, Faculty of Economics and Management, Königsworther Platz 1, 30167 Hannover, ORCID: 0009-0005-1090-4139, Email: roemer@inec.uni-hannover.de.

§Leibniz University Hanover, Faculty of Economics and Management, Königsworther Platz 1, 30167 Hannover and CESifo, ORCID: 0000-0002-3265-6317, Email: schroeder@inec.uni-hannover.de.

1 Introduction

Progress typically starts with an innovative idea, i.e., with an idea that is both original and of high quality. It is important for organizations to understand the effect of incentives on the generation of such innovative ideas. Previous research provides insights on the effect of incentives in the context of routine tasks (see Prendergast, 1999, for an overview of the literature), however, there are some indications that incentives do not affect performance in the same way when considering more complex tasks such as idea generation (see Ederer & Manso, 2011, for an overview). In this paper, we study the effect of PR incentives on the generation of innovative ideas. To do so, we introduce a novel experimental task and study the effect of piece-rate (PR) incentives on the generation of innovative ideas. Further, we study whether and how this effect depends on the degree to which incentives are aligned with the desired outcome.

We introduce a novel experimental task, the word illustration task (WIT). In WIT, ideators are asked to illustrate words using a given set of materials within a predetermined working time. Ideators are instructed to illustrate as many words as possible that provide value for customers (quality), and that are statistically infrequent (originality).[1] We take a product innovation approach and define ideas as innovative when they are at the same time original and create value to customers. WIT captures dimensions of performance that are typically relevant in the context of innovation. Quality and originality are, for example, requirements for patentability.

To be successful, ideators in WIT have to come up with an underlying idea (a word they want to illustrate) and have to provide effort to realize this idea. While originality is directly related to the underlying idea, the number of ideas and their quality are related to both the underlying ideas and effort provision. WIT allows us to analyze effects of incentives on both effort provision and underlying ideas, and thus, to derive important insights into how incentives affect performance in idea generation.

In this paper, we provide results from two experimental studies implementing a lab and an online version of WIT. Between treatments, we vary the incentives faced by ideators. In

[1] Customers in WIT are additional study participants who are incentivized to identify the exact word illustrated by an ideator. Thus, the utility of customers increases in how easily the illustrated words can be identified.

both the lab and the online version of WIT, we implement a treatment in which ideators receive a fixed payment that is not contingent on performance (No PR) and a treatment in which ideators receive a bonus payment that is contingent only on the number of ideas generated independent of the quality and originality of these ideas (Unweighted PR). By comparing performance in the respective No PR and Unweighted PR treatments, we derive insights into the effectiveness of unweighted PR incentives in the context of idea generation.

In the next step, we analyze whether the alignment of PR incentives matters. In routine tasks, firms typically seek a close alignment of incentives and desirable outcomes (Holmstrom, 1979; Kerr, 1975). Due to challenges in performance measurement (Cordero, 1990; Smith, 2006; Ederer & Manso, 2011), such an alignment is hard to achieve in the context of idea generation. In any case, it is unclear whether a stronger alignment actually has a desirable impact on performance. In fact, recent empirical research reveals that for complex tasks a closer alignment of incentives does not enhance the effectiveness of incentives but may actually undermine it (Azoulay et al., 2011; Butschek et al., 2019). Why would alignment have an adverse effect on performance? On the one hand, adverse effects of control (see e.g., Frey & Jegen, 2001; Falk & Kosfeld, 2006; Belot & Schröder, 2013; Herz & Zihlmann, 2021) are likely to increase in the degree to which incentives are aligned with the desirable outcome. On the other hand, alignment may be related to distortion of effort. When incentives cannot be fully aligned, it may be better to provide less aligned incentives in order to avoid inefficient distortions in effort provision (Holmstrom & Milgrom, 1991; Rubin et al., 2018).

To understand the relevance of alignment, we vary the degree to which PR incentives are contingent on additional performance dimensions between treatments. In the lab experiment, we study PR incentives that are more closely but still imperfectly aligned to the desired outcome of generating innovative ideas. In the Quality-Weighted PR treatment, we implement imperfectly aligned incentives where the level of the PR paid for each idea is contingent on the quality (while originality is not payoff relevant). In the Originality-Weighted PR treatment, the PR is paid only for ideas that are original within a reference group but incentives are not contingent on the quality of the respective ideas. We predict that compared to the Unweighted PR, both the Quality-Weighted PR and the Originality-Weighted PR may lead to distortions of effort and to adverse effects on effort provision due

to increased control. In the online experiment, we study PR incentives that are perfectly aligned with the desirable outcome, i.e., the PR is only provided for innovative ideas. As incentives in the Innovation-Weighted PR treatment are perfectly aligned with the desirable outcome, we do not expect to observe inefficiencies arising due to distortions. However, we may observe adverse effects of control.

We can show that unweighted PR incentives have a positive effect on the number of innovative ideas generated. We also find evidence of a shift in the types of ideas that are generated under PR incentives. Specifically, PR incentives lead to a stronger focus on easy solutions that can be realized in less time and to the generation of innovative ideas that are less complex, i.e., involve the use of fewer materials. Additionally, we can show that under PR incentives it is harder for individuals to identify their best ideas.

We find that aligning incentives more closely with the desired outcome does not lead to an increase in the number of innovative ideas. In fact, we find adverse effects of a quality weight and an innovation weight. In both the Innovation-Weighted PR and the Quality-Weighted PR treatments, we find evidence for adverse effects of control, where ideators provide less effort when incentives are more aligned with the desirable outcome. For the quality weight, we find additional evidence for distortions in effort provision. Interestingly, we do not find an adverse effect of an originality weight. This result suggests that incentives targeting specific types of ideas (originality) affect behavior differently than incentives targeting effort provision.

In summary, we show that incentives are effective in fostering the generation of innovative ideas. When incentivizing idea generation, it seems to be enough to reward trying (generating many ideas) while a closer alignment of incentives does not provide additional benefits. In line with previous research (Englmaier et al., 2024), we show that incentives have an impact on the problem-solving approach. Specifically, we find that ideators generate ideas that are less complex when faced with PR incentives. More research is needed to understand the implications of this shift in the types of ideas.

Our research is most closely related to previous research in creativity, which reveals that creativity is linked to economic success (Gill and Prowse, 2023) but provides mixed results

with respect to the effectiveness of incentives in the context of fostering creativity.[2] The impact of incentives seems to depend on the underlying performance measure. For instance, Bradler et al. (2019) show that gifts are not effective in enhancing performance in an ideation task while they are effective in a routine task. Bradler et al. (2019) attribute differences between routine and ideation tasks to differences in the ambiguity of performance measures. While ideators know what is expected from them in the routine task, this may not be the case for the creative task. In line with this interpretation, Kachelmeier et al. (2008) show that incentives contingent on the number of ideas lead to an increase in the number of ideas, while incentives contingent on the creativity (a performance measure that is rather vaguely defined) do not. Further studies that are consistent with the interpretation that performance-related incentives are only effective when the desirable outcome is clearly defined include Eckartz et al. (2012), Erat & Gneezy (2016), Gibbs et al. (2017), and Charness & Grieco (2019).[3] One benefit of the experimental design presented in this paper is that it allows us to derive clear quantitative measures of performance that can be easily communicated. To rule out that treatment differences that arise due to ambiguity in the interpretation of desired behavior, all ideators in our experiment are informed about the relevant performance dimensions and the procedure of measuring performance. In the absence of ambiguity with respect to the relevant performance measures, we can show that unweighted PR incentives actually outperform incentives that are more closely aligned to the desired outcome.

Another contribution of this paper is the introduction of WIT, a relatively simple experimental task that can be used to test various interventions in the context of idea generation. We introduce a lab and an online version of WIT. To facilitate the implementation of WIT for future researchers, we provide detailed accompanying materials and an oTree toolbox accompanying this paper. WIT can be implemented in a wide range of contexts. For example, WIT is applied to study collaboration in idea generation (Grözinger et al., 2020),

[2]Different to this closely related research, we refer to idea generation instead of creativity. We chose the wording idea generation to underline the fact that we focus on performance in a specific context in which the quality measure is pre-determined, i.e., it is related to clearly defined value created for customers.

[3] Other related research reveals that framing incentives as either losses or gains has little impact on ideation performance (see Kleine, 2021; Lagarde & Blaauw, 2021). For more general overviews on the research on the impact of incentives on idea generation (see Beckers et al., 2010; Attanasi et al., 2021).

biases in support for innovation (Römer & Schröder, 2024), and subjective performance evaluations (Petters & Schröder, 2020).

Our insights are highly relevant for policymakers and managers seeking to promote innovation by fostering idea generation. Our results suggest that simple incentives targeting an increase in the number of ideas are suitable for fostering a generation of innovative ideas. We can show that more complex incentives schemes that are more closely aligned with the desirable outcome do not provide additional benefits, despite the additional costs they are likely to infer due to the required performance measurement.

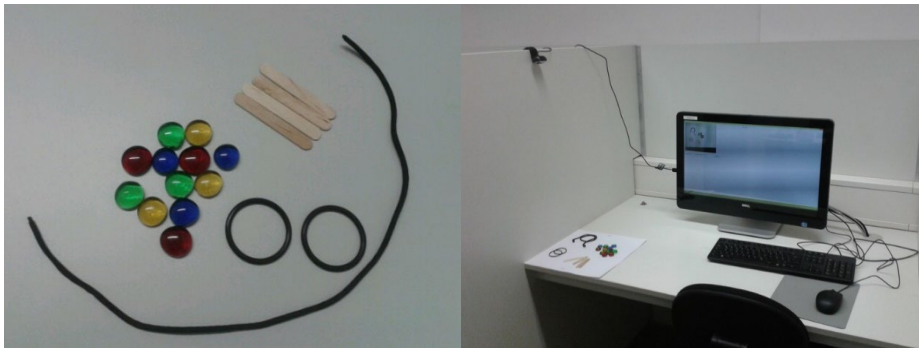
2 Lab Experiment

2.1 Design and Procedure (lab)

In WIT, ideators illustrate single words using a set of materials. The set of materials provided for each ideator consists of one string, two O-rings, four wooden sticks, and 12 colored glass pebbles (see Figure 1, left panel). Ideators can use some or all of these materials to illustrate words (see Figure 2 for example illustrations). Ideators do not receive a list or any specifications of words they illustrate. WIT requires ideators to come up with an underlying idea (the word they want to illustrate) and ideators have to provide effort to realize this underlying idea. Thus, WIT captures an important aspect that distinguishes idea generation from routine tasks, namely the fact that both effort provision and the underlying idea are relevant to success. In our experiment, ideators can illustrate as many words as they want within a predetermined time frame of 20 minutes. After finalizing an illustration, ideators are instructed to take a picture using a pre-installed camera and to type in the illustrated word. We provide a detailed description of the procedure in Appendix C.

All ideators are instructed to generate as many ideas as possible that can be identified by customers (quality) and that are original (originality). The design mimics a context of product innovation. Ideators receive clear instructions with respect to customers' needs (i.e., the need for ideas to be identifiable) and are asked to generate original ideas to meet customers' needs. Obviously, ideas generated in this experiment are not in all dimensions comparable to patentable innovations, but the design captures some of the important aspects. For example, quality and originality are requirements for patentability. Thus, using

WIT, we can quantify performance dimensions that are typically relevant in the context of innovation (see Table 1 for an overview of the relevant outcome measures of WIT).



Note: The left picture shows the available material that ideators can use to generate ideas. The right picture shows the experimental setup in the lab.

Figure 1: Set of materials and experimental setup (lab)

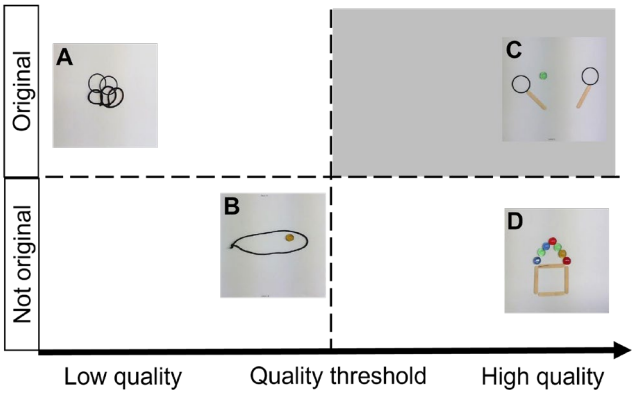
To measure the number of ideas we only consider valid ideas generated by an ideator. Before working on the task, ideators receive clear instructions on the regulations for valid words. Ideators are informed that they should illustrate single words (e.g., "tree" or "face"), while phrases consisting of more than one word (e.g., "tree in the woods," "happy face") are not valid. Ideators are further informed that they can illustrate each word only once. Therefore, any duplicate illustration of the same word (e.g., illustrating the word "house" a second time) is invalid.[4] The predefined rules further prohibit the use or illustration of any symbol found on the keyboard (e.g., "→", "8", "b", "@", ">", "+"). For example, displaying "7" to illustrate the word "seven" is considered to be a violation of the predefined rules. See Appendix C for experimental instructions. Only valid ideas are considered for payment and are used in this study to determine the relevant measures.

To measure the quality of ideas we incentivize customers in an additional survey to identify the illustrated word. These customers are provided with the illustration but not the word that is illustrated. Customers receive €0.10 for each correctly identified word.[5] Thus,

[4] In these cases, the first illustration of the word was counted as a valid illustration.

[5] When eliciting the quality, we did not account for synonyms since we explicitly informed ideators in the lab experiment and customers in the online experiment that only exact matches of the illustrated word would be considered for payment. Spelling errors were not corrected. The experiment was run in German. Special characters ä, ö, ü and ß were standardized to a, o, u and ss, respectively. Capitalization of letters was not taken into account.

customers' utility increases in the quality of the illustration. Each illustration was presented to at least 10 customers, and each customer was asked to identify a random sample of 50 illustrations. We derive the quality of an idea as the fraction of customers who correctly identify the illustrated word. For instance, an idea for which 10 out of 10 customers identify the illustrated word has the maximum quality of 1. An idea where only 1 out of 10 customers identifies the illustrated word has a quality of 0.1, while the quality of an idea which no customer could correctly identify has the minimum quality of 0. See Figure 2 for examples of ideas of high quality (C and D) or low quality (A and B). To determine the avg. quality, we estimate the average over all ideas generated by an ideator (see Table 1 for a summary of the relevant measures).



Note: A illustrates the word 'olympics' (quality: 0.1, originality: 1), B illustrates the word 'fish' (quality: 0.4, originality: 0), C illustrates the word 'tennis' (quality: 0.9, originality: 1) and D the word 'house' (quality: 1, originality: 0). All ideas in the area shaded in gray are considered to be innovative ideas.

Figure 2: Examples of ideas by quality and originality (lab)

We measure originality as the statistical infrequency of an illustrated word within 100 randomly chosen ideas.[6] Originality equals 0 when the illustrated word is among these 100 randomly chosen ideas, and 1 if it is not. See Figure 2 for examples of original ideas (A and C), and ideas that are not original (B and D). Our measure of avg. originality refers to the average of originality over all ideas generated by an ideator.

The relevant outcome measure in our design is the number of innovative ideas. We consider ideas that are both original and of high quality as innovative. Specifically, we define an idea as innovative when it is among the best quartile (75% value percentile) of the

[6] We draw reference ideas from the entire sample, excluding illustrations of the respective ideator.

product of original and quality. Due to the binary nature of the originality measure an idea has to be original in order to be innovative. Additionally, an idea has to meet a certain quality threshold to be considered as innovative. Thus, original ideas of high quality are considered to be innovative ideas. These innovative ideas are illustrated in the shaded area of Figure 2. In the Appendix, we provide robustness checks for a number of alternative specifications of innovative ideas (see Table A1 in Appendix A).

Table 1: Summary of relevant measures in WIT

Aggregate measures	
No. of innovative ideas	Number of ideas by an ideator that score in the top quartile with respect to the product of quality and originality.
No. of ideas	Number of valid ideas per ideator.
Avg. quality	Average quality across all ideas by an ideator.
Avg. originality	Average of the indicator for originality across all ideas generated by an ideator. For each idea i originality$_i$ is defined as $originality_i = \begin{cases} 1 \; if \; i \; is \; not \; among \; a \; sample \; of \; 100 \; randomly \; chosen \; ideas \\ 0 \; if \; i \; is \; among \; a \; sample \; of \; 100 \; randomly \; chosen \; ideas \end{cases}$.
Characteristics of recurring ideas	
Avg. idea time	Average over idea time across all recurring ideas by an ideator. Idea time is the average time across all illustrations of the same word in our sample.[7]
Avg. idea quality	Average over idea quality across all recurring ideas by an ideator. Idea quality is the average quality across all illustrations of the same word in our sample.
Effort recurring ideas	
Avg. rel. time	Avg. rel. time is the average over rel. time for all recurring ideas (j) generated by an ideator (k). The relative time of an idea *j* illustrated by ideator *k* is defined as $rel.time_{j\,k} = \frac{(time_{jk} - idea\ time_j)}{idea\ time_j}$.
Avg. rel. quality	Avg rel. quality is the average over rel. quality for all recurring ideas (j) generated by an ideator (k). The relative quality of an idea j illustrated by ideator k is defined as $rel.quality_{j\,k} = \frac{(quality_{jk} - idea\ quality_j)}{idea\ quality_j}$.

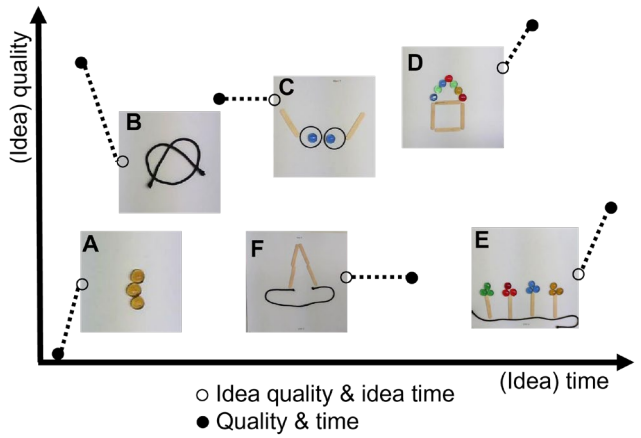
[7] When referring to the sample, we always refer to the sample of the specific experiment, that is, the lab sample for the lab experiment and the online sample for the online experiment.

Incentives may have an effect on both the underlying idea and effort provision. To distinguish between these two effects, we analyze performance for ideas that are generated multiple times within our sample (recurring ideas). In our sample, 2,216 of 2,632 ideas are illustrations of recurring ideas, i.e., illustrations of words that other ideators in the sample also illustrated. On average, each recurring idea is illustrated by roughly eight distinct ideators. In total, we find that ideators in our sample illustrate 261 words that are recurring in our sample and 416 words that are illustrated only once within our sample (Table 5 provides summary statistics on recurring ideas).[8]

For these recurring ideas, we analyze the time needed to illustrate the idea and the quality. We refer to the average time that all ideators generating the same word require to generate an illustration of this specific word as the idea time. We focus on idea time as this is directly related to the number of ideas generated by an individual. We refer to the average quality of all illustrations of the same word in our sample as the idea quality. Idea quality and idea time are measures for the characteristics of the underlying idea. We find that the quality potential of ideas varies strongly depending on the underlying idea. For example, the idea quality of a house is 0.89, the idea quality of pretzels is 0.60 (compare B and C in Figure 3). The idea time of a banana is 30 seconds, while the idea time of flowers is more than twice as much (76 seconds). See Figure 3 for some examples of recurring ideas and the idea quality and idea time of these recurring ideas. We derive avg. idea quality and avg. idea time as the averages over the idea quality and the idea time of all recurring ideas illustrated by an ideator (see Table 1 for details on how we derive these measures). Comparing avg. idea quality and avg. idea time between treatments provides us with insights with respect to systematic differences in the characteristics of the underlying ideas. Does the treatment have an effect on how time-consuming the underlying ideas an ideator illustrates are? Does the treatment have an effect on the quality potential of the underlying ideas an ideator chooses to illustrate?

[8] Due to the measures' definitions, we can only include those individuals that illustrate at least one idea that other ideators have also illustrated in our sample. Thus, our analysis in this section is restricted to a subsample of ideators who illustrate at least one recurring idea. Specifically, we have to exclude one observation from No PR when analyzing recurring ideas because this ideator only illustrated words that occurred only once within our sample.

It is not only the underlying ideas that may be relevant, but also the effort an ideator invests when illustrating them. For instance, the idea quality of a house is 0.88 and the idea time is 66 seconds. On average, nearly 90% of customers identified houses (idea quality = 0.88) and ideators spent a bit more than a minute illustrating houses (idea time = 66 seconds). For separated houses, the quality and the time vary quite strongly. To demonstrate this variation, we display some examples of illustrated houses and the corresponding times and qualities in Figure 4. Illustrations A and E in Figure 4 have the same quality, while the illustration of A is much quicker (28 seconds) compared to E (154 seconds). To capture this variation, we calculate the rel. time and the rel. quality of the recurring ideas per ideator. The rel. time is defined as the difference between the time the relevant ideator needs to illustrate a specific recurring idea and the idea time of this recurring idea, divided by the idea time of this recurring idea, i.e., the relative time of an idea j illustrated by ideator k is defined as $rel.time_{j\,k} = \frac{(time_{jk} - idea\ time_j)}{idea\ time_j}$. We determine the avg. rel. time as the average of the rel. time over all recurring ideas that a specific ideator illustrated.

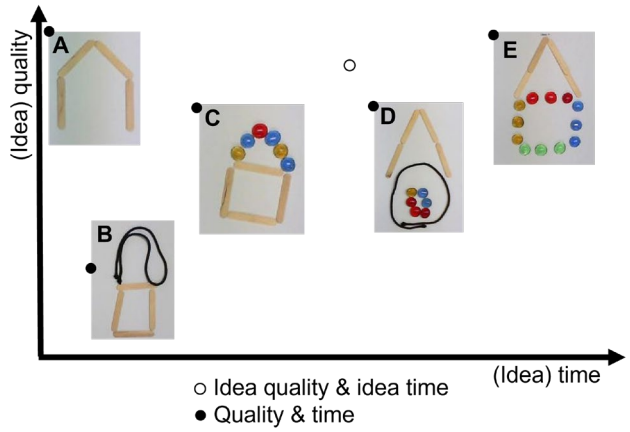


Note: This figure shows examples of different types of ideas that are typical for the respective idea quality and idea time. The figure is not drawn to scale. A illustrates the word 'banana' (idea quality=0.1, idea time=30, quality=0, time=25), B illustrates the word 'pretzel' (idea quality=0.6, idea time=38, quality=0.8, time=13), C illustrates the word 'glasses' (idea quality=0.8, idea time=65, quality=0.8, time=50), D illustrates the word 'house' (idea quality=0.88, idea time= 66, quality=1, time=86), E illustrates the word 'flowers' (idea quality=0.3, idea time=76, quality=0.4, time=78), and F illustrates the word 'hat' (idea quality=0.3, idea time=55, quality=0.3, time=96). (Idea) time is reported in seconds.

Figure 3: Examples of different types of ideas (lab)

Analogously, we derive the rel. quality as the difference between the quality achieved by an ideator in illustrating a specific idea and the idea quality of this idea divided by the idea

quality of the specific recurring idea, i.e., we define the relative quality of an idea *j* illustrated by ideator *k* is defined as $rel.quality_{j\,k} = \frac{(quality_{jk} - idea\ quality_j)}{idea\ quality_j}$. We determine the avg. rel. quality as the average of the rel. quality over all recurring ideas that a specific ideator illustrated. We ascribe treatment differences in avg. rel. quality and avg. rel. time to differences in effort provision.[9]



Note: These are examples of ideas that illustrate the word 'house' but differ in their quality and time. The figure demonstrates the variation in quality and time for illustrations of the same word. The figure is not drawn to scale. For A, quality=1, time=28, rel. quality=0.1 and rel. time=−0.57, for B, quality=0.3, time=33, rel. quality=−0.7 and rel. time=−0.5, for C, quality=0.7, time=49, rel. quality=−0.2 and rel. time=−0.3, for D, quality=0.7, time=108, rel. quality=−0.2 and rel. time=0.64 and for E, quality=1, time=154, rel. quality=0.1 and rel. time=1.3. The idea quality of a house is 0.9 and the idea time is 65. (Idea) time is reported in seconds.

Figure 4: Variations in quality and time of illustrations of the word house (lab)

In a between-subjects design, we vary whether or not PR incentives are implemented and the alignment of PR incentives. Table 2 summarizes the treatments of the lab experiment. In No PR, all ideators receive a fixed payment of €10, independent of their performance. In Unweighted PR, ideators are paid based on the number of ideas. For each valid idea, they receive €0.60, regardless of the quality or originality of the idea. By comparing No PR to Unweighted PR, we obtain insights into the effect of unweighted PR incentives on performance.

[9] On an individual level. Both measures may also be related to ideator-specific characteristics such as ability. Given that we randomize treatment assignments, such individual specific differences should not be relevant in the analysis of treatment differences.

Table 2: Treatments (lab)

Treatment	Payment	Amount	Ideators
No PR	Fixed payment	€10	32
Unweighted PR	Number of ideas	€0.60 per idea	31
Quality-Weighted PR	Number of customers who correctly identify an idea	€0.10 per correct identification of each idea per customer	30
Originality-Weighted PR	Number of original ideas (unique in a group of four)	€0.85 per unique idea	32

Note: The table summarizes the conducted treatments for the lab experiment, where treatment refers to the name of the treatment, payment shows the payoff-relevant performance measure for the ideator, amount the amount they were paid, and ideators shows the number of participants.

In the lab experiment, we study PR incentives that are less lenient. Specifically, in the lab experiment, incentives are more closely but still imperfectly aligned with the desirable outcome (i.e., with the number of innovative ideas generated). Such situations often occur in practice where some aspects of innovative ideas may be easier to quantify than others. To test for the effect of aligment, we introduce two additional treatments. In Quality-Weighted PR, the PR for each idea is contingent on its quality, i.e., the number of customers who correctly identify the illustrated word. Ideators in Quality-Weighted PR receive €0.10 for each out of 10 customers who correctly identifies the illustrated word of their idea. In Quality-Weighted PR, the PR is paid regardless of the originality of the illustration.

In Originality-Weighted PR, ideators' payment is contingent on the number of ideas that are unique within a group of four ideators.[10] For each idea based on a unique word, the ideator receives €0.85.

We calibrated incentives for the three treatments based on the performance in No PR. That is, we set the PR such that the average payment would have been equal to €10 (No PR) in each of the treatments given the performance in No PR. In this approach, we hold the overall incentives for a given performance level constant, naturally implying that the payment may vary between treatments. In the online experiment, we use the contrary approach holding the final payment constant. Using this approach, we can replicate our main findings

[10] We used uniqueness within a group of four instead of originality as an incentive measure because of procedural reasons in running the experiment. We were able to calculate uniqueness after each session and thus ensure timely payment of ideators. In Appendix A we provide an analysis based on the originality measure used to incentivize ideators and show that the results are robust to an analysis based on this measure.

indicating that our results are robust to different procedures of calibrating incentives. The experiment was conducted at the Cologne Laboratory for Economic Research. Ideators were recruited with the online recruiting system ORSEE (Greiner, 2004). The level of incentives is in line with the typical earnings in an experimental study at this location.

We ran eight sessions in May 2014 with two sessions for each treatment. Ideators were randomly seated in separated cubicles in the lab. They were given written instructions about the task, which were then read aloud by the experimenter. After the experimenter had answered all questions in private, the set of materials was handed to the ideators. All illustrations had to be placed within a designated area on the desks. We told ideators to place all materials that were not relevant for the illustration outside this area. Additionally, ideators were instructed not to use any materials other than those provided by the experimenter. Once an ideator generated an illustration, she pressed a button on the computer screen so that the software would automatically take a picture of the illustration within the designated area. If ideators were satisfied with the picture, they were asked to type in the word that they illustrated and would then proceed with their next illustration. If they were not satisfied, ideators could take another picture before proceeding. Figure 1 (right side) illustrates a cubicle in the lab, including the designated area in which ideators provided illustrations and the camera taking the pictures. As soon as the 20-minute working time was over, the experimental software automatically stopped and initiated a questionnaire with some general demographic questions.

Our sample consists of 125 ideators in the lab experiment.[11] On average, each session lasted 40 minutes, and the average payoff was €14.43 (€12.50 in No PR, €17.87 in Unweighted PR, €12.94 in Quality-Weighted PR and €14.83 in Originality-Weighted PR). The final payoff for each ideator consisted of the money earned during the experiment and a standard show-up fee of €2.50. In all treatments, the money was paid out two weeks after the experiment, and ideators could choose whether they preferred to collect the money in cash at the university or have it transferred directly to their bank account.

[11] We find a slight but insignificant imbalance in gender composition (Chi^2-test: p=0.11). We can show that the main results of this paper are robust when controlling for gender. Nevertheless, to avoid differences in the gender composition, we recruited a gender balanced sample for the online experiment.

The online survey to elicit quality was conducted two weeks after the lab experiment using the software SoSciExperiment (Leiner, 2014). See Appendix C for details on the survey. Customers did not take part in any previous related experiments and were blind to treatments. The online experiment lasted about 20 minutes, and the average earnings per customer were €4.50, including a €2.00 show-up fee, where customers could choose between collecting the money in cash or a bank transfer.

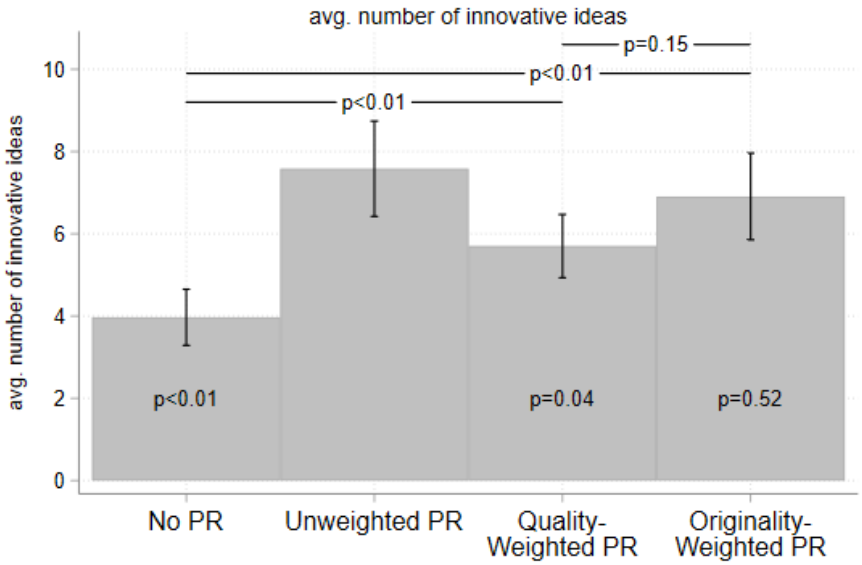
2.2 Results

The main outcome variable in our design is the number of innovative ideas per ideator. This measure mimics the number of ideas generated by an ideator that can potentially lead to innovation. In Figure 5, we illustrate the avg. number of innovative ideas per ideator by treatment. Compared to No PR, both weighted and unweighted PR incentives lead to a substantial increase in the avg. number of innovative ideas (pairwise MWU-test: $p<0.01$). Thus, we find that PR incentives have a positive effect on the avg. number of innovative ideas. Comparing weighted and unweighted PR incentives, we find that a quality weight significantly decreases the avg. number of innovative ideas (MWU-test: $p=0.04$), while an originality weight does not have a significant effect (MWU-test: $p=0.52$). Using a wide range of alternative specifications of innovative ideas, we can show that the positive effect of PR incentives on the number of innovative ideas is very robust. Qualitatively, we can also replicate the negative impact of a quality weight, depending on the specification of innovative ideas. This effect does not always meet conventional levels of significance. See Table A1 in Appendix A for an overview of the alternative innovation measures applied.

Finding 1 (lab): PR incentives lead to an increase in the number of innovative ideas. This positive effect is mitigated by a quality weight.

To understand the source of this positive effect of incentives on the avg. number of innovative ideas, we analyze performance in the different performance dimensions, i.e., no. of ideas, avg. quality, and avg. originality. See Table 3 for summary statistics on the avg. quality and the avg. originality per ideator and Appendix A (see Table A3 and Figure A1-Figure A5) for further descriptives regarding the separate performance dimensions.

We find a significant increase in the number of ideas in Unweighted PR compared to No PR (MWU-test: p<0.01). Analyzing the effect of a stronger alignment of incentives, we observe that, compared to Unweighted PR, a quality weight significantly decreases the number of ideas (MWU-test: p<0.01). Due to the quality weighting, the number of ideas in Quality-Weighted PR is reduced to the point where it is no longer statistically different from No PR (MWU-test: p=0.30). We can show that an originality weighting does not affect the number of ideas compared to Unweighted PR (MWU-test: p=0.76).



Note: Graph displays the avg. no. of innovative ideas by ideators in the respective treatment. P-values indicated within the bars indicate the results of pairwise MWU-tests comparing Unweighted PR to other treatments. P-values indicated above bars refer to pairwise comparisons of the respective PR treatments. Error bars indicate 90% confidence intervals.

Figure 5: Number of innovative ideas by treatment (lab)

We find that adding a quality weight increases the avg. quality of ideas, while unweighted PR incentives or originality-weighted PR incentives do not have a significant effect on the avg. quality of ideas. Compared to No PR, avg. quality slightly but insignificantly decreases in Unweighted PR (MWU-test: p=0.38). Adding a quality weight significantly increases the avg. quality of ideas in Quality-Weighted PR compared to Unweighted PR (MWU-test: p=0.01). Adding an originality weight does not have a significant effect on the avg. quality of ideas compared to Unweighted PR (MWU-test: p=0.43). With respect to avg. originality, we find no significant treatment differences (pairwise MWU-tests: p≥0.19).

Table 3: Different dimensions of performance (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-Weighted PR
No of ideas	16.59***	24.90	18.13***	24.53 n.s.
	(8.17)	(8.04)	(5.32)	(9.58)
Avg. quality	0.50 n.s.	0.46	0.55**	0.48 n.s.
	(0.18)	(0.15)	(0.13)	(0.14)
Avg. originality	0.51 n.s.	0.56	0.51 n.s.	0.53 n.s.
	(0.20)	(0.17)	(0.13)	(0.17)
Ideators	32	31	30	32

Note: We provide means and standard deviations in parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where ***p<0.01, **p<0.05, *p<0.10 and n.s. p>0.10.

In regression analyses (see Table A5-Table A7 in Appendix A), we can show that the positive effect of unweighted PR incentives on the number of ideas seems at most marginally driven by distortions, but rather due to an increase in effort or a change in the types of ideas. Analyzing the negative effect of a quality weight on the number of ideas, we can show that about a quarter of the effect can be explained through distortions of effort, where ideators provide fewer but higher quality ideas. However, we observe a large and significant negative impact of a quality weight on the number of ideas even after controlling for avg. quality and avg. originality. We can show that the entire gain in quality due to a quality weight seems to be due to distortions.

Summarizing, we find that Unweighted PR incentives lead to an increase in the number of ideas generated per ideator, while they have no significant effect on the avg. quality or avg. originality of ideas. With respect to alignment, we find that a quality weight mitigates the positive effect of PR incentives on the number of ideas but increases the avg. quality compared to unweighted PR incentives. In contrast, an originality weight does not have a significant effect on performance.

Finding 2 (lab): PR incentives lead to an increase in the number of ideas. A quality weight mitigates this effect and some of the decrease in the number of ideas due to a quality weight can be explained by distortions of effort.

How relevant are the above-discussed performance differences in explaining the effect of PR incentives on the number of innovative ideas? In Table 4, we provide results from OLS regressions with the number of innovative ideas as a dependent variable. In the regression model, Unweighted PR serves as reference category, while No PR, Quality-Weighted PR and

Originality-Weighted PR are dummy variables that equal 1 for the respective treatment. In the model presented in Col. (1), we replicate the earlier finding that unweighted PR incentives have a positive effect on the number of innovative ideas (negative coefficient for No PR). Compared to Unweighted PR, we find that without PR incentives, ideators illustrate 3.6 fewer innovative ideas in No PR. We also replicate the negative effect of a quality weight. Compared to Unweighted PR, ideators in Quality-Weighted PR illustrate on average 1.9 fewer innovative ideas. As in the non-parametric analysis provided above, we do not find a significant difference in the number of innovative ideas comparing Unweighted PR and Originality-Weighted PR. In the model presented in Col. (2), we control for the number of ideas. Adding this control substantially closes the gap between Unweighted PR and No PR or Quality-Weighted PR by more than 50% and more than 80%, respectively. It seems that an increase in the number of ideas is an important driver of the positive effect of PR incentives and the decrease in number of ideas seems to be a main driver of the negative effect of a quality weight. In the model presented in Col. (3), we control for avg. originality and avg. quality. Adding these controls does not have a substantial impact on the treatment coefficients.[12]

In the model presented in Col. (4), we control for interactions between the different dimensions of performance. We find that adding these controls has a significant explanatory power for understanding the effect of unweighted PR incentives on the number of innovative ideas. Thus, incentives have an impact not only on the average performance in the separate dimensions, but also on the interactions between these performance dimensions. In fact, we find that the correlation between quality and originality is less negative in Unweighted PR compared to No PR (MWU-test: p=0.05, see Table A3 in Appendix A for details). Thus, it seems that under PR incentives, ideators are less likely to specialize in one of the dimensions of performance. We can qualitatively replicate this finding when focusing on obvious ideas, i.e., ideas that are of high quality but that many individuals come up with (specialization in quality) and bad ideas, i.e., original ideas that are not identifiable (specialization in originality). See Table A2 in Appendix A for summary statistics on obvious and bad ideas.

[12] We observe a slightly larger treatment effect of a quality weight as compared to a model that just controls for the no. of ideas. The intuition behind this finding is that part of the observed decrease in the no. of ideas due to a quality weight is due to a distortion of effort provision between the different dimensions of performance.

Finding 3 (lab): The positive effect of PR incentives on the number of innovative ideas is mainly due to an increase in the number of ideas.

Table 4: Regression number of innovative ideas (lab)

	DV: Number of innovative ideas			
	(1)	(2)	(3)	(4)
No PR	−3.612***	−1.698**	−1.456***	−0.872**
	(0.793)	(0.659)	(0.420)	(0.348)
Quality-Weighted PR	−1.881**	−0.300	−0.983**	−0.749*
	(0.819)	(0.697)	(0.475)	(0.397)
Originality-Weighted PR	−0.674	−0.482	−0.502	−0.706**
	(0.923)	(0.760)	(0.505)	(0.351)
Number of ideas		0.217***	0.238***	−0.809***
		(0.042)	(0.035)	(0.146)
Avg. Quality			17.016***	−15.586***
			(1.630)	(3.482)
Avg. Originality			13.106***	−10.877***
			(1.366)	(3.304)
Avg. quality x Avg. originality				15.369***
				(3.761)
Number of ideas x Avg. quality				1.247***
				(0.151)
Number of ideas x Avg. originality				0.895***
				(0.149)
Constant	7.581***	2.067*	−13.635***	10.562***
	(0.683)	(1.052)	(1.594)	(2.880)
Observations	125	125	125	125
R-squared	0.171	0.458	0.738	0.862

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1.

Above, we see that a main driver of the observed treatment effects is an increase in the number of ideas. As the total working time is fixed by design, a raise in the number of ideas is equivalent to a decrease in the average time per idea. The question arising is whether this increase in the number of ideas is due to an increase in effort provision, i.e., due to a more efficient use of the allocated time, or due to differences in the types of ideas, i.e., by choosing to illustrate ideas which require less time in the realization. To analyze the effect of incentives on the types of ideas and on effort provision in more detail we analyze performance focusing on recurring ideas. We can show that the subsample of ideators used

when analyzing recurring ideas is not systematically different from the full sample (see Table A8–Table A11 in Appendix A).[13]

Table 5 provides summary statistics on recurring ideas. Focusing on idea time, we find that PR incentives shift the types of ideas that are generated toward less time-intensive ideas. This effect is partially mitigated by a quality weight. We find that the avg. idea time in Unweighted PR is significantly lower compared to No PR (MWU-test: p=0.07). Under quality-weighting, the idea time is slightly but insignificantly larger compared to Unweighted PR (MWU-test: p=0.17). As a result, we observe no significant difference in the avg. idea time comparing the Quality-Weighted PR to No PR (MWU-test: p=0.17). An originality weight has no effect on avg. idea time (MWU-test: p=0.72). We find no significant treatment effects with respect to the avg. idea quality (pairwise MWU-tests: p>0.15). Thus, incentives seem to lead to a shift toward illustrating ideas that are less time intensive without leading to substantial changes in the quality potential of the underlying idea.

Finding 4 (lab): PR incentives lead to a shift in the underlying ideas toward less time-intensive ideas. This effect is mitigated when a quality weight is introduced.

To analyze the effect of incentives on effort provision, we focus on the avg. rel. time and the avg. rel. quality.[14] We find a significant negative effect of Unweighted PR on avg. rel. time compared to No PR (MWU-test: p<0.01). This effect is partly diminished due to a quality weight (MWU-test: p<0.01), while an originality weight does not have a significant effect on avg. rel. time (MWU-test: p=0.52) compared to PR incentives. The treatment differences in avg. rel. time are accompanied by changes in the avg. rel. quality. Thus, effects may partially be due to distortion in effort and partially be due to differences in the level of effort. However, we observe a positive effect of an unweighted PR and a negative effect of a quality weighted

[13] In Appendix A we provide further evidence of the fact that the increase is not only due to the increase in no. of innovative ideas, but also due to an increase in the likelihood of generating an innovative idea using random effects regressions on the idea level.

[14] Note that the distribution of avg. rel. time is skewed. Positive values are much more dispersed (it is easier to take longer than the average) while negative values are bounded (one cannot take less than 0 seconds to illustrate a word and it is harder to do so than to spend more time on the illustration). Therefore, the means by treatment do not add up to zero.

PR on the avg. rel. time even when controlling for avg. rel. quality (see Table A4 in Appendix A for detailed results).

Table 5: Summary statistics recurring ideas (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-Weighted PR
Avg. idea time	55.10*	52.22	54.51 n.s.	52.61 n.s.
	(6.87)	(4.58)	(5.87)	(5.44)
Avg. idea quality	0.577 n.s.	0.55	0.58 n.s.	0.54 n.s.
	(0.10)	(0.09)	(0.09)	(0.10)
Avg. rel. time	0.48***	−0.06	0.18***	0.03 n.s.
	(1.22)	(0.27)	(0.22)	(0.45)
Avg. rel. quality	0.06***	−0.08	0.05**	0.02*
	(0.18)	(0.18)	(0.19)	(0.18)
Ideators illustrating recurring ideas	31	31	30	32

Note: We provide means and standard deviations in parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where *** p<0.01, ** p<0.05, * p<0.1 and n.s. p≥0.10. All results for recurring ideas are based on a limited sample of those ideators who illustrate at least one recurring ideas. See Table 1 for the exact definition of the reported measures.

Thus, we can show that unweighted PR incentives lead to an increase in effort provision and a shift toward illustrating less time-intensive ideas. In line with an interpretation of adverse effects of control, a quality weight mitigates this positive effect of PR incentives on effort provision. Interestingly, an originality weight does not have a significant effect on effort provision. We conjecture that an originality weight (which targets only the underlying idea but not effort provision) may not be perceived as an act of control. However, it is beyond the scope of this paper to analyze the cause of the differences between a quality weight and an originality weight in more detail.

Finding 5 (lab) PR incentives lead to an increase in effort provision. This effect is mitigated when a quality weight is introduced.

Above, we could show that incentives have an effect on the types of recurring ideas. Next, we want to explore whether incentives also have an impact on the types of innovative ideas that are generated. See Table 6 for some characteristics of innovative ideas. Note that in the analysis of innovative ideas, we can only include those ideators who illustrate at least one innovative idea.[15] In line with our previous analysis, we find that ideators invest significantly

[15] We can show that there are no treatment differences in the fraction of ideators who have at least one innovative idea (Chi^2-test: p=0.90).

less time to generating innovative ideas in Unweighted PR in comparison to No PR (MWU-test: $p<0.01$). This effect is mitigated by a quality weight (MWU-test: $p<0.01$), while an originality weight does not have a significant effect (MWU-test: $p=0.62$). We find no significant treatment differences in the avg. quality of innovative ideas (pairwise MWU-tests: $p>0.27$). Given the originality requirement for innovative ideas, we cannot disentangle the relevance of differences in effort provision and differences in the types of ideas when analyzing innovative ideas. Applying the insights from recurring ideas, however, it is plausible to assume that the shift in time invested in generating innovative ideas is due to both a shift in the types of innovative ideas and a shift in effort provision.

Table 6: Characteristics of innovative ideas (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-Weighted PR
Avg. time	68.70***	47.73	63.37***	52.30 n.s.
	(32.06)	(17.71)	(24.75)	24.01
Avg. quality	0.53 n.s.	0.54	0.57 n.s.	0.57 n.s.
	(0.15)	(0.10)	(0.16)	(0.11)
Avg. complexity	7.27 n.s.	6.38	6.31 n.s.	6.64 n.s.
	(3.18)	(2.47)	(2.13)	(2.78)
Ideators generating innovative ideas	30	30	29	31

Note: We provide means and standard deviations in parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where *** $p<0.01$, ** $p<0.05$, * $p<0.1$ and n.s. $p\geq0.10$. These results are based on a limited sample of those ideators who generated at least one innovative idea. See Table 1 for the exact definition of the reported measures.

Finally, we focus on the avg. complexity of innovative ideas. We define avg. complexity as the number of materials used in generating an idea. We argue that incentives may have an effect on the avg. complexity of ideas through spillover effects. We find that PR incentives (weighted or unweighted) lead to a slight but insignificant increase in the avg. complexity of innovative ideas. Analyzing recurring ideas (see Table A12, Appendix A), our results suggest that this may be due to both differences in the underlying ideas and shifts in the number of materials used when illustrating the same idea.

Finding 6 (lab): Unweighted PR incentives have a slight (insignificant) effect on the types of innovative ideas, where innovative ideas tend to require less time to be illustrated and are less complex.

3 Online Experiment

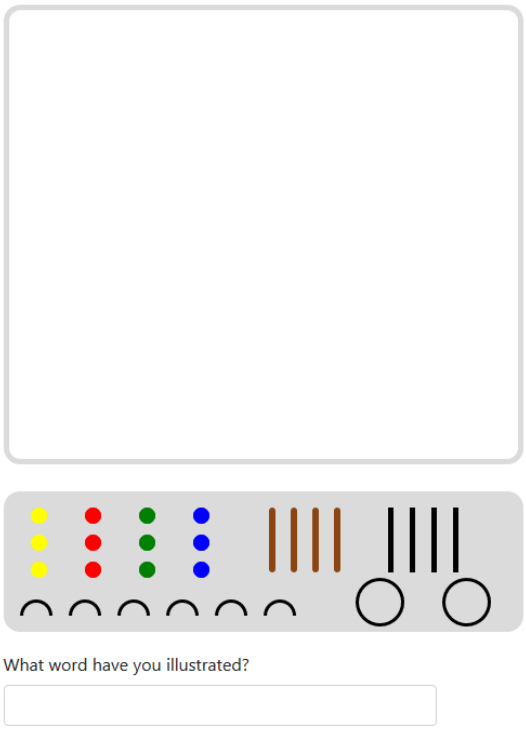
3.1 Design and Procedure (online)

For the second experiment we introduce an online version of the WIT. We provide the code and documentation for this online version in the companion materials to this paper.[16] The online version mimics the lab version in most relevant aspects. As in the lab version, ideators are provided with a set of materials, i.e., icons, and are asked to use these icons to illustrate words. The set of icons is chosen to mimic the set of materials provided in the offline version of WIT. The set of icons consists of 12 colored dots (three each in yellow, red, green, and blue), four bold brown lines, four black lines, six quarter circles in black, and two full circles in black. See Figure 6 for an overview of the icons available to ideators. As in the lab version, ideators have to come up with an underlying idea (the word they want to illustrate) and have to provide effort to realize this idea.

To illustrate words, ideators drag materials into the working area (white area with gray frame illustrated in Figure 6) and can resize, rotate, and change the layer of materials. Resizing is an additional function that was not available in the lab experiment. Thus, in the online version, we have expanded the scope of what ideators can do with the given icons. Another difference between the lab and the online version of WIT is the haptics of the task. While ideators in the lab experiment can touch materials, ideators in the online version work using a computer interface.

As in the lab version, ideators are instructed to illustrate as many ideas as possible that are of high quality and original (see Appendix D for instructions). Within a limited time frame of 20 minutes, ideators can generate as many ideas as they want. Once they have illustrated a word, ideators insert the word they illustrated in the respective text field (see lower part of Figure 6) and submit the idea. Once an idea has been submitted, materials return to the initial position and ideators can work on the next idea.

[16] The code is available on GitHub (https://github.com/nathalieroemer/WIT).

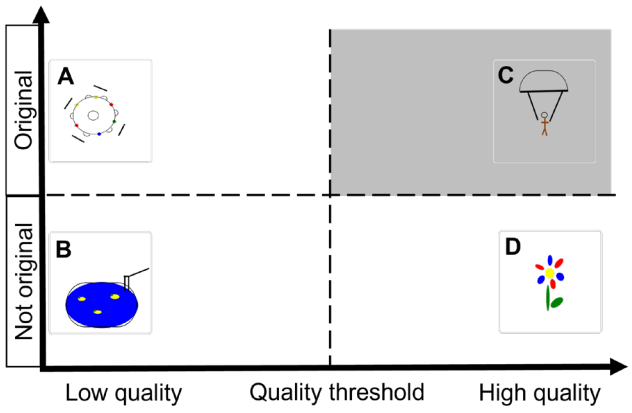


Note: They gray box includes the material that ideators can use to generate ideas in the working area (white box with gray frame). They can drag the material, resize, and rotate it, and change the layer. They indicate the word they have illustrated in the text field below.

Figure 6: Set of provided materials (online)

We quantify performance using the same measures as in the lab experiment. See Table 1 for an overview of the relevant performance measures. As in the lab version, we consider only the number of valid ideas generated by an ideator when measuring the number of ideas generated by an ideator. The requirements that an idea has to meet to be considered as valid are the same as in the lab experiment.

To measure the quality of an illustration we incentivize customers in an additional survey to identify the exact illustrated words. Customers earned £0.10 for each correct answer. Each customer was shown a random sample of 50 different illustrations. We displayed each illustration to at least 10 customers. Equivalent to the lab experiment, we compute the quality of an idea as the fraction of customers who correctly identify the illustrated word. See Figure 7 for examples of ideas of high quality (C and D) and low quality (A and B) in the online experiment.



Note: A illustrates the word 'carousel' (quality=0, originality=1), B illustrates the word 'fish' (quality=0, originality=0), C illustrates the word 'parachute' (quality=0.9, originality=1), and D illustrates the word 'flower' (quality=1, originality=0). All ideas in the area shaded in gray are considered to be innovative ideas.

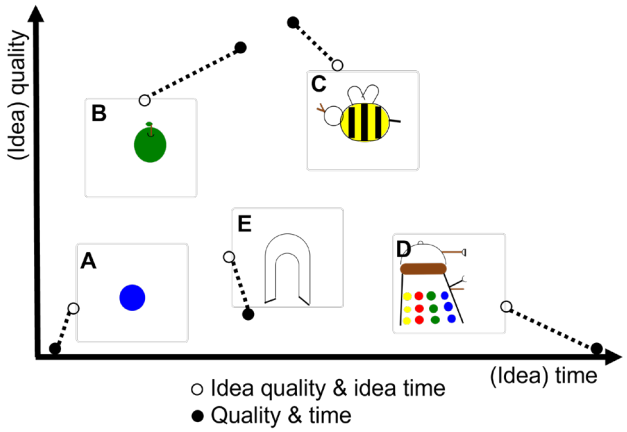
Figure 7: Examples of ideas by quality and originality (online)

To measure the originality of an idea, we again check whether the word is included in a random sample of 100 words illustrated by other ideators in the online experiment. If the word is not among this subsample, it is defined as original. See Figure 7 for examples of original ideas (A and C) and ideas that are not original (B and D) from the online experiment.

Our relevant outcome measure is again the number of innovative ideas, which we define equivalent to the lab experiment. Ideas are innovative if they are among the top quartile (75% value percentile) with respect to the product of quality and originality. For the online experiment we provide robustness checks using alternative specifications in Table A13, Appendix B.

In the online experiment, 7,239 of 7,853 ideas are illustrations of recurring ideas. In the online sample, on average, each recurring idea is illustrated by 11 distinct ideators. In total, ideators illustrate 7,239 words that are recurring and 614 words that are illustrated only once within the sample.

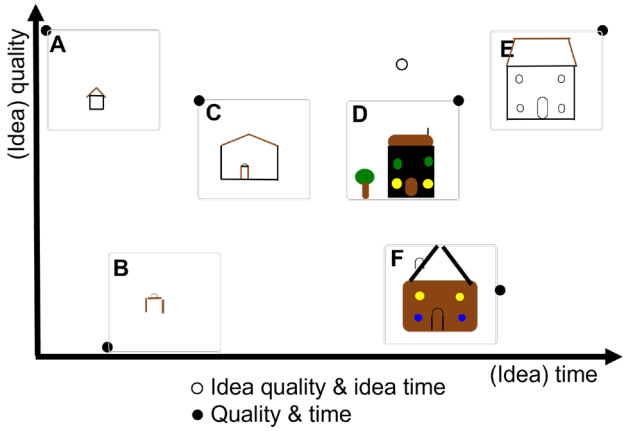
We again use idea quality and idea time to measure the characteristics of underlying ideas. We show in Figure 8 that, similar to the lab version of WIT, ideas can strongly vary with respect to how time-consuming they are (idea time) and their quality potential (idea quality).



Note: This figure shows examples of different types of ideas that are typical for the respective idea quality and idea time. The figure is not drawn to scale. A illustrates the word 'ball' (idea quality=0.17, idea time=78, quality=0, time=18), B illustrates the word 'apple' (idea quality=0.7, idea time=80, quality=0.9, time=123), C illustrates the word 'bee' (idea quality=0.9, idea time=155, quality=1, time=132), D illustrates the word 'alien' (idea quality=0.2, idea time=192, quality=0, time=424) and E illustrates the word 'arch' (idea quality=0.3, idea time=111, quality=0.2, time=126). (Idea) time is reported in seconds.

Figure 8: Examples of different types of ideas (online)

Similar to the lab experiment, we find that, depending on how much effort the ideator invests, the time and quality of the illustration of the same word can substantially vary. We provide one example for this in Figure 9, where we again show different illustrations of the word house and their corresponding time and quality. To capture this variation, we again calculate the avg. rel. time and the avg. rel. quality of recurring ideas per ideators (see Table 1) to capture treatment differences in effort provision.



Note: These are examples of ideas that illustrate the word 'house' but differ in their quality and time. The figure demonstrates the variation in quality and time for illustrations of the same word. The figure is not drawn to scale. For A, quality=1, time=35, rel. quality=0.2 and rel. time=-0.77, for B, quality=0, time=89, rel. quality=-1 and rel. time=-0.4, for C, quality=0.7,time=95, rel. quality=-0.2 and rel. time=-0.4, for D, quality=0.7, time=250, rel. quality=-0.2 and rel. time=0.62, for E quality=1, time=448, rel. quality=0.2 and rel. time=1.9 and for F quality=0.1, time=278, rel. quality=-0.9 and rel. time=0.8. The idea quality of a house is 0.9 and the idea time is 153. (Idea) time is reported in seconds.

Figure 9: Variations in quality and time of illustrations of the word house (online)

Using a between-subjects design, we randomly allocate ideators to one of three treatments (see Table 7 for an overview of the treatments).[17] In No PR, ideators received £10.00, independent of performance. In Unweighted PR, ideators received £1.00 for each valid submitted illustration, regardless of the quality or originality of these illustrations. In an additional treatment, the Innovation-Weighted PR, ideators receive £10.00 for each innovative idea, where we define the top 10% of ideas as innovative.[18] This way, ideators in the Unweighted PR and the Innovation-Weighted PR both earn in expectations £1.00 per idea. For No PR, we calibrate incentives by first sampling 100 ideators from Unweighted PR. We determined the fixed payment in No PR such that the realized payment would be equalized in both treatments based on this sample of 100 ideators.[19] To avoid distortions due to differences in the displayed bonus, the bonus displayed to ideators was 10 tokens in all treatments. However, the exchange rate for realized payment varied between treatments.

Table 7: Treatments (online)

Treatment	Payment	Amount	Ideators
No PR	fixed payment	£10.00	298
Unweighted PR	number of ideas	£1.00 per idea	299
Innovation-Weighted PR	number of innovative ideas	£10.00 per innovative idea (ideas among the best 10%)	305

Note: The table summarizes the conducted treatments for the online experiment, where treatment refers to the name of the treatment, payment shows the payoff-relevant performance measure for the ideator and the amount they were paid. Ideators shows the number of ideators (participants in the experiment) in each treatment.

After the completion of WIT, ideators of the online experiment took part in a second experiment that was explained to them only after completing WIT. In this second experiment, ideators were provided with all the ideas that they had generated in WIT and

[17] See https://osf.io/azs3b/?view_only=fff97053d8044db78ca243d20cafbe78 for the pre-registration of this experiment and https://gfew.de/ethik/d6PLgtXR for ethical approval.

[18] We used this more restrictive measure to incentivize ideators to generate ideas that are as innovative as possible. We show in Table A13, Appendix B, that results go in the same direction for the incentivized measure. However, due to the more restrictive measure and the fewer no. of ideas that qualify as innovative using it, not all of them meet levels of statistical significance. We preregistered the measure based on the 75% percentile we are using in the main paper.

[19] Thus, we use different approaches to calibrate incentives in the online compared to the lab experiment. Using these different approaches, we can test whether the main results of our study are robust to differences in the procedure of calibrating incentives.

were asked to select one of them. Ideators were incentivized to choose innovative ideas, i.e., to choose their best idea (see Appendix D for details).

The experiment was conducted in the English language on Prolific using a gender-balanced sample of UK residents.[20] We collected the data in September 2023. As pre-registered, we excluded ideators who had participated in other experiments involving WIT and ideators who did not pass our attention check and/or our trial task. Overall, ideators spent an average of 48 minutes to complete the online experiment, of which they spent 20 minutes working on generating ideas. On average, ideators earned £12.99 (£13.50 in No PR, £13.32 in Unweighted PR and £12.16 in Innovation-Weighted PR) including a £3.50 show-up fee.

For the quality elicitation, we recruited 1,681 customers from Prolific. We again recruited a gender-balanced sample of UK residents. Ideators were not eligible to participate as customers in the quality elicitation. Customers took about 15 minutes to complete the experiment and earned £3.42 on average, including a £1.00 show-up fee.

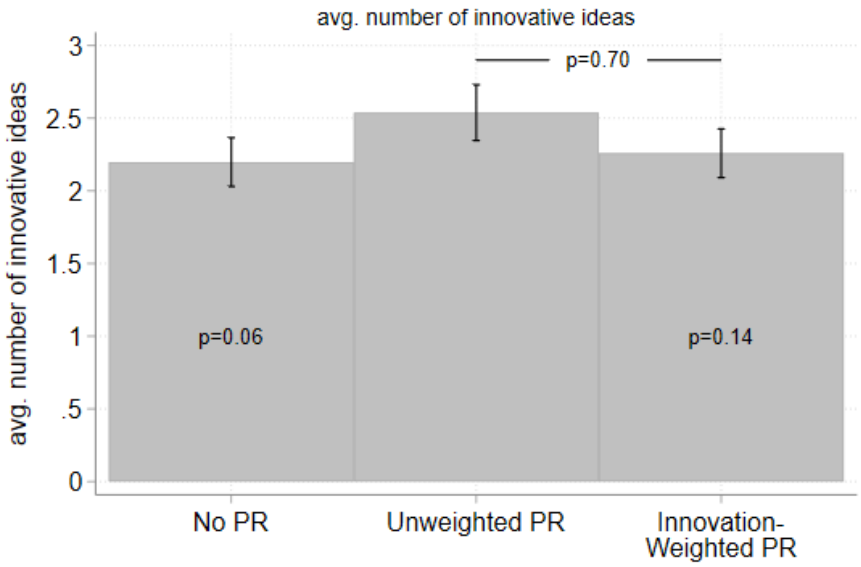
3.2 Results (online)

Analyzing the number of innovative ideas, we find that compared to No PR, Unweighted PR significantly increases the number of innovative ideas (MWU-test: p=0.07). See Figure 3 for a graphical illustration. Thus, we can replicate our Finding 1 from the lab experiment and can show that unweighted PR incentives significantly increase the number of innovative ideas.[21] Adding an innovation weight mitigates the observed effect (MWU-test: p=0.14) to the point where the number of innovative ideas in Innovation-Weighted PR is not statistically significantly different from No PR (MWU-test: p=0.70). We can show that our findings are robust to using alternative specifications of innovative ideas (see Table A13, Appendix B).

[20] Another difference between the lab and the online experiment lies in the language of data collection. The lab experiment was run in German, and the online experiment was run in the English language.

[21] While we replicate the same direction of effects, one notable difference between the lab and the online experiment is the effect size. However, given that the designs vary in multiple dimensions, we cannot and do not aim at explaining the differences in effect size. Rather, it is reassuring that we observe the same direction of effects in both experiments.

Finding 1 (online): PR incentives lead to an increase in the number of innovative ideas. This positive effect is mitigated when an innovation weight is introduced.



Note: Graph displays the average number of innovative ideas generated by the ideators in the respective treatment. P-values indicated within the bars indicate the results of pairwise MWU-tests compared to Unweighted PR. P-values indicated above bars refers to pairwise comparisons of the respective PR treatments. Error bars indicate 90% confidence intervals.

Figure 10: Number of innovative ideas by treatment (online)

Investigating different performance dimensions, we replicate our finding from the lab experiment that Unweighted PR incentives increase the number of ideas, while a stronger alignment of incentives can have adverse effects. Table 8 summarizes performance across different dimensions.[22] We find that Unweighted PR incentives lead to a significant increase in the number of ideas compared to No PR (MWU-test: $p<0.01$). Comparing the Unweighted PR to the Innovation-Weighted PR, we can show that an innovation weight mitigates the positive effect of a PR on the number of ideas (MWU-test: $p<0.01$) to the extent that the number of ideas does not differ from No PR (MWU-test: $p=0.90$). We find no treatment effects on avg. quality (pairwise MWU-tests: $p>0.81$) or avg. originality (pairwise MWU-tests: $p>0.33$). We also find no evidence of significant distortions in effort provision between the different dimensions of performance (see Table A17–Table A19 in Appendix B).

[22] In Appendix B we provide further descriptives regarding the separate performance dimensions (see Table A15 and Figure A5–Figure A8.

Table 8: Different dimensions of performance (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Number of ideas	8.15**	9.83	8.15**
	(3.78)	(4.69)	(4.01)
Avg. quality	0.48 n.s.	0.49	0.49 n.s.
	(0.20)	(0.18)	(0.19)
Avg. originality	0.66 n.s.	0.65	0.67 n.s.
	(0.23)	(0.21)	(0.21)
N Ideators	298	299	305

Note: We provide means in standard deviations in parenthesis. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments where *** p<0.01, **p<0.5, *p<0.10 and n.s. p>0.10. See Table 1 for the exact definition of the reported measures.

In an OLS regression model (Table 9), we replicate our findings from the non-parametric analysis showing that Unweighted PR incentives significantly increase the number of innovative ideas. In the provided regression model, Unweighted PR serves as a reference category and we consider the number of innovative ideas as the dependent variable. The negative coefficients in Col. (1) show that relative to Unweighted PR, ideators have fewer innovative ideas in both No PR and Innovation-Weighted PR. These effects are mitigated when controlling for the number of ideas (Col. (2) in Table 9). These findings again confirm that the increase in the number of ideas is an important driver of the positive effect of PR incentives.

Finding 2 (online): The increase in the number of innovative ideas is mainly due to an increase in the number of ideas.

Focusing on recurring ideas, we analyze differences in the types of ideas and in effort provision (see Table 10 for summary statistics).[23] We find that the avg. idea time in Unweighted PR is significantly lower compared to No PR (MWU-test: p<0.01). In Innovation-Weighted PR avg. idea time is slightly but insignificantly higher as compared to the Unweighted PR (MWU-test: p=0.16). With respect to avg. idea time, we do not observe a significant difference between No PR and Innovation-Weighted PR (MWU-test: p=0.22). We

[23] For all measures based on recurring ideas, we restrict the analysis to the subsample of ideators who illustrated at least one recurring idea. We can show that the subsample of ideators used when analyzing recurring ideas is not systematically different from the full sample of this study (see Table A8–Table A11 in Appendix B).

find no significant treatment differences with respect to average idea quality (pairwise MWU-tests: p>0.38).[24]

Finding 3 (online): PR incentives lead to a shift in the underlying ideas toward less time-intensive ideas. This effect is mitigated when an innovation weight is introduced.

Table 9: Regression number of innovative ideas (online)

	DV: Number of innovative ideas			
	(1)	(2)	(3)	(4)
No PR	−0.340**	0.057	0.131	0.102
	(0.154)	(0.134)	(0.086)	(0.064)
Innovation-Weighted PR	−0.279*	0.118	0.130	0.045
	(0.155)	(0.128)	(0.087)	(0.064)
Number of ideas		0.237***	0.257***	−0.590***
		(0.016)	(0.012)	(0.035)
Avg. quality			6.401***	−3.097***
			(0.252)	(0.408)
Avg. originality			4.177***	−2.184***
			(0.223)	(0.278)
Avg. quality x Avg. originality				4.431***
				(0.426)
Number of ideas x Avg. quality				0.908***
				(0.038)
Number of ideas x Avg. originality				0.628***
				(0.035)
Constant	2.538***	0.205	−5.862***	1.613***
	(0.116)	(0.159)	(0.285)	(0.270)
Observations	902	902	902	902
R-squared	0.006	0.292	0.691	0.832

Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1.

Equivalent to the lab experiment, we find that unweighted PR incentives lead to a decrease in the avg. rel. time, while a stronger alignment of incentives offsets this effect. Ideators in No PR need, on average, significantly more time to illustrate specific recurring ideas as compared to those in Unweighted PR (MWU-test: p<0.01). Comparing time efficiency between Innovation-Weighted PR and Unweighted PR reveals that weighting significantly reduces the gains in time efficiency through PR incentives (MWU-test: p<0.01). We find no

[24] Again, we see effects on the types of ideas reflected in the number of distinct ideas generated by treatment. The number of distinct ideas of those which are recurring is highest in No PR with 519 ideas that are repeatedly illustrated, followed by the number of distinct recurring ideas generated in the Innovation-Weighted PR, while it is lowest in No PR.

significant differences in the avg. rel. quality of recurring ideas (pairwise MWU-tests: p>0.11).

Finding 4 (online) PR incentives lead to an increase in effort provision. This effect is mitigated when an innovation weight is introduced.

Table 10: Summary statistics recurring ideas (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Avg. idea time	135.51***	131.54	133.81 n.s.
	(27.81)	(30.35)	(29.55)
Avg. idea quality	0.50 n.s.	0.51	0.51 n.s.
	(0.14)	(0.13)	(0.13)
Avg. rel. time	0.24**	−0.10	0.27***
	(0.55)	(0.54)	(0.67)
Avg. rel. quality	0.01 n.s.	0.00	0.02 n.s.
	(0.39)	(0.36)	(0.34)
N Ideators	297	299	301

Note: We provide means in standard deviations in parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments where ***p<0.01, **p<0.05, *p<0.10 and n.s. p≥0.10. All results for recurring ideas are based on a limited sample of those ideators who illustrate at least one recurring idea.

See Table 11 for an overview of some characteristics of innovative ideas. Note that our analysis in this section is based on a reduced sample of ideators, because we can only include those ideators who illustrated at least one innovative idea.[25] As for the lab experiment, we find that unweighted PR incentives lead to a significant reduction in the avg. time needed to illustrate innovative ideas (MWU-test: p<0.01). An innovation weight mitigates this effect (MWU-test: p<0.01), so that we observe no significant difference between No PR and Innovation-Weighted PR (MWU-test: p=0.89) with respect to the time needed to illustrate innovative ideas. We find no significant differences in the avg. quality of innovative ideas (pairwise MWU-tests: p>0.49).

Analyzing the avg. complexity of innovative ideas (i.e., avg. number of materials used), we find that unweighted PR incentives lead to a significant reduction in complexity compared to No PR (MWU-test: p=0.02). We find that, compared to Unweighted PR, an Innovation-Weighted PR leads to a slight but insignificant increase in the complexity of innovative ideas

[25] We find no significant treatment differences in the fraction of ideators who provide at least one innovative idea (Chi^2 test: p=0.65).

(MWU-test: p=0.41). Overall, we do not observe a significant difference in the avg. complexity of innovative ideas in Innovation-Weighted PR compared to No PR (MWU-test: p=0.13). Thus, an innovation weight seems to partially mitigate the effect of unweighted PR incentives with respect to complexity. Again, given innovative ideas, it is hard to isolate whether the shift in complexity is due to the use of more materials to illustrate the same word or due to differences in the underlying ideas. When analyzing recurring ideas (see Table A24 in Appendix B), we observe both.

Finding 5 (online) PR incentives have an effect on the types of innovative ideas, where innovative ideas require less time to illustrate and are less complex. This effect is mitigated by an innovation weight.

Table 11: Characteristics of innovative ideas (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Avg. idea time	166.53***	143.55	170.21***
	(121.91)	(97.64)	(131.37)
Avg. quality	0.75$^{n.s.}$	0.76	0.76$^{n.s.}$
	(0.12)	(0.12)	(0.13)
Avg. complexity	7.00**	6.28	6.67$^{n.s.}$
	(3.47)	(2.71)	(3.38)
N Ideators	247	256	256

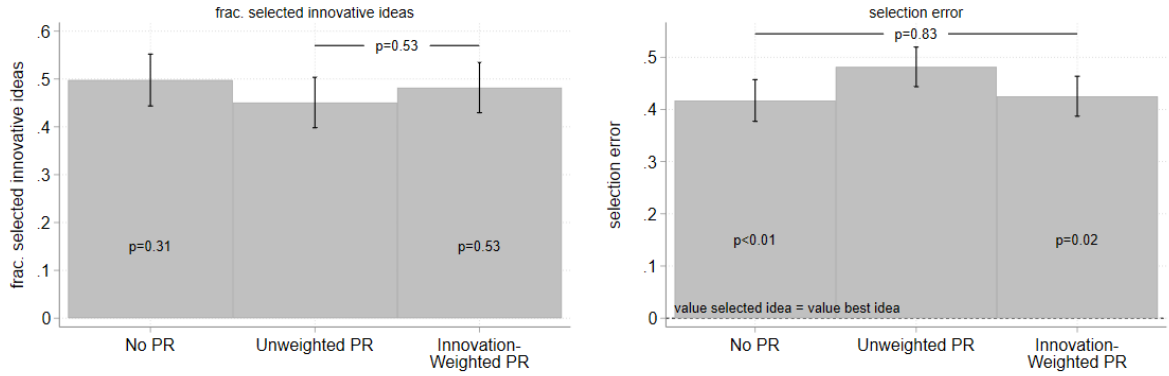
Note: We provide means and standard deviations in parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where *** p<0.01, ** p<0.05, * p<0.1. These results are based on a limited sample of those ideators who generated at least one innovative idea. For the exact definitions of the reported measures see Table 1.

To foster innovation through the generation of innovative ideas, it is crucial that ideators select and pursue these ideas. To better understand what types of ideas are pursued, we introduce a selection stage in the online experiment. In this selection stage, we incentive ideators to choose their best idea. As illustrated in Table 11, we find strong biases in the selection. Only roughly half of the ideators in our experiment who have an innovative idea also select an innovative idea. We find a slight but insignificant tendency for ideators who have an innovative idea in Unweighted PR to be less likely to select this as compared to No PR or Innovation-Weighted PR (pairwise Chi^2-test: p>0.30).

To obtain better insights into why PR incentives may have an impact on the selection of ideas, we compute a continuous measure for the selection error which we define as the difference of the value (product of quality and originality) of the best idea and of the selected

idea (see right panel of Figure 4). We find that unweighted PR incentives lead to a significantly higher selection error as compared to No PR and to Innovation-Weighted PR (pairwise MWU-test, $p<0.01$ and $p=0.02$, respectively). Using regression analysis (see Table A25 Appendix B), we can show that this increase in selection error seems to be driven by the increase in the number of ideas in Unweighted PR. It seems that the mere quantity of ideas that ideators in Unweighted PR can choose from makes it harder for them to choose their best idea.

Finding 6 (online) Due to an increase in the number of ideas, ideators in Unweighted PR perform less well in selecting their best ideas as compared to ideators in No PR or Innovation-Weighted PR.



Note: The right graph plots the fraction of ideators who select an innovative idea. The left graph plots the selection error, that is, the difference between the value (product of originality and quality) of the best idea minus the selected idea. We display values by treatment. P-values indicated within the bars indicate the results of pairwise Fisher-exact-tests for the fraction of selected ideas, and MWU-tests for the selection error, compared to Unweighted PR. P-values indicated above bars refers to pairwise comparisons of the respective PR treatments. Error bars indicate 90% confidence intervals.

Figure 11: Selection of innovative ideas (online)

4 Conclusion and Discussion

In this paper, we introduce the word illustration task (WIT). Using WIT, we can experimentally study the early phase of any innovation process, namely idea generation. The advantage of WIT is that we can derive objective measures for performance in multiple dimensions that are relevant in the context of innovation, i.e., number of ideas, quality (how well customers can identify the illustrated word), and originality (statistical infrequency of the illustrated word). Using WIT, we evaluate the effectiveness of PR incentives in an experimental idea generation context. Focusing on ideas that are recurring in our sample,

we can isolate the effect of incentives on effort provision and the effect of incentives on the underlying ideas.

We believe that WIT is a valuable tool also for future researchers seeking to understand the effect of organizational design on performance in idea generation and more general in complex tasks. WIT is easy to adapt and has therefore already been implemented in a wide range of research contexts. In this version of WIT, we measure and incentivize originality and quality. Our approach mimics a context of product innovation. However, one strength of WIT lies in its adaptability. Process innovations could, for example, be captured by asking ideators to find a new way to illustrate a given idea (word), or to illustrate a specific word using fewer materials while maintaining the quality standard. A similar approach has already been implemented by Petters & Schröder (2020). Other adaptations would be to include a measure for the aesthetic value of ideas (Grözinger et al., 2020), varying the materials that are available to ideators (Blaufus et al. 2024), or studying support for innovative ideas (Römer & Schröder, 2024).

Our results show that unweighted PR incentives are suitable to foster idea generation. In both the lab and the online experiment, unweighted PR incentives have a positive effect on the number of innovative ideas generated by ideators. We can show that this positive effect is mainly due to an increase in the number of ideas. We further observe that incentives impact both effort provision and the types of ideas. We find some indications that unweighted PR incentives lead to a shift in the types of innovative ideas, where under PR incentives ideators generate less complex ideas.

In the online experiment, we find that the positive effect of unweighted PR incentives on idea generation may be accompanied by a negative effect in the context of idea selection. The increase in the number of ideas under unweighted PR incentives makes it harder for ideators to select their best ideas.

With respect to alignment, we find evidence of both reduced effort provision and distortions arising due to a quality weight. Overall, a quality or an innovation weight mitigates the effect of unweighted PR incentives on performance. In line with an interpretation of adverse effects of control, a quality or an innovation weight leads to a decrease in effort provision. In addition, imperfectly aligned quality weighting can distort

performance by shifting ideators' effort toward illustrating fewer ideas that are, on average, of higher quality.

Interestingly, we find no effect of an originality weight on performance in WIT. Our results suggest that incentivizing, and thus controlling effort provision (quality), may have a very different effect on performance as compared to incentivizing and thus controlling types of ideas (originality). However, it may also be the case that originality is perceived as less influenceable by the ideators and thus, does not impact behavior. More research is needed to fully understand differences in incentives targeting effort provision and different types of ideas.

Our findings are highly relevant for the economic literature on complex tasks (Ederer & Manso, 2013; Butschek et al., 2019; Englmaier et al., 2024). We can show that incentives have an effect both on effort provision and on the types of ideas pursued. In line with our results, evidence from the field that incentives influence both the level of effort exerted and the nature of ideas explored. For example, the design of research funding is shown to have an impact on the likelihood of academics performing breakthrough research (Azoulay et al., 2011). Furthermore, previous research provides evidence of a link between incentives and the novelty of user-generated content (Gross, 2020; Burtch et al., 2022). Our insights are complementary to this line of research. Given the experimental setup implemented in this study, we can provide causal evidence of the effect of incentives on the types of ideas generated. Furthermore, we can analyze the underlying mechanisms in more detail.

In our design, we can show that only half of the ideators who generate an innovative idea also choose to pursue this idea. These inefficiencies in the early stages of an innovation process are typically not visible in research using outcome measures such as patents or publications to quantify innovation. However, they have real implications on innovation outcome as our results suggest that reducing biases in the very early stages of an innovative process may have strong economic implications. More research is needed to better understand the very early selection of innovative ideas.

References

Attanasi, G., Chessa, M., Gil-Gallen, S., & Llerena, P. (2021). A survey on experimental elicitation of creativity in economics. *Revue d'économie Industrielle*, *174*, 273–324. https://doi.org/10.4000/rei.10448

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *The RAND Journal of Economics*, *42*(3), 527–554. https://doi.org/10.1111/j.1756-2171.2011.00140.x

Beckers, N., Cools, M., & Van den Abbeele, A. (2010). The impact of incenctives on creativity: A literature review. *Review of Business and Economic Literature*, *55*(4), 463–484.

Belot, M., & Schröder, M. (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. *Journal of Economic Behavior & Organization*, *93*, 233–238. https://doi.org/10.1016/j.jebo.2013.03.019

Blaufus, K., Milde, M., Schaefer, M. & Schröder, M. (2024). Time to Innovate: The Role of Input Incentives in Creative Production (Working Paper)

Bradler, C., Neckermann, S., & Warnke, A. J. (2019a). Incentivizing creativity: A large-scale experiment with performance bonuses and gifts. *Journal of Labor Economics*, *37*(3), 793–851. https://doi.org/10.1086/702649

Bradler, C., Neckermann, S., & Warnke, A. J. (2019b). Incentivizing creativity: A large-scale experiment with performance bonuses and gifts. *Journal of Labor Economics*, *37*(3), 793–851. https://doi.org/10.1086/702649

Burtch, G., He, Q., Hong, Y., & Lee, D. (2022). How Do Peer Awards Motivate Creative Content? Experimental Evidence from Reddit. *Manage. Sci.*, *68*(5), 3488–3506. https://doi.org/10.1287/mnsc.2021.4040

Butschek, S., Amor, R. G., Kampkötter, P., & Sliwka, D. (2019). *Paying gig workers – Evidence from a field experiment* (12667; IZA DP).

Charness, G., & Grieco, D. (2019). Creativity and incentives. *Journal of the European Economic Association*, *17*(2), 454–496. https://doi.org/10.1093/jeea/jvx055

Cordero, R. (1990). The measurement of innovation performance in the firm: An overview. *Research Policy*, 19(2), 189–192. https://doi.org/10.1016/0048-7333(90)90048-B

Eckartz, K., Kirchkamp, O., & Schunk, D. (2012). How do incentives affect creativity? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2198760

Ederer, F., & Manso, G. (2011). Incentives for innovation: Bankruptcy, corporate governance, and compensation systems. *Handbook on Law, Innovation and Growth*, 90–111. https://doi.org/10.4337/9780857930545.00009

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, *59*(7), 1496–1513. https://doi.org/10.1287/mnsc.1120.1683

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., & Schudy, S. (2024). The effect of incentives in nonroutine analytical team tasks. *Journal of Political Economy*, *132*(8), https://doi.org/10.1086/729443.

Erat, S., & Gneezy, U. (2016). Incentives for creativity. *Experimental Economics*, *19*(2), 269–280. https://doi.org/10.1007/s10683-015-9440-5

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, *96*(5), 1611–1630. https://doi.org/10.1257/aer.96.5.1611

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611. https://doi.org/10.1111/1467-6419.00150

Gibbs, M., Neckermann, S., & Siemroth, C. (2017). A field experiment in motivating employee ideas. *The Review of Economics and Statistics*, *99*(4), 577–590. https://doi.org/10.1162/REST_a_00631

Gill, D., & Prowse, V. (2024). The creativity premium: Exploring the link between childhood creativity and life outcomes. Journal of Political Economy Microeconomics, 2(3), https://doi.org/10.1086/730453.

Greiner, B. (2004). An online recruitment system for economic experiments. *Forschung Und Wissenschaftliches Rechnen*, *63*, 79–93.

Gross, D. P. (2020). Creativity under fire: The effects of competition on creative production. *The Review of Economics and Statistics*, *102*(3), 583–599. https://doi.org/10.1162/rest_a_00831

Grözinger, N., Irlenbusch, B., Laske, K., & Schröder, M. (2020). Innovation and communication media in virtual teams – An experimental study. *Journal of Economic Behavior & Organization*, *180*, 201–218. https://doi.org/10.1016/j.jebo.2020.09.009

Herz, H., & Zihlmann, C. (2021). Adverse effects of monitoring: Evidence from a field experiment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3788710

Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, *10*(1), 74. https://doi.org/10.2307/3003320

Holmstrom, B., & Milgrom, P. (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, *7*(special_issue), 24–52. https://doi.org/10.1093/jleo/7.special_issue.24

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. https://doi.org/10.1111/j.1475-679X.2008.00277.x

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management journal*, *18*(4), 769-783. https://doi.org/10.5465/255378

Kleine, M. (2021). No eureka! Incentives hurt creative breakthrough irrespective of the incentives' frame. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3848440

Lagarde, M., & Blaauw, D. (2021). Effects of incentive framing on performance and effort: evidence from a medically framed experiment. *Journal of the Economic Science Association*, *7*(1), 33–48. https://doi.org/10.1007/s40881-021-00100-0

Leiner, D. J. (2014). *Convenience samples and respondent pools: A case study of the SoSci panel* (Working Paper).

Petters, L. M., & Schröder, M. (2020). Negative side effects of affirmative action: How quotas lead to distortions in performance evaluation. *European Economic Review*, *130*, 103500. https://doi.org/10.1016/j.euroecorev.2020.103500

Römer, N., & Schröder, M. (2024). *Effects of written self-promotions on gender bias and decision quality*.

Rubin, J., Samek, A., & Sheremeta, R. M. (2018). Loss aversion and the quantity–quality tradeoff. *Experimental Economics*, *21*(2), 292–315. https://doi.org/10.1007/s10683-017-9544-1

Smith, K. (2006). Measuring innovation. In J. Fagerberg & D. C. Mowery (Eds.), *The Oxford handbook of innovation* (pp. 148-180). Oxford Academic. https://doi.org/10.1093/oxfordhb/9780199286805.003.000.

Appendix

Appendix A: Additional Analysis Laboratory Experiment

A.1 Alternative Specifications

Table A1 provides robustness checks using a range of different measures to specify innovative ideas (Panel A), the quality of ideas (Panel B), and the originality of ideas (Panel C).

In Panel A, Table A1, we focus on different measures to determine innovative ideas. The first measure classifies ideas as innovative whenever the product of originality and quality is among the top 25% of ideas. This is the same measure as used in the main analyses. The second and third measures classify ideas as innovative when they belong to the top 10% of ideas or to the top 33% of ideas. For all of these measures, the idea has to be original in order to be innovative, while the threshold for quality changes depending on how restrictive we are in the classification. We find equivalent results showing a significant increase of innovative ideas in the Unweighted PR as compared to No PR for all three measures. We also find qualitatively similar results when it comes to a quality weight, however, these are not significant when using the more restrictive measure of only considering the top 10% of ideas. We additionally report results using different definitions of originality and employing different quality thresholds.

In the first measure, we adapt the quality thresholds employed. Avg. number original & non-zero quality and avg. number original & above median quality classify ideas as innovative when they are original and they meet a certain quality threshold. The quality threshold is quality >0 for the first measure and quality>0.5 for the latter.[26] Again, we replicate the positive effect of unweighted PR incentives compared to no PR, while we find qualitative evidence on the negative effect of a quality weight, this is, however, not always significant.

[26] Note that classifying ideas based on the top tercile results in the same specification as classifying them as innovative when their quality is > 0 and they are original, given that the resulting quality threshold for the top tercile is 0.1 (and thus, corresponds to non-zero quality).

Table A1: Alternative specifications of ideas (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-weighted PR
Panel A: Measures for innovative ideas				
Avg. number top quartile	3.97***	7.58	5.70**	6.91 n.s.
	(2.28)	(3.80)	(2.48)	(3.51)
Avg. number top percentile	1.50***	3.00	2.33 n.s.	3.03 n.s.
	(1.39)	(1.90)	(1.75)	(2.06)
Avg. number top tercile	5.13***	9.16	6.83**	8.63 n.s.
	(2.78)	(4.36)	(2.52)	(4.04)
Avg. number original & non-zero quality	5.13***	9.16	6.83**	8.63 n.s.
	(2.78)	(4.36)	(2.52)	(4.04)
Avg. number original & above median quality	2.06***	3.77	3.00 n.s.	3.34 n.s.
	(1.72)	(2.12)	(1.93)	(2.07)
Avg. number top quartile (original200)	3.78***	6.48	5.10*	6.28 n.s.
	(2.55)	(2.94)	(2.73)	(3.26)
Avg. number top quartile (below median frequency)	4.53***	7.13	5.33**	7.56 n.s.
	(2.84)	(3.13)	(2.64)	(3.75)
Avg. number top quartile (cont. originality)	3.81***	5.94	4.97 n.s.	6.34 n.s.
	(2.26)	(2.67)	(2.67)	(3.14)
Avg. number top quartile (uniqueness)	5.03 n.s.	6.29	5.77 n.s.	6.44 n.s.
	(2.49)	(3.20)	(2.65)	(2.92)
Panel B: Quality measures				
Avg. quality	0.50 n.s.	0.46	0.55**	0.48 n.s.
	(0.18)	(0.15)	(0.13)	(0.14)
Avg. quality (above-zero quality)	0.78 n.s.	0.77	0.85*	0.79 n.s.
	(0.21)	(0.20)	(0.15)	(0.14)
Avg. quality (above median quality)	0.51 n.s.	0.46	0.55**	0.46 n.s.
	(0.23)	(0.15)	(0.16)	(0.16)
Panel C: Originality measures				
Avg. originality	0.51 n.s.	0.56	0.51 n.s.	0.53 n.s.
	(0.20)	(0.17)	(0.13)	(0.17)
Avg. originality200	0.41 n.s.	0.45	0.40 n.s.	0.43 n.s.
	(0.24)	(0.16)	(0.16)	(0.14)
Avg. below median frequency	0.47 n.s.	0.49	0.41*	0.49 n.s.
	(0.23)	(0.17)	(0.16)	(0.14)
Avg. cont. originality	0.27 n.s.	0.26	0.20*	0.25 n.s.
	(0.21)	(0.15)	(0.11)	(0.11)
Avg. uniqueness	0.66*	0.59	0.57 n.s.	0.59 n.s.
	(0.20)	(0.17)	(0.16)	(0.12)
N Ideators	32	31	30	32

Note: We provide means and standard deviations (in parenthesis). Superscripts refer to results of MWU-tests comparing Unweighted PR to the other treatments, where ***p<0.01, **p<0.5, *p<0.10 and n.s. p>0.10. Panel A provides the avg. no. of innovative ideas according to different measures. Avg. no. top quartile, -percentile and -tercile classifies ideas as innovative if they score in the top quartile, top percentile or top tercile, respectively of the product of originality and quality. For Avg. no. original & non-zero quality and Avg. no. original & above median quality, ideas are classified as innovative if they are original and have non-zero or above median quality, respectively. Avg. no. top quartile (original200) classifies ideas as innovative if they score in the top quartile of the product of quality and originality, where an idea is defined as original if it is not among a random draw of 200 other ideas. Avg. no. top quartile (below median frequency) similarly classifies ideas as innovative, changing the definition of originality to below median frequency. Avg. no. top quartile uses the same definition but defines originality=1/frequency. Avg. no top quartile (based on inc. originality) shows the equivalent avg. no. of innovative ideas when using the incentivized originality measure (uniqueness within a group of 4). Panel B shows the average quality of idea, where avg. quality is the quality according to the measure from the main paper, frac. above-zero- and above median quality show the fraction of ideas with above zero and above median quality, respectively. Panel C shows the average originality according to different measures. Avg. originality shows the originality according to the measure used in the main paper, avg. originality200 uses the classification based on 200 randomly drawn ideas (see above), below median frequency reports the average originality when classifying ideas as original when they have below median frequency, avg. cont. originality is based on originality=1/frequency and avg. uniqueness shows the originality based on the incentivized measure (see above).

Next, we increase the number of ideas considered as a reference group to determine whether an idea is original when computing our main measure of innovative ideas (innovative if they score in the top 25% of ideas with respect to the product of quality and originality). Therefore, we compute avg. number top quartile (original200) based on a random set of 200 ideas instead of 100 ideas as a reference group to determine originality. We find similar results as for the initial measures based on 100 reference ideas.

Another way to determine originality is to consider originality as a continuous measure by focusing on the frequency of illustrated words. We determine the frequency as the number of times a specific word is illustrated within the entire sample. In our sample, median frequency is equal to 14. To compute avg. number top quartile (below median frequency) we again use our main measure but consider all ideas that occur less than 14 times in our sample as original.

Another way to consider originality as a continuous measure is to compute originality as 1 over the frequency of the illustrated word (original=1/frequency). Based on this measure we define avg. number top quartile (cont. originality). Again, we find a significant positive effect of unweighted PR incentives as compared to No PR for both measures. We that a quality weight mitigates this effect, but the difference is only statistically significant for Avg. number top quartile (cont. originality).

Lastly, we analyze the number of innovative ideas (uniqueness), using the measure of originality which was used to incentivize ideators in the lab experiment. That is, an idea was considered as original when it was unique in a group with four other ideators. We again compute the measure used in the main analysis (top 25% of the product of originality and quality) using this specification of originality. We replicate the positive effect of PR incentives on the number of innovative ideas. Using this measure, the difference between the Unweighted PR and No PR are not statistically significant ($p=0.14$). Uniqueness among four is a much less restrictive measure for originality than based on a random draw of 100 other ideas. Based on this classification, there will be more innovation in ideas that in fact would not be considered original using other measures, making this classification more noisy. Therefore, it is unsurprising that the effect is not statistically significant. However, given that results go in to the same direction we are confident that the difference does not affect the robustness of our main result. We further back this claim through the ample alternative

definitions shown above and our online experiment in which ideators were incentivized to use our main originality measure, supporting this confidence.

Overall, the finding that Unweighted PR incentives lead to a larger number of innovative ideas as compared to no PR is robust across all the other specifications of the number of innovative ideas provided in Table A1. With respect to the negative effect of a quality weight on the number of innovative ideas, our results are qualitatively consistent with the interpretation that a quality weight leads to a decrease in the number of innovative ideas. However, this difference is not always statistically significant.

In Panel B of Table A1, we focus on the quality of ideas. We show the average quality based on the quality thresholds used for the alternative definitions of innovative ideas above, and the measure from our main analyses (avg. quality, defined as the fraction of customers who correctly identify the illustrated word) as a comparison. Avg. quality (above-zero quality) and (above median quality) show the fraction of ideas with a quality greater than 0 or greater than the median, respectively. For both alternative measures Table A1 shows a significant increase in quality in the Quality-Weighted PR compared to Unweighted PR, which is in line with the effect observed in the main text (see avg. quality). Note, that these are the thresholds we used for the alternative classifications of innovative ideas reported above.

In Panel C of Table A1 we focus on the average originality of ideas. We report the measures used for the alternative definitions of innovative ideas and the measure from our main analyses (avg. originality, where ideas are original when they are unique within a random draw of 100 ideas) as a comparison. First, in our measure for avg. number of original based on 200 draws (avg. original200), we increase the number of ideas which are used as a reference group to determine originality from 100 (in the measure for avg. number original) to 200. Second, for the measure of avg. number below median frequency, we consider only illustrations of words that occur less than 14 times (median frequency) within our sample. Third, avg. cont. originality shows the average of 1/frequency (see above). Lastly, avg. uniqueness reports the average originality according to the incentivized measure in the lab (uniqueness within a group of 4). For the latter, we observe a slight increase in avg. originality in No PR that is significant at the 10% level.

A.2 Specialization

In Table A2 we provide further summary statistics on different types of ideas. We classify ideas as being 'obvious' and 'bad' to show that treatments can result in ideators specializing in certain types of ideas. We classify an idea as obvious if it illustrates a word that is among the top 10% most illustrated words and when the quality is ≥0.50. In the lab experiment, these are words that were illustrated seven or more times within the experimental sample. We classify an idea as bad if none of the customers could identify the illustrated word (quality=0).

We find that PR incentives insignificantly lower the fraction of obvious ideas, while a quality weight leads to a significant increase in obvious ideas. Absent PR incentives, the fraction of obvious ideas is slightly but insignificantly lower as compared to the unweighted PR treatment (MWU-test: p=0.27). We find that a quality weight leads to a significant increase in the number of obvious ideas compared to unweighted PR incentives (MWU-test: p=0.02), while an originality weight does not have an effect (MWU-test: p=0.98).

We do not find evidence of PR incentives affecting the fraction of bad ideas. Slightly less than 20% of the ideas generated in our experiment are bad. We find that this fraction is not significantly different between the No PR and the Unweighted PR treatments (MWU-test: p=0.78), or the Unweighted and the Originality-Weighted PR treatments (MWU-test: p=0.78). We find a slight decrease in the number of bad ideas when comparing the Quality-Weighted PR treatment to the unweighted PR treatment (MWU-test: p=0.11).

Another way to examine the effects of incentives on the types of ideas is to focus on the number of different recurring ideas by treatment. Table A2 shows that PR incentives do raise the overall number of distinct ideas that are generated. When comparing the fraction of recurring ideas by ideators, we find, in line with the other results, that this fraction is only significantly higher in Quality-Weighted PR.

Table A2: Specialization (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-weighted PR
Frac. obvious ideas	0.51 n.s.	0.45	0.52**	0.46 n.s.
	(0.23)	(0.16)	(0.15)	(0.15)
Frac. bad ides	0.21 n.s.	0.19	0.13 n.s.	0.18 n.s.
	(0.20)	(0.19)	(0.13)	(0.13)
Frac. recurring ideas	0.81 n.s.	0.84	0.90*	0.85 n.s.
	(0.22)	(0.15)	(0.10)	(0.10)
N distinct recurring ideas	154	188	159	199
N Ideators	32	31	30	32

Note: We provide means and standard deviations ins parenthesis. Superscripts refer to results of MWU-test comparing unweighted PR to the other treatments, where ***p<0.01, **p<0.5, *p<0.10 and n.s. p>0.10. Frac. obvious ideas reports the fraction of ideas that are in the top 10% of illustrated words with respect to their frequency and at least median quality. Frac. bad ideas the fraction of ideas that have a quality of 0. Frac. recurring ideas is the fraction of recurring ideas (that is, the average over the number of ideas based on words that are illustrated multiple times in the sample per ideator). Number of distinct recurring ideas shows the number of distinct words that are recurring ideas per treatment.

A.3 Additional descriptive statistics on performance measures

Table A3 shows the correlation of performance measures across treatments. In the first panel of Table A3, we focus on correlations on the ideator level. In the second panel, we provide correlations on the idea level. On both the ideator level and idea level, we find a positive correlation of number of ideas and (avg.) originality and a negative correlation of number of ideas and avg. quality. Furthermore, we find a negative correlation of quality and originality.

Table A3: Spearman correlations across performance dimensions (lab)

	Fixed Pay	Unweighted PR	Quality-Weighted PR	Originality-Weighted PR	Overall
Panel A: Ideator level					
Number of ideas and avg. originality	0.30*	0.35*	0.29 n.s	0.39***	0.35***
Number of ideas and avg. quality	-0.49***	-0.28 n.s	-0.46**	-0.39***	-0.42***
Avg. quality and avg. originality	-0.83***	-0.69***	-0.59***	-0.49***	-0.67***
N Ideators	32	31	30	32	125
Panel B: Idea level					
Number of ideas and originality	0.11**	0.13***	0.10**	0.12***	0.12***
Number of ideas and quality	-0.24***	-0.15***	-0.15***	-0.18***	-0.18***
Quality and originality	-0.56***	-0.47***	-0.51***	-0.49***	-0.50***
Time and originality	-0.07*	-0.05 n.s	-0.04 n.s	-0.01 n.s	-0.05**
Time and quality	0.14***	0.11***	0.12***	0.11***	0.13***
Complexity and originality	0.02 n.s	0.00 n.s	-0.04 n.s	0.05 n.s	0.01 n.s
Complexity and quality	0.04 n.s	0.08**	0.04 n.s	-0.03 n.s	0.03*
N ideas	531	772	544	785	2,632

Note: We report Spearman correlation coefficients (ρ) and significant levels in superscripts, where *** p<0.01, ** p<0.05, * p<0.1 and n.s. refers to p>0.10. Panel A reports correlation on ideator level, where number of ideas is the number of ideas that an ideator generated, avg. originality and avg. quality are the means of quality and originality over all these ideas. Panel B reports correlations on the idea level, and thus referring to the quality and originality of the respective idea. Time is the seconds the ideator took to generate an idea and complexity the number of materials used for the idea.

We visualize the reported correlations on the ideator level (Figure A1) in the scatter plots shown below.

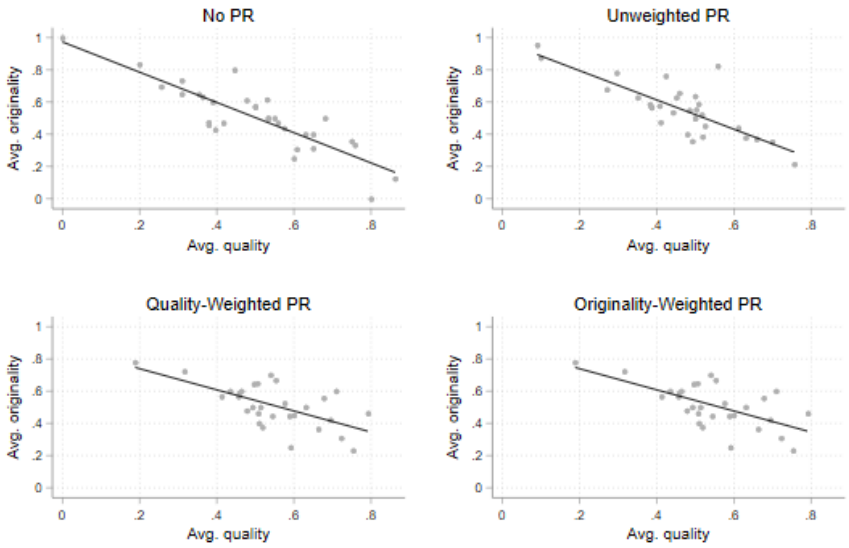


Figure A1: Scatter plot of avg. quality and avg. originality (lab)

To visualize the relation of quality and originality on the idea level, we provide histograms of the quality of all ideas in Figure A2, and conditioning on the originality of ideas in Figure A3 and Figure A4. Figure A3 shows the quality of ideas with originality=0 and Figure A4 for ideas with originality=1. Note that due to the binary nature of originality scatter plots are not suitable for displaying this relationship on the idea level.

In line with the negative correlation of quality and originality on the aggregate level the histograms confirm that the distribution of quality of original ideas (originality=1) is significantly different (higher frequency of values around 0) to that of ideas with originality=0. The distributions also differ significantly (pairwise Kolmogorov-Smirnov tests $p<0.01$).

The distribution of quality in Quality-Weighted PR differs significantly from all other treatments (Kolmogorov-Smirnov tests: $p<0.05$) with one exception. Only for ideas with originality=0 the distribution of quality between the Quality-Weighted PR and Originality-Weighted PR is not statistically different (Kolmogorov-Smirnov tests: $p=0.16$), although it shows a similar tendency.

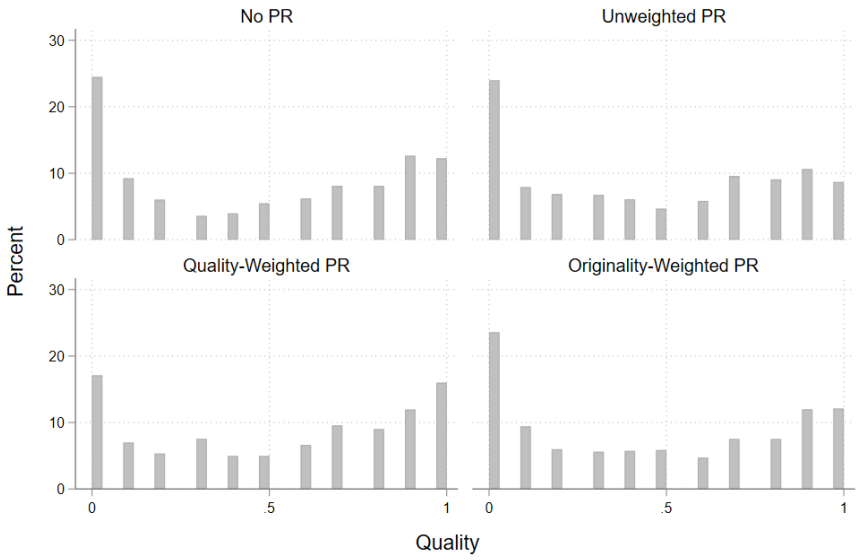


Figure A2: Histogram of quality of ideas by treatment (lab)

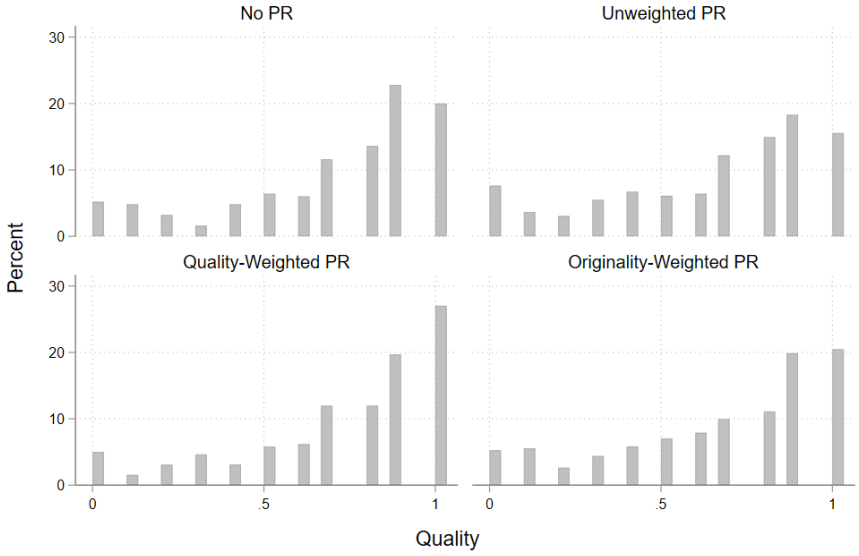


Figure A3: Histogram of quality of ideas with originality=0 (lab)

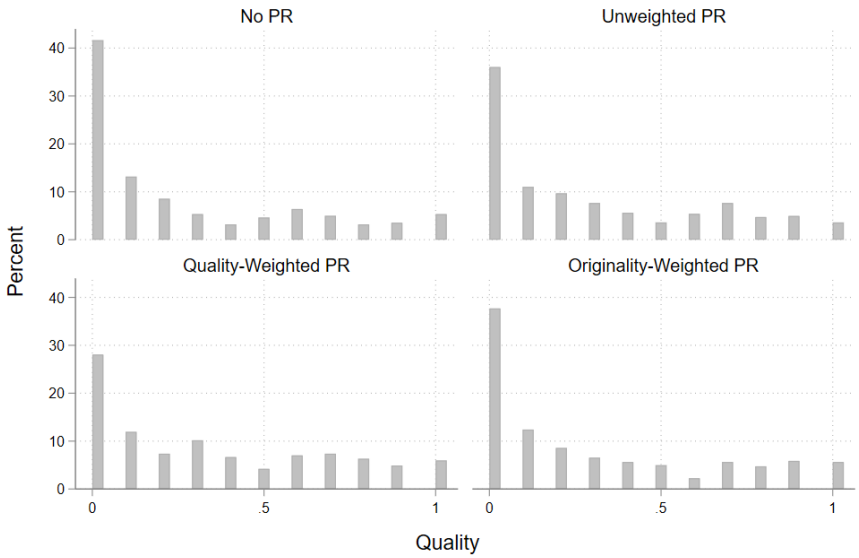


Figure A4: Histogram of quality of ideas with originality=1 (lab)

A.4 Effort provision

Table A4: Regression treatment differences in avg. rel. time (lab)

	Avg. rel. time	
	(1)	(2)
No PR	0.542**	0.401**
	(0.225)	(0.156)
Quality-Weighted PR	0.240***	0.127
	(0.064)	(0.092)
Originality-Weighted PR	0.092	0.033
	(0.093)	(0.088)
Avg. originality		-1.722*
		(0.952)
Avg. quality		0.290
		(0.419)
Avg. rel. quality		0.065
		(0.234)
Constant	-0.061	0.771
	(0.049)	(0.630)
Observations	124	124
R-squared	0.088	0.289

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1 of the main paper.

In the regression models reported in Table A4, we show that unweighted PR incentives lead to higher time efficiency. Therefore, we run OLS regressions with the avg. rel. time as a dependent variable and add indicators for the respective treatment conditions as well as controls for performance dimensions. The positive and significant coefficient of No PR in the model presented in Col. (1) shows that the avg. rel. time is significantly higher in No PR than in Unweighted PR. In line with our results reported in the main paper the Quality-Weighted PR also leads to significant higher rel. avg. time, as the positive and significant coefficient shows, while there is no difference for the Originality-Weighted PR. We control for the avg. rel. quality in the model presented in Col. (2), confirming that PR incentives have a positive impact on efficiency even when controlling for quality differences. Furthermore, the model presented in Col. (2) confirms that the decrease in efficiency gains through a quality weight are explained through quality differences.

A.5 Distortions across performance dimensions

Below, we provide regression results for the separate dimensions of performance, i.e., number of ideas (Table A5), quality (Table A6), and originality (Table A7).

For number of ideas (Table A5), we confirm that observed positive effect of unweighted PR incentives on the number of ideas seems at most marginally driven by distortions, but rather due to an increase in effort or a change in the types of ideas. The significant negative effect of No PR relative to Unweighted PR incentives in the model in Col. (1), if anything, only slightly changes when controlling for performance in other dimensions, confirming that their effect is only marginal. In contrast, the effect is reduced by 60% when controlling for efficiency (by adding controls on avg. rel. time and avg. rel. quality to account for quality differences) in the model presented in Col. (4), as well as types of ideas in the model in Col. (5), and it loses significance.

We provide the same models on avg. quality in Table A6, confirming the previously shown relationship between number of ideas and avg. quality, as well as for avg. originality in Table A7. All of the presented models show that distortions do not matter much for explaining the increase in number of ideas through PR incentives, and confirm the role of efficiency gains and different types of ideas.

Table A5: Regression of number of ideas (lab)

	Number of ideas				
	(1)	(2)	(3)	(4)	(5)
No PR	-8.323***	-6.605***	-6.657***	-4.128**	-3.798**
	(2.220)	(1.875)	(1.929)	(1.718)	(1.469)
Quality-Weighted PR	-7.286***	-5.236***	-5.143***	-4.006***	-4.239***
	(1.979)	(1.709)	(1.668)	(1.438)	(1.327)
Originality-Weighted PR	-0.888	-0.172	-0.060	0.559	-0.043
	(2.418)	(2.102)	(2.028)	(1.661)	(1.563)
Avg. originality		11.278*	19.515	32.486	21.538
		(6.402)	(22.704)	(22.196)	(20.291)
Avg. quality		-17.958**	-9.350	18.135	23.663
		(7.186)	(19.476)	(21.026)	(19.919)
Avg. quality x Avg. originality			-14.646	-56.484	-34.091
			(36.821)	(36.589)	(34.370)
Avg. rel. time				-6.393**	-4.501*
				(2.812)	(2.343)
Avg. rel. quality				-0.619	-0.518
				(4.526)	(4.367)
Avg. idea time					-0.642***
					(0.115)
Avg. idea quality					-13.894
					(12.818)
Constant	25.419***	27.417***	22.347*	11.985	51.395***
	(1.724)	(6.331)	(12.701)	(13.104)	(14.552)
Observations	124	124	124	124	124
R-squared	0.172	0.377	0.380	0.534	0.654

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1 of the main paper.

Table A6: Regression of avg. quality (lab)

	Avg. quality				
	(1)	(2)	(3)	(4)	(5)
No PR	0.053	-0.013	-0.009	-0.040*	-0.018
	(0.039)	(0.026)	(0.026)	(0.021)	(0.016)
Quality-Weighted PR	0.085**	0.034	0.034	0.006	0.016
	(0.035)	(0.026)	(0.026)	(0.021)	(0.015)
Originality-Weighted PR	0.023	0.005	0.005	-0.023	-0.002
	(0.036)	(0.027)	(0.026)	(0.021)	(0.013)
Number of ideas		-0.003***	0.002	0.001	0.003
		(0.001)	(0.003)	(0.003)	(0.002)
Avg. originality		-0.547***	-0.414***	-0.436***	-0.174**
		(0.079)	(0.109)	(0.107)	(0.071)
Number of ideas x Avg. originality			-0.008**	-0.005	-0.003
			(0.004)	(0.004)	(0.003)
Avg. rel. time				0.004	0.012
				(0.011)	(0.008)
Avg. rel. quality				0.312***	0.260***
				(0.050)	(0.027)
Avg. idea time					0.000
					(0.002)
Avg. idea quality					0.896***
					(0.086)
Constant	0.462***	0.855***	0.773***	0.777***	0.777***
	(0.026)	(0.040)	(0.063)	(0.073)	(0.073)
Observations	124	124	124	124	124
R-squared	0.047	0.550	0.562	0.695	0.695

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1 of the main paper.

Table A7: Regression of avg. originality (lab)

	Avg. originality				
	(1)	(2)	(3)	(4)	(5)
No PR	-0.068	-0.006	-0.000	-0.022	-0.020
	(0.044)	(0.031)	(0.033)	(0.032)	(0.034)
Quality-Weighted PR	-0.047	0.035	0.037	0.013	0.016
	(0.038)	(0.029)	(0.030)	(0.030)	(0.032)
Originality-Weighted PR	-0.027	-0.008	-0.008	-0.024	-0.023
	(0.042)	(0.031)	(0.031)	(0.030)	(0.031)
Number of ideas		0.003*	-0.000	0.002	0.002
		(0.002)	(0.003)	(0.002)	(0.003)
Avg. quality		-0.718***	-0.871***	-0.792***	-0.843***
		(0.068)	(0.162)	(0.153)	(0.208)
Number of ideas x Avg. quality			0.008	-0.002	-0.002
			(0.007)	(0.006)	(0.006)
Avg. rel. time				-0.053***	-0.052***
				(0.016)	(0.016)
Avg. rel. quality				0.228***	0.231***
				(0.075)	(0.086)
Avg. idea time					0.001
					(0.003)
Avg. idea quality					0.056
					(0.236)
Constant	0.558***	0.817***	0.882***	0.924***	0.835***
	(0.030)	(0.065)	(0.082)	(0.079)	(0.183)
Observations	124	124	124	124	124
R-squared	0.023	0.525	0.531	0.603	0.604

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper.

A.6 Robustness to inclusion of ideators without recurring ideas

To separate effort provision from the types of ideas generated, we generate measures based on recurring ideas (ideas that are generated multiple times in our sample). In fact, around 84% of ideas are recurring ideas, covering a large majority of generated ideas. To include measures based on recurring ideas, we have to rely on the subsample of ideators that generated at least one recurring idea. In the lab experiment, there is only one ideator that did not generate at least one recurring idea. In Table A8, Table A9, Table A10, and Table A11, we provide regression results on the relevant performance dimensions for the used subsample (only ideators with at least one recurring idea) and without. Coefficients only differ marginally across all regressions, showing that our results are robust to the inclusion of the full sample.

Table A8: Regression of number of ideas of full sample vs. subsample (lab)

	Number of ideas					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	-8.323***	-6.605***	-6.657***	-8.826***	-7.818***	-7.528***
	(2.220)	(1.875)	(1.929)	(2.250)	(2.164)	(2.142)
Quality-Weighted PR	-7.286***	-5.236***	-5.143***	-7.286***	-5.668***	-5.751***
	(1.979)	(1.709)	(1.668)	(1.978)	(1.748)	(1.761)
Originality-Weighted PR	-0.888	-0.172	-0.060	-0.888	-0.313	-0.477
	(2.418)	(2.102)	(2.028)	(2.417)	(2.129)	(2.133)
Avg. originality		11.278*	19.515		9.592	-3.959
		(6.402)	(22.704)		(6.735)	(24.541)
Avg. quality		-17.958**	-9.350		-13.792*	-28.950
		(7.186)	(19.476)		(7.931)	(21.233)
Avg. quality x Avg. originality			-14.646			24.599
			(36.821)			(39.471)
Constant	25.419***	27.417***	22.347*	25.419***	26.434***	35.114**
	(1.724)	(6.331)	(12.701)	(1.724)	(6.482)	(13.997)
Observations	124	124	124	125	125	125
R-squared	0.172	0.377	0.380	0.180	0.321	0.330

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A9: Regression of avg. quality of full sample vs. subsample (lab)

	Avg. quality					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	0.053	-0.013	-0.009	0.037	-0.017	-0.015
	(0.039)	(0.026)	(0.026)	(0.042)	(0.026)	(0.027)
Quality-Weighted PR	0.085**	0.034	0.034	0.085**	0.039	0.039
	(0.035)	(0.026)	(0.026)	(0.035)	(0.026)	(0.026)
Originality-Weighted PR	0.023	0.005	0.005	0.023	0.004	0.005
	(0.036)	(0.027)	(0.026)	(0.036)	(0.027)	(0.027)
Number of ideas		-0.003***	0.002		-0.002*	0.001
		(0.001)	(0.003)		(0.001)	(0.003)
Avg. originality		-0.547***	-0.414***		-0.606***	-0.536***
		(0.079)	(0.109)		(0.086)	(0.132)
Number of ideas x Avg. originality			-0.008**			-0.005
			(0.004)			(0.004)
Constant	0.462***	0.855***	0.773***	0.462***	0.863***	0.816***
	(0.026)	(0.040)	(0.063)	(0.026)	(0.041)	(0.071)
Observations	124	124	124	125	125	125
R-squared	0.047	0.550	0.562	0.040	0.565	0.569

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A10: Regression of avg. originality of full sample vs. subsample (lab)

	Avg. originality					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	-0.068	-0.006	-0.000	-0.052	-0.005	0.002
	(0.044)	(0.031)	(0.033)	(0.046)	(0.031)	(0.032)
Quality-Weighted PR	-0.047	0.035	0.037	-0.047	0.034	0.037
	(0.038)	(0.029)	(0.030)	(0.038)	(0.029)	(0.030)
Originality-Weighted PR	-0.027	-0.008	-0.008	-0.027	-0.007	-0.008
	(0.042)	(0.031)	(0.031)	(0.042)	(0.031)	(0.031)
Number of ideas		0.003*	-0.000		0.002	-0.001
		(0.002)	(0.003)		(0.002)	(0.003)
Avg. quality		-0.718***	-0.871***		-0.766***	-0.936***
		(0.068)	(0.162)		(0.068)	(0.135)
Number of ideas x Avg. quality			0.008			0.010
			(0.007)			(0.007)
Constant	0.558***	0.817***	0.882***	0.558***	0.856***	0.920***
	(0.030)	(0.065)	(0.082)	(0.030)	(0.064)	(0.067)
Observations	124	124	124	125	125	125
R-squared	0.023	0.525	0.531	0.015	0.547	0.558

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A11: Regression of number of innovative ideas of full sample vs. subsample (lab)

	DV: Number of innovative ideas							
	Subsample				Full sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No PR	-3.484***	-1.688**	-1.476***	-0.915***	-3.612***	-1.698**	-1.456***	-0.872**
	(0.789)	(0.658)	(0.416)	(0.340)	(0.793)	(0.659)	(0.420)	(0.348)
Quality-Weighted PR	-1.881**	-0.309	-0.964**	-0.786**	-1.881**	-0.300	-0.983**	-0.749*
	(0.819)	(0.700)	(0.472)	(0.394)	(0.819)	(0.697)	(0.475)	(0.397)
Originality-Weighted PR	-0.674	-0.483	-0.510	-0.751**	-0.674	-0.482	-0.502	-0.706**
	(0.923)	(0.760)	(0.502)	(0.351)	(0.923)	(0.760)	(0.505)	(0.351)
Number of ideas		0.216***	0.247***	-0.795***		0.217***	0.238***	-0.809***
		(0.043)	(0.036)	(0.143)		(0.042)	(0.035)	(0.146)
Avg. quality			12.899***	-14.395***			13.106***	-10.877***
			(1.376)	(4.110)			(1.366)	(3.304)
Avg. originality			17.441***	-17.070***			17.016***	-15.586***
			(1.664)	(3.614)			(1.630)	(3.482)
Avg. quality x Avg. originality				20.791***				15.369***
				(5.233)				(3.761)
Number of ideas x Avg. quality				1.186***				1.247***
				(0.159)				(0.151)
Number of ideas x Avg. originality				0.925***				0.895***
				(0.152)				(0.149)
Constant	7.581***	2.096*	-13.938***	11.805***	7.581***	2.067*	-13.635***	10.562***
	(0.683)	(1.078)	(1.622)	(2.984)	(0.683)	(1.052)	(1.594)	(2.880)
Observations	124	124	124	124	125	125	125	125
R-squared	0.161	0.444	0.734	0.861	0.171	0.458	0.738	0.862

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, Quality-Weighted PR and Originality-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(4) show results based on the subsample of ideators who have at least one recurring idea. Col. (5)-(8) show results based on the full sample of ideators.

A.7 Complexity of recurring ideas

Table A12 provides summary statistics on the avg. idea complexity and the avg. rel. complexity of recurring ideas. When referring to complexity, we refer to the number of materials used in an idea. We define avg. idea complexity and avg. rel. complexity for recurring ideas in a manner that is similar to how we define avg. idea quality and avg. rel. quality of recurring ideas. That is, avg. idea complexity is the average complexity across all recurring ideas generated by an ideator. Where idea complexity is the average complexity across all illustrations of the same word in our sample. We define avg. rel. complexity as the average of rel. complexity for all recurring ideas (j) generated by individual (i). Where $rel.complexity_{j\,k} = \frac{(complexity_{jk} - idea\ complexity_j)}{idea\ complexity_j}$.

Qualitatively, we show that PR incentives decrease both the avg. complexity and the avg. rel. complexity of recurring ideas. However, these effects are not significant.

Table A12: Complexity of recurring ideas (lab)

	No PR	Unweighted PR	Quality-Weighted PR	Originality-Weighted PR
Avg. idea complexity	6.32 $^{n.s.}$	6.16	6.37 $^{n.s.}$	6.06 $^{n.s.}$
	(0.94)	(0.84)	(0.96)	(0.95)
Avg. rel. complexity	0.06 $^{n.s.}$	-0.01	0.01$^{n.s.}$	-0.02$^{n.s.}$
	(0.27)	(0.17)	(0.19)	(0.21)
N ideators illustrating recurring ideas	31	31	30	32

Note: We provide means and standard deviations ins parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where *** p<0.01, ** p<0.05, * p<0.1 and $^{n.s.}$ p>0.10. All results for recurring ideas are based on a limited sample of those ideators who illustrate at least one recurring idea.

Appendix B: Additional Analysis Online Experiment

B.1 Alternative specifications

Table A13 provides robustness checks using a range of different measures to specify innovative ideas (Panel A), the quality of ideas (Panel B), and the originality of ideas (Panel C). Note that we use the same definitions as the robustness check, like we did for the lab experiment.[27]

In Panel A, Table A13, we again focus on different measures to determine innovative ideas. The first measure classifies ideas as innovative whenever the product of originality and quality is among the top 25% of ideas (measure used in the main analyses). The second and third measures classify ideas as innovative when they belong to the top 10% of ideas or to the top 33% of ideas. For all of these measures, the idea has to be original in order to be innovative, while the threshold for quality changes depending on how restrictive we are in the classification. We find equivalent results showing a significant increase of innovative ideas in the Unweighted PR as compared to No PR for all three measures. We find qualitatively similar results when classifying ideas as original based on the top percentile. These are, however, not significant when using the more restrictive measure of only considering the top 10% of ideas. In line with adverse effects of control, we find that the number of innovative ideas is lower in the Innovation-Weighted PR compared to Unweighted PR. The effect is not always statistically significant compared to Unweighted PR. However, we find that for all measures the number of innovative ideas in Innovation-Weighted PR is not different from No PR (pairwise MWU-tests: $p \geq 0.39$).

[27] With the exception of the measure for uniqueness, as this is not applicable for the online experiment, where we incentivized ideators already using the measure applied in the main paper (the idea is original if not among a random set of 100 other ideas from the sample).

Table A13: Alternative specifications of ideas (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Panel A: Measures for innovative ideas			
Avg. number top quartile	2.20*	2.54	2.26 n.s.
	(1.75)	(2.01)	(1.78)
Avg. number top percentile	0.84 n.s.	0.98	0.87 n.s.
	(1.04)	(1.19)	(1.01)
Avg. number top tercile	2.88***	3.49	3.03**
	(2.07)	(2.44)	(2.09)
Avg. number original & non-zero quality	3.95***	4.78	4.14**
	(2.33)	(2.87)	(2.53)
Avg. number original & above median quality	1.84**	2.16	1.89 n.s.
	(1.60)	(1.81)	(1.60)
Avg. number top quartile (original 200)	2.19***	2.51	2.21***
	(1.72)	(2.00)	(1.78)
Avg. number top quartile (below median frequency)	2.18 n.s.	2.54	2.21 n.s.
	(1.69)	(1.96)	(1.72)
Avg. number top quartile (based on cont. originality)	2.04**	2.38	2.18 n.s.
	(1.66)	(1.86)	(1.70)
Panel B: Quality measures			
Avg. quality	0.48 n.s.	0.49	0.49 n.s.
	(0.20)	(0.18)	(0.19)
Avg. quality (above-zero quality)	0.81 n.s.	0.82	0.83 n.s.
	(0.21)	(0.18)	(0.20)
Avg. quality (above median quality)	0.46 n.s.	0.46	0.45 n.s.
	(0.25)	(0.23)	(0.24)
Panel C: Originality measures			
Avg. originality	0.66 n.s.	0.65	0.67 n.s.
	(0.23)	(0.21)	(0.21)
Avg. originality200	0.67*	0.64	0.66 n.s.
	(0.22)	(0.21)	(0.22)
Avg. originality (below median frequency)	0.52*	0.48	0.50 n.s.
	(0.25)	(0.23)	(0.25)
Avg. cont. originality	0.18 n.s.	0.16	0.17 n.s.
	(0.17)	(0.14)	(0.17)
N Ideators	298	299	305

Note: We provide means and standard deviations ins parentheses. Superscripts refer to results of MWU-test comparing unweighted PR to the other treatments, where ***p<0.01, **p<0.5, *p<0.10 and n.s. p>0.10. Panel A provides the number of innovative ideas according to different measures. Avg. number top quartile, - percentile and -tercile classifies ideas as innovative if they score in the top quartile, top percentile or top tercile, respectively of the product of originality and quality. For Avg. number original & non-zero quality and Avg. number original & above median quality, ideas are classified as innovative if they are original and have non-zero or above median quality, respectively. Avg. number top quartile (original200) classifies ideas as innovative if they score in the top quartile of the product of quality and originality, where an idea is defined as original if it is not among a random draw of 200 other ideas. Avg. number top quartile (below median frequency) similarly classifies ideas as innovative but changes the definition of originality to having below median frequency. Avg. number top quartile also uses the same definition but defines originality=1/frequency. Panel B shows the average quality of idea, where avg. quality is the quality according to the measure from the main paper, frac. above-zero- and above median quality show the fraction of ideas that have above zero and above median quality, respectively. Panel C shows the average originality according to different measures. Avg. originality shows the originality according to the measure used in the main paper, avg. originality200 uses the classification based on 200 randomly drawn ideas (see above), below median frequency reports the average originality when classifying ideas as original when they have below median frequency and avg. cont. originality is based on the average over originality=1/frequency.

We additionally report results using different definitions of originality and employing different quality thresholds. First, we move the quality thresholds employed. Avg. number

original & non-zero quality and avg. number original & above median quality classify ideas as innovative when they are original and they meet a certain quality threshold. The quality threshold is quality >0 for the first measure and quality>0.5 for the latter. Again, we replicate the positive effect of Unweighted PR incentives compared to No PR and adverse effects of innovation-weighting, which is not always significant. However, again we find that the number of innovative ideas in Innovation-Weighted PR does not differ from No PR (pairwise MWU-tests: $p \geq 0.46$).

Next, we increase the number of ideas considered as a reference group to determine whether an idea is original when computing our main measure of innovative ideas (innovative if they score in the top 25% of ideas with respect to the product of quality and originality). Therefore, we compute avg. number top quartile (original200) based on a random set of 200 ideas instead of 100 ideas as a reference group to determine originality. We again find results that are similar to the initial measures based on 100 reference ideas.

Another way to determine originality is to consider originality as a continuous measure by focusing on the frequency of illustrated words. We determine the frequency as the number of times a specific word is illustrated within the entire sample. In our sample, median frequency is equal to 26. To compute avg. number top quartile (below median frequency) we again use our main measure but consider all ideas that occur less than 26 times in our sample as original.

Another way to consider originality as a continuous measure is to compute originality as 1 over the frequency of the illustrated word (original=1/frequency). Based on this measure we define avg. number top quartile (cont. originality). Again, we find a significant positive effect of Unweighted PR incentives as compared to No PR for both measures, while the effect for avg. number top quartile (below median frequency) is not statistically significant (MWU-test: $p=0.11$) but shows a similar increase in the number of innovative ideas. We find that an innovation weight mitigates this effect, but the difference is not statistically significant. However, again, the Innovation-Weighted PR does not differ from No PR for both measures (pairwise MWU-test: ≥ 0.92).

Overall, the finding that Unweighted PR incentives lead to a larger number of innovative ideas as compared to No PR is robust across all the other specifications of the number of innovative ideas provided in Table A13. With respect to the negative effect of an innovation

weight on the number of innovative ideas, our results are qualitatively consistent with the interpretation that an innovation weight leads to a decrease in the number of innovative ideas to the point in which results do not significantly differ from No PR.

In Table A13, we focus on the quality of ideas. We show the average quality based on the quality thresholds used for the alternative definitions of innovative ideas above, and the measure from our main analyses (avg. quality, defined as the fraction of customers who correctly identify the illustrated word) as a comparison. Avg. quality (above-zero quality) and (above median quality) show the fraction of ideas with a quality greater than 0 or greater than the median, respectively. For both alternative measures, we show that our treatments have no effect on the quality of ideas. Note, that these are the thresholds we used for the alternative classifications of innovative ideas reported above.

In Panel C, Table A13, we focus on the average originality of ideas. We report the measures used for the alternative definitions of innovative ideas and the measure from our main analyses (avg. originality, where ideas are original when they are unique within a random draw of 100 ideas) as a comparison. First, in our measure for avg. number of original based on 200 draws (avg. original200), we increase the number of ideas which are used as a reference group to determine the originality from 100 (in the measure for avg. number original) to 200. Second, for the measure of avg. number below median frequency, we consider only illustrations of words that occur less than 14 times (median frequency) within our sample. Third, avg. cont. originality shows the average of 1/frequency (see above).

We mainly observe that originality is unaffected by PR incentives. If anything, there is some slight indication of higher originality in No PR, which is in line with adverse effects of control. However, effects are small and only significant at the 10% level for two of the employed measures.

B.2 Specialization

Similar to the lab experiment, we analyze different types of ideas in Table A14. We classify an idea as obvious if it illustrates a word that is among the top 10% most often illustrated words and when the quality is ≥0.50. In the online experiment, obvious ideas are illustrations of words that were illustrated 14 or more times within the experimental sample. We classify an idea as bad if none of the customers could identify the illustrated word (quality=0). For the online experiment, we do not find any effects on obvious or bad ideas. In line with the lab experiment, we find that the number of distinct recurring ideas is somewhat higher when PR are introduced. When comparing the fraction of recurring ideas by ideators, we do not find significant differences between treatments.

Table A14: Specialization (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Frac. obvious ideas	0.41 $^{n.s.}$	0.43	0.43 $^{n.s.}$
	(0.25)	(0.23)	(0.24)
Frac. bad ides	0.17 $^{n.s.}$	0.16	0.16 $^{n.s.}$
	(0.20)	(0.18)	(0.19)
Frac. recurring ideas	0.90 $^{n.s.}$	0.93	0.92 $^{n.s.}$
	(0.16)	(0.12)	(0.16)
N distinct recurring ideas	495	519	511
N Ideators	298	299	305

Note: We provide means and standard deviations in parentheses. Superscripts refer to results of MWU-test comparing unweighted PR to the other treatments, where ***p<0.01, **p<0.5, *p<0.10 and $^{n.s.}$ p>0.10. Frac. obvious ideas reports the fraction of ideas that are in the top 10% of illustrated words with respect to their frequency and at least median quality. Frac. bad ideas the fraction of ideas that have a quality of 0. Frac. recurring ideas is the fraction of recurring ideas (that is, the average over the number of ideas based on words that are illustrated multiple times in the sample per ideator). Number of distinct recurring ideas shows the number of distinct words that are recurring ideas per treatment.

B.3 Additional descriptive statistics on performance measures

Table A15 shows Spearman correlations between performance dimensions across treatments in the online experiment. In the first panel of Table A15, we focus on correlations on the ideator level. In the second panel, we provide correlations on the idea level. On both the ideator and idea level, we negative correlations of originality and quality.

Table A15: Spearman correlations across performance dimensions (online)

	Fixed pay	Unweighted PR	Innovation-Weighted PR	Overall
Panel A: Ideator level				
Number of ideas and avg. originality	-0.07 n.s.	-0.04 n.s.	-0.13**	-0.09***
Number of ideas and avg. quality	0.00 n.s.	-0.11*	0.04 n.s.	-0.02
Avg. quality and avg. originality	-0.41***	-0.39***	-0.41***	-0.40***
N Ideators	298	299	305	902
Panel B: Idea level				
Number of ideas and originality	-0.01 n.s.	0.02 n.s.	-0.01 n.s.	0.00 n.s.
Number of ideas and quality	-0.04*	-0.06***	0.00	-0.04***
Quality and originality	-0.34***	-0.37***	-0.35***	-0.35***
Time and originality	0.03 n.s.	-0.01 n.s.	0.05***	0.02**
Time and quality	0.04*	0.06***	0.01 n.s.	0.04***
Complexity and originality	0.01 n.s.	-0.03*	0.01 n.s.	-0.01 n.s.
Complexity and quality	0.14***	0.13***	0.11***	0.13***
N ideas	2,429	2,938	2,486	7,853

Note: We report Spearman correlation coefficients (ρ) and significant levels in superscripts, where *** p<0.01, ** p<0.05, * p<0.1 and n.s. refers to p>0.10. Panel A reports correlation on ideator level, where number of ideas is the number of ideas that an ideator generated, avg. originality and avg. quality are the means of quality and originality over all these ideas. Panel B reports correlations on the idea level, and thus referring to the quality and originality of the respective idea. Time is the seconds the ideator took to generate an idea and complexity the amount of materials used for the idea.

We visualize the reported correlation from Table A15 in the scatter plots shown below.

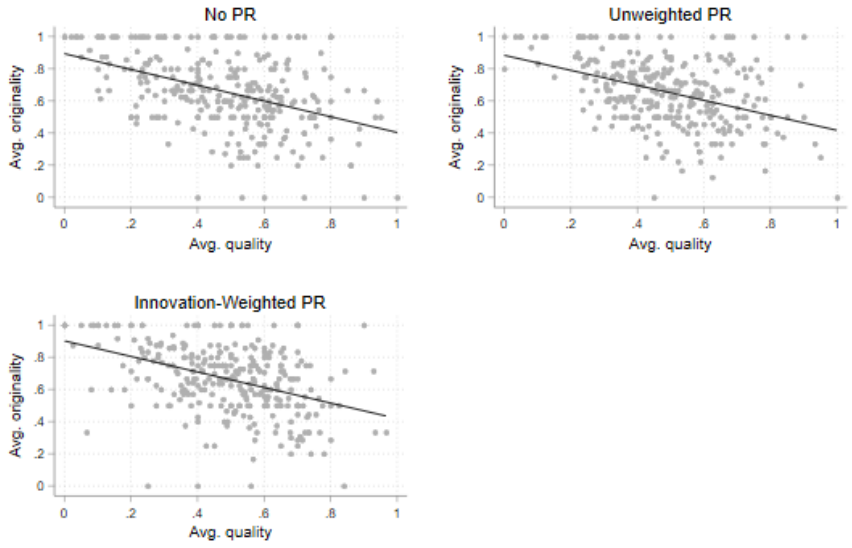


Figure A5: Scatter plot of avg. quality and avg. originality (online)

To visualize the relationship between quality and originality by treatment on the idea level, we provide histograms of the quality of all ideas in Figure A6, and conditioning on the originality of ideas in Figure A7 and Figure A8. Figure A7 shows the quality of ideas with originality=0 and Figure A8 for ideas with originality=1. Note that due to the binary nature of originality scatter plots are not suitable for displaying this relationship on the idea level. Clearly, the distributions look very similar to those in the lab experiment.

In line with the negative correlation of quality and originality on the aggregate level the histograms confirm that the distribution of quality of original ideas (originality=1) is significantly different (higher frequency of values around 0) to that of ideas with originality=0. In the online experiment we do not find significant differences in the distributions by treatments (Komolgorov-Smirnoff tests: $p \geq 0.20$).

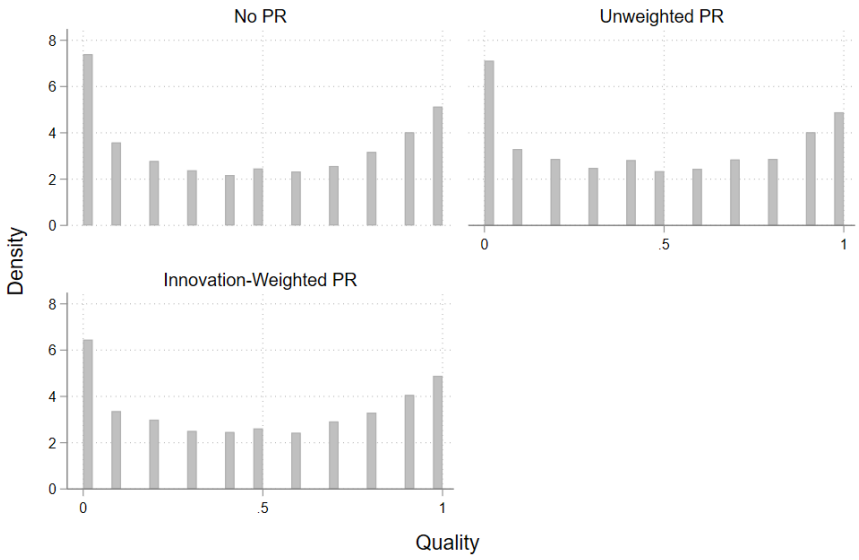


Figure A6: Histogram of quality of ideas by treatment (online)

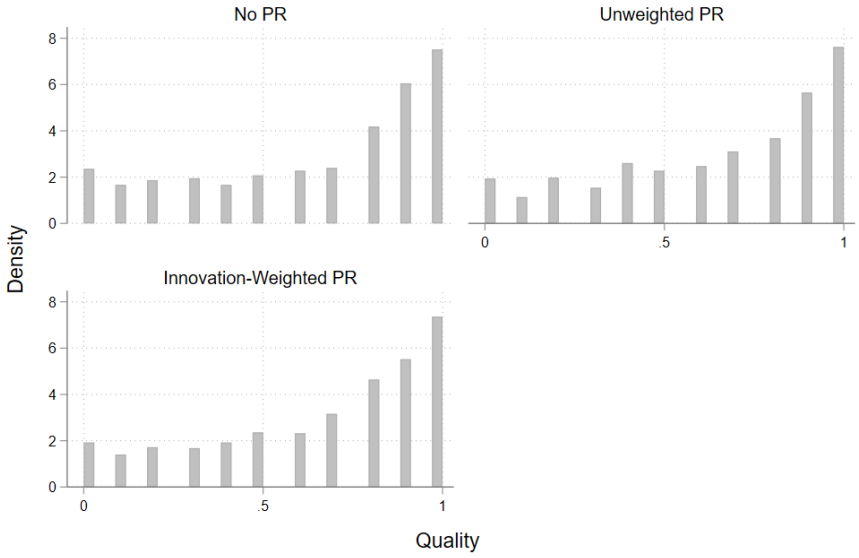


Figure A7: Histogram of quality of ideas with originality=0 (online)

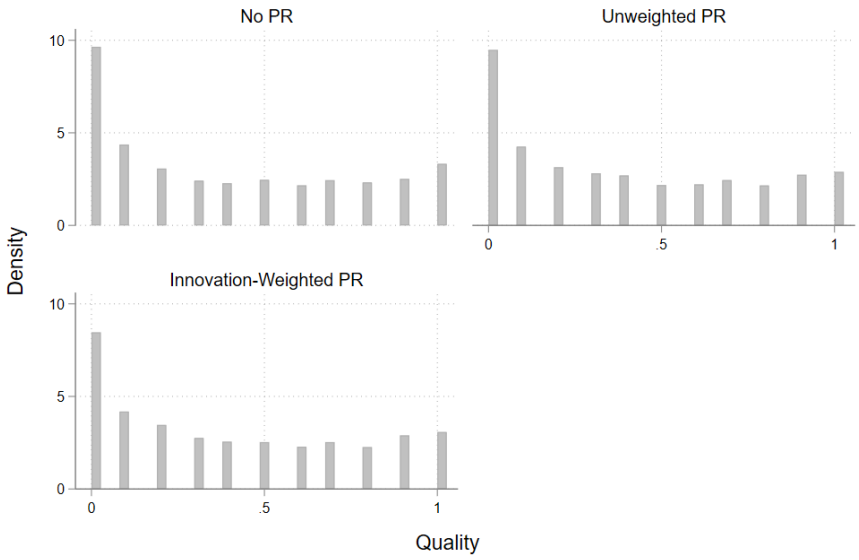


Figure A8: Histogram of quality of ideas with originality=1 (online)

B.4 Effort provision

In the regression models reported in Table A16, we show for the online experiment that unweighted PR incentives lead to higher time efficiency. In the OLS regressions avg. rel. time is the dependent variable and we add indicators for the respective treatment conditions as well as controls for performance dimensions as independent variables. The positive and significant coefficient of No PR in the model presented in Col. (1) shows that the avg. rel. time is significantly higher in No PR than in Unweighted PR. We control for the avg. rel. quality in the model presented in Col. (2), confirming that PR incentives have a positive impact on efficiency even when controlling for quality differences. If anything, the effect is even stronger when controlling for quality differences.

Table A16: Regression treatment differences in avg. rel. time (online)

	DV: Avg. rel. time	
	(1)	(2)
No PR	0.149***	0.153***
	(0.044)	(0.044)
Innovation-WeightedPR	0.172***	0.169***
	(0.050)	(0.049)
Avg. originality		0.095
		(0.173)
Avg. quality		0.185
		(0.229)
Avg. rel. quality		0.100
		(0.104)
Constant	0.096***	-0.058
	(0.031)	(0.211)
Observations	897	895
R-squared	0.017	0.028

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1 of the main paper.

B.5 Models' performance dimensions

Below, we provide regression results for the separate dimensions of performance, i.e., number of ideas (Table A17), quality (Table A18), and originality (Table A19).

For the number of ideas, we again confirm that the observed positive effect of unweighted PR incentives on the number of ideas seems at most marginally driven by distortions, but rather due to an increase in effort or a change in the types of ideas. The significant negative effect of No PR relative to unweighted PR incentives in the model in Col. (1) if anything, only slightly changes when controlling for performance in other dimensions, confirming that their effect is only marginal. The significant negative effect of No PR relative to unweighted PR incentives in the model in Col. (1) if anything, only slightly changes when controlling for performance in other dimensions, confirming that their effect is only marginal. In contrast, the effect is reduced by around 46% when controlling for efficiency (by adding controls on avg. rel. time and avg. rel. quality to account for quality differences) in the model presented in Col. (4), as well as types of ideas in the model in Col. (5).

We provide the same models on avg. quality (see Table A18) and avg. originality (Table A19). All of the presented models show that distortions do not matter much for explaining the increase in number of ideas through PR incentives and confirm the role of efficiency gains. For the online experiment, we find that the role of the type of ideas for our main treatment effects is rather small.

Table A17: Regression of number of ideas (online)

	Number of ideas				
	(1)	(2)	(3)	(4)	(5)
No PR	-1.651***	-1.662***	-1.614***	-0.876***	-0.804***
	(0.348)	(0.347)	(0.343)	(0.241)	(0.216)
Innovation-Weighted PR	-1.590***	-1.579***	-1.580***	-0.733***	-0.740***
	(0.355)	(0.354)	(0.353)	(0.248)	(0.220)
Avg. quality		-1.933**	-9.942***	-6.217***	-4.957**
		(0.828)	(2.447)	(2.318)	(2.329)
Avg. originality		-1.089	-7.030***	-4.980***	-3.557**
		(0.741)	(1.961)	(1.717)	(1.588)
Avg. quality x Avg. originality			11.341***	9.123***	8.726***
			(3.161)	(3.207)	(2.943)
Avg. rel. time				-4.925***	-4.295***
				(0.403)	(0.378)
Avg. rel. quality				-0.413	-0.222
				(0.332)	(0.437)
Avg. idea time					-0.047***
					(0.005)
Avg. idea quality					-1.638
					(1.473)
Constant	9.826***	11.488***	15.839***	13.818***	19.390***
	(0.271)	(0.809)	(1.613)	(1.366)	(1.361)
Observations	897	897	897	895	895
R-squared	0.033	0.039	0.054	0.523	0.613

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper.

Table A18: Regression of avg. quality (online)

	Avg. quality				
	(1)	(2)	(3)	(4)	(5)
No PR	-0.006	-0.011	-0.011	-0.008	-0.005
	(0.016)	(0.015)	(0.014)	(0.011)	(0.006)
Innovation-Weighted PR	0.002	-0.001	-0.001	-0.003	-0.009
	(0.015)	(0.014)	(0.014)	(0.011)	(0.006)
Number of ideas		-0.003**	0.011*	0.005	0.002
		(0.001)	(0.006)	(0.005)	(0.003)
Avg. originality		-0.361***	-0.222***	-0.317***	-0.026
		(0.031)	(0.073)	(0.056)	(0.038)
Number of ideas x Avg. originality			-0.021**	-0.007	-0.002
			(0.008)	(0.006)	(0.004)
Avg. rel. time				0.011	0.005
				(0.018)	(0.008)
Avg. rel. quality				0.287***	0.261***
				(0.026)	(0.015)
Avg. idea time					0.000
					(0.000)
Avg. idea quality					0.994***
					(0.034)
Constant	0.491***	0.759***	0.666***	0.694***	-0.013
	(0.011)	(0.026)	(0.051)	(0.045)	(0.041)
Observations	897	897	897	895	895
R-squared	0.000	0.175	0.183	0.478	0.834

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper.

Table A19: Regressions of avg. originality (online)

	Avg. originality				
	(1)	(2)	(3)	(4)	(5)
No PR	0.001	-0.006	-0.006	-0.008	-0.003
	(0.018)	(0.016)	(0.017)	(0.016)	(0.015)
Innovation-Weighted PR	0.007	0.004	0.004	0.002	0.011
	(0.017)	(0.016)	(0.016)	(0.015)	(0.014)
Avg. quality		-0.476***	-0.453***	-0.705***	-0.235*
		(0.036)	(0.096)	(0.088)	(0.126)
Number of ideas		-0.002	-0.001	-0.002	0.004
		(0.002)	(0.004)	(0.005)	(0.004)
Number of ideas x Avg. quality			-0.003	0.002	0.002
			(0.010)	(0.009)	(0.008)
Avg. rel. time				0.006	0.014
				(0.027)	(0.025)
Avg. rel. quality				0.206***	0.076**
				(0.025)	(0.029)
Avg. idea time					0.001***
					(0.000)
Avg. idea quality					-0.586***
					(0.116)
Constant	0.655***	0.912***	0.901***	1.008***	0.832***
	(0.012)	(0.026)	(0.046)	(0.050)	(0.070)
Observations	897	897	897	895	895
R-squared	0.000	0.172	0.172	0.250	0.314

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper.

B.6 Robustness to inclusion of ideators without recurring ideas

For the online experiment we can again show that our results are robust to the inclusion of the full sample. In the online experiment, 92% of all generated ideas are recurring ideas. Overall, again only a very small fraction of the sample is excluded from the analysis due to the lack of recurring ideas. Specifically, there are five ideators who have no recurring ideas. In Table A20, Table A21, Table A22, and Table A23, we again provide results from regressions on performance dimensions for the full sample, and excluding the ideators without recurring ideas. As in the lab experiment, coefficients, if anything, only marginally change. Thus, our results are robust to the inclusion of the full sample and are not driven by the exclusion of few ideators without recurring ideas.

Table A20: Regression of number of ideas of full sample vs. subsample (online)

	Number of ideas					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	-1.651***	-1.662***	-1.614***	-1.675***	-1.684***	-1.623***
	(0.348)	(0.347)	(0.343)	(0.349)	(0.347)	(0.343)
Innovation-Weighted PR	-1.590***	-1.579***	-1.580***	-1.675***	-1.668***	-1.647***
	(0.355)	(0.354)	(0.353)	(0.356)	(0.355)	(0.353)
Avg. quality		-1.933**	-9.942***		-1.484*	-10.888***
		(0.828)	(2.447)		(0.831)	(2.419)
Avg. originality		-1.089	-7.030***		-1.205	-8.070***
		(0.741)	(1.961)		(0.742)	(1.938)
Avg. quality x Avg. originality			11.341***			13.158***
			(3.161)			(3.104)
Constant	9.826***	11.488***	15.839***	9.826***	11.343***	16.428***
	(0.271)	(0.809)	(1.613)	(0.271)	(0.809)	(1.606)
Observations	897	897	897	902	902	902
R-squared	0.033	0.039	0.054	0.034	0.039	0.060

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A21: Regression of avg. quality of full sample vs. subsample (online)

	Avg. quality					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	-0.006	-0.011	-0.011	-0.008	-0.011	-0.011
	(0.016)	(0.015)	(0.014)	(0.016)	(0.015)	(0.015)
Innovation-Weighted PR	0.002	-0.001	-0.001	-0.004	-0.004	-0.005
	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)
Number of ideas		-0.003**	0.011*		-0.003*	0.008
		(0.001)	(0.006)		(0.001)	(0.006)
Avg. originality		-0.361***	-0.222***		-0.375***	-0.273***
		(0.031)	(0.073)		(0.031)	(0.074)
Number of ideas x Avg. originality			-0.021**			-0.015*
			(0.008)			(0.008)
Constant	0.491***	0.759***	0.666***	0.491***	0.761***	0.693***
	(0.011)	(0.026)	(0.051)	(0.011)	(0.026)	(0.051)
Observations	897	897	897	902	902	902
R-squared	0.000	0.175	0.183	0.000	0.183	0.187

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A22: Regression of avg. originality of full sample vs. subsample (online)

	Avg. originality					
	Subsample			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
No PR	0.001	-0.006	-0.006	0.002	-0.006	-0.006
	(0.018)	(0.016)	(0.017)	(0.018)	(0.016)	(0.016)
Innovation-Weighted PR	0.007	0.004	0.004	0.012	0.005	0.005
	(0.017)	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)
Avg. quality		-0.476***	-0.453***		-0.483***	-0.477***
		(0.036)	(0.096)		(0.035)	(0.088)
Number of ideas		-0.002	-0.001		-0.003*	-0.002
		(0.002)	(0.004)		(0.002)	(0.004)
Number of ideas x Avg. quality			-0.003			-0.001
			(0.010)			(0.009)
Constant	0.655***	0.912***	0.901***	0.655***	0.918***	0.915***
	(0.012)	(0.026)	(0.046)	(0.050)	(0.070)	(0.042)
Observations	897	897	897	902	902	902
R-squared	0.000	0.172	0.172	0.001	0.182	0.182

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(3) show results based on the subsample of ideators who have at least one recurring idea. Col. (4)-(6) show results based on the full sample of ideators.

Table A23: Regression of number of innovative ideas of full sample vs. subsample (online)

	DV: Number of innovative ideas							
	Subsample				Full sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No PR	-0.333**	0.057	0.131	0.103	-0.340**	0.057	0.131	0.102
	(0.154)	(0.135)	(0.086)	(0.064)	(0.154)	(0.134)	(0.086)	(0.064)
Innovation-Weighted PR	-0.249	0.126	0.121	0.041	-0.279*	0.118	0.130	0.045
	(0.155)	(0.128)	(0.087)	(0.065)	(0.155)	(0.128)	(0.087)	(0.064)
Number of ideas		0.236***	0.259***	-0.588***		0.237***	0.257***	-0.590***
		(0.016)	(0.012)	(0.035)		(0.016)	(0.012)	(0.035)
Avg. quality			6.474***	-3.112***			6.401***	-3.097***
			(0.255)	(0.407)			(0.252)	(0.408)
Avg. originality			4.163***	-2.306***			4.177***	-2.184***
			(0.224)	(0.287)			(0.223)	(0.278)
Avg. quality x Avg. originality				4.594***				4.431***
				(0.434)				(0.426)
Number of ideas x Avg. quality				0.899***				0.908***
				(0.039)				(0.038)
Number of ideas x Avg. originality				0.633***				0.628***
				(0.035)				(0.035)
Constant	2.538***	0.219	-5.912***	1.645***	2.538***	0.205	-5.862***	1.613***
	(0.116)	(0.160)	(0.289)	(0.270)	(0.116)	(0.159)	(0.285)	(0.270)
Observations	897	897	897	897	902	902	902	902
R-squared	0.006	0.287	0.690	0.831	0.006	0.292	0.691	0.832

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR, and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the exact definition of the other measures see Table 1 of the main paper. Col. (1)-(4) show results based on the subsample of ideators who have at least one recurring idea. Col. (5)-(8) show results based on the full sample of ideators.

B.7 Complexity of recurring ideas

Table A24 provides summary statistics on the avg. idea complexity and the avg. rel. complexity of recurring ideas. When referring to complexity, we refer to the number of materials used in an idea. We define avg. idea complexity and avg. rel. complexity for recurring ideas in a manner that is similar to how we define avg. idea quality and avg. rel. quality of recurring ideas. That is, avg. idea complexity is the average complexity across all recurring ideas generated by an ideator. Where idea complexity is the average complexity across all illustrations of the same word in our sample. We define avg. rel. complexity as the average of rel. complexity for all recurring ideas (j) generated by individual (i). Where $rel.complexity_{j\,k} = \frac{(complexity_{jk} - idea\ complexity_j)}{idea\ complexity_j}$.

Equivalent, but more pronounced as in the lab experiment, we find that PR incentives significantly reduce the relative complexity of ideas, while stronger aligned incentives offset this effect. The average complexity of ideas in No PR is significantly higher than in Unweighted PR (MWU-test: p<0.01). Adding an innovation-weight leads to significantly more complex illustrations compared to unweighted PR incentives (MWU-test: p<0.01), such that the relative complexity is not different from No PR anymore (MWU-test: p=0.69).

Table A24: Complexity of recurring ideas (online)

	No PR	Unweighted PR	Innovation-Weighted PR
Avg. complexity	6.20**	5.99	5.89$^{n.s.}$
	(1.30)	(1.23)	(1.13)
Avg. rel. complexity	0.08***	0.01	0.09***
	(0.32)	(0.25)	(0.29)
N ideators illustrating recurring ideas	247	256	256

Note: We provide means and standard deviations ins parentheses. Superscripts refer to results of MWU-test comparing Unweighted PR to the other treatments, where *** p<0.01, ** p<0.05, * p<0.1 and $^{n.s.}$ p>0.10. All results for recurring ideas are based on a limited sample of those ideators who illustrate at least one recurring idea.

B.8 Idea selection

Why are ideators worse in selecting their best idea in Unweighted PR? Our earlier results demonstrated that ideators' average performance does not differ between treatments. However, ideators in Unweighted PR generate significantly more ideas than ideators in other treatments.

We analyze the deviation in value (product of originality and quality) of the selected idea from the best in more detail using regression analysis using OLS regression where this difference is the dependent variable. Unweighted PR serves as the base category, such that positive coefficients of the treatment indicators show that the difference between the value of the selected idea from the best idea becomes smaller, and thus, ideators are better at selecting a high value idea.

The positive coefficients in Col. (1) replicates that the deviation in value from the best idea is smaller in No PR and Innovation-Weighted PR than in Unweighted PR and thus, that ideators are worse at selecting a high value idea under Unweighted PR incentives.

Controlling for number of ideas in Col. (2) explains these differences, suggesting that having more ideas to select from makes it harder for ideators to select the best idea.

Controlling for performance across other dimensions in Col. (3)–Col. (4) does not further affect coefficients of the treatment indicators, suggesting that the difference is mainly driven by the number of ideas, which is in line with the absence of performance difference along other dimensions.

Table A25: Regression of difference in value of selected ideas

	Diff. in value			
	(1)	(2)	(3)	(4)
No PR	0.072**	0.041	0.038	0.037
	(0.030)	(0.029)	(0.029)	(0.029)
Innovation-Weighted PR	0.064**	0.033	0.033	0.032
	(0.030)	(0.029)	(0.029)	(0.029)
Number of ideas		-0.019***	-0.020***	-0.030*
		(0.003)	(0.003)	(0.016)
Avg. quality			-0.179***	-0.101
			(0.064)	(0.218)
Avg. originality			0.001	0.085
			(0.057)	(0.163)
Avg. quality x Avg. originality				-0.247
				(0.228)
Number of ideas x Avg. quality				0.014
				(0.017)
Number of ideas x Avg. originality				0.007
				(0.017)
Constant	-0.438***	-0.251***	-0.160**	-0.183
	(0.021)	(0.033)	(0.068)	(0.153)
Observations	863	863	863	863
R-squared	0.008	0.059	0.068	0.070

Note: Coefficients are from an OLS regression. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted reference category is Unweighted PR. No PR and Innovation-Weighted PR are dummy variables that equal 1 for the respective treatments and 0 otherwise. For the definition of the other measures see Table 1 of the main paper.

Appendix C: Experimental Instructions Laboratory Experiment

C.1 Lab experiment instructions (translation from German)

Instructions

Welcome to this experiment!

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

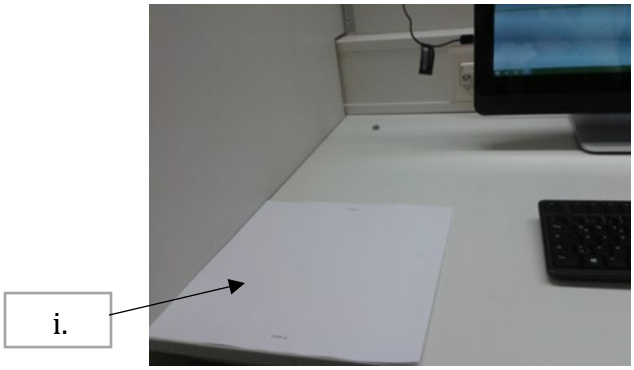
Task. – Immediately before the start of the task, you will receive various materials. The task consists of illustrating words with the provided set of materials. The goal is:

- To illustrate as many different words as possible,
- The words must be identifiable by others,
- And the illustrated words should be unique, meaning that they were not illustrated by any of the ideators in the randomly selected four-person group.

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

Please proceed to illustrate each word in the following manner:

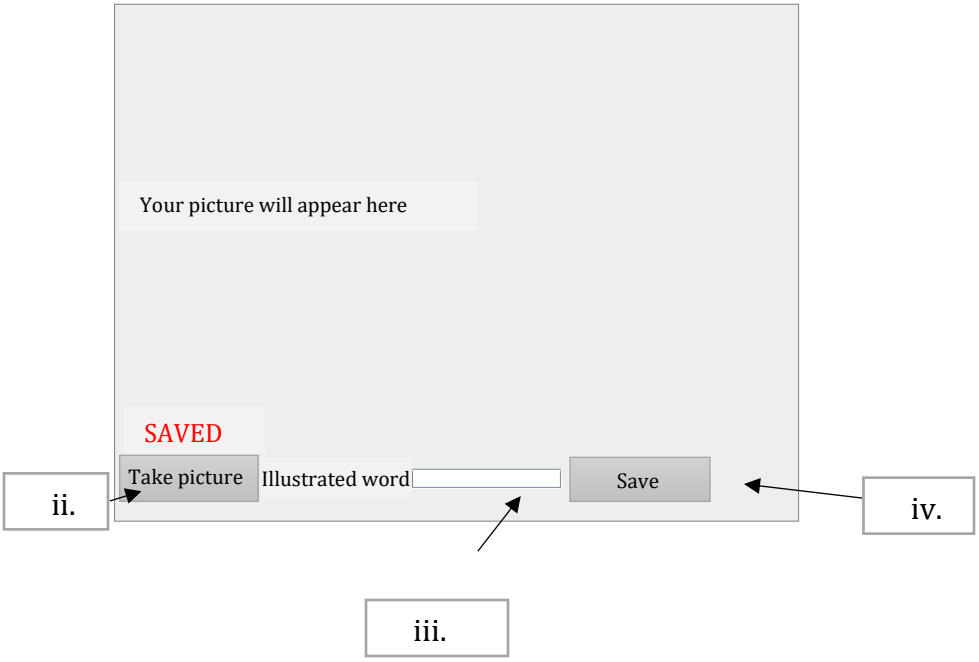
i. Illustrate the word in the designated area using the provided materials.



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.



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. Once this time has expired, we ask that you fill out the questionnaire before the end of the experiment.

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

Fixed pay treatment: You will be paid €10 for this task. In addition, you will receive a show-up payment of €2.50. Your payment will be received 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.

Unweighted PR: You will be paid €0.60 for each admissible word that you illustrate. You will 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.

Quality-Weighted PR: 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 the exact same word that you saved along with the respective picture.

Each word will be presented to 10 other people. We will measure how many of these 10 people correctly identify the respective word. For each illustrated word, you will be 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 10 people. In addition, you will be paid 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.

Originality-Weighted PR: After this experiment, you will be randomly assigned to a group of four people who also participated in the same experiment. For each admissible word that you alone in the group illustrated, you will be paid €0.85. If at least one other person in the group illustrated the same word, then you will receive €0 for illustrating this word. In addition, you will 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.

C.2 Instructions for online survey to assess (materialize) quality (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 via e–mail: 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 ideators in a prior experiment. These ideators' 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 ideator 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 ideator, then you will 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 ideators. This means that it is possible to see more than one illustration of the same word.

Payment. – You will receive your payment only if you complete the entire experiment. You will receive €2 for participating in the experiment. In addition, you will receive €0.10 for each picture that you correctly identify. At the end of the experiment, you can choose whether you would like to have an electronic transfer or pick up the payment in cash.

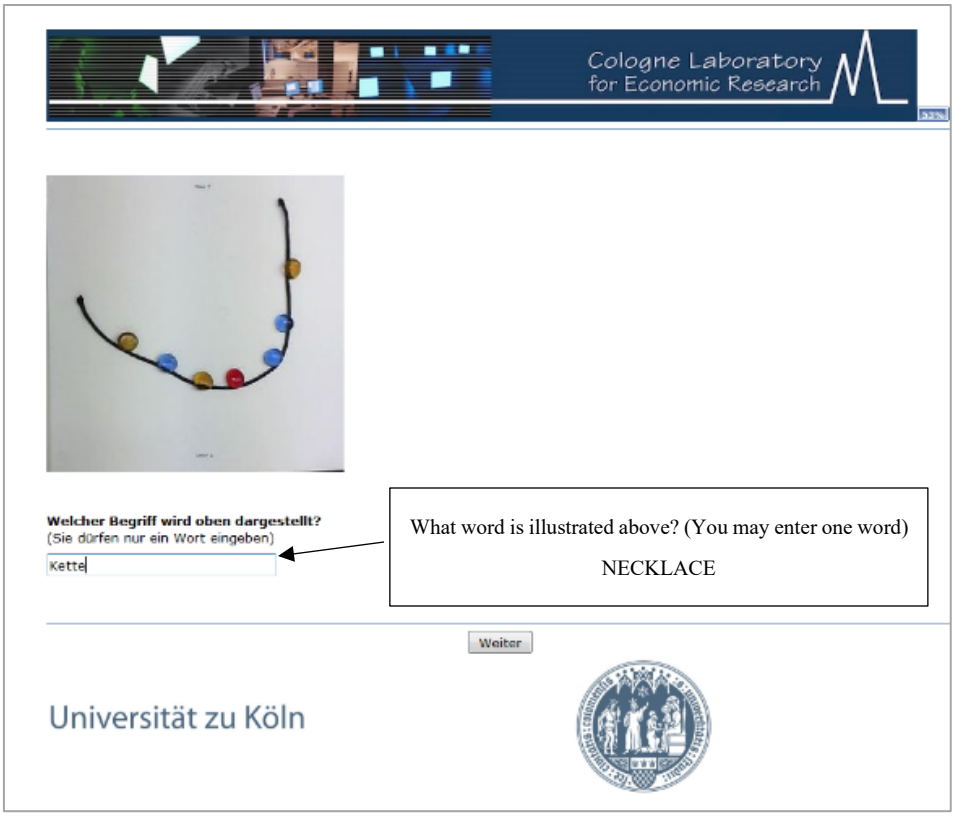


Figure A9: Screen of questionnaire (example)

Appendix D: Experimental Instructions Online Experiment

D.1 Instructions for ideators

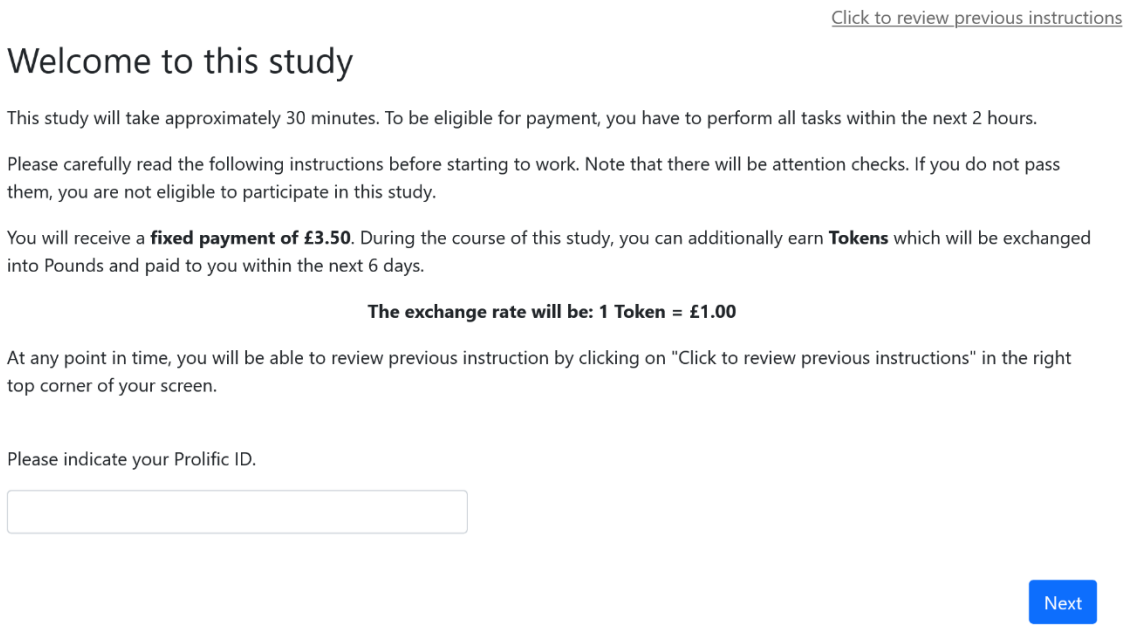


Figure A10: Treatment Fix Pay: Screen 1 of 14

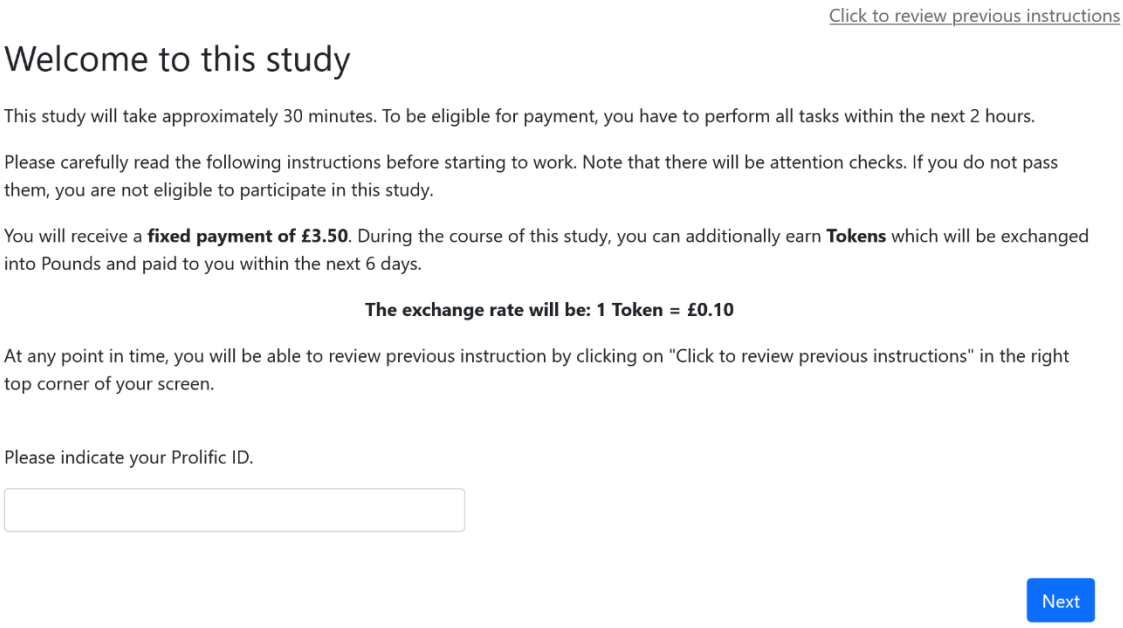


Figure A11: Treatment Piece Rate: Screen 1 of 14

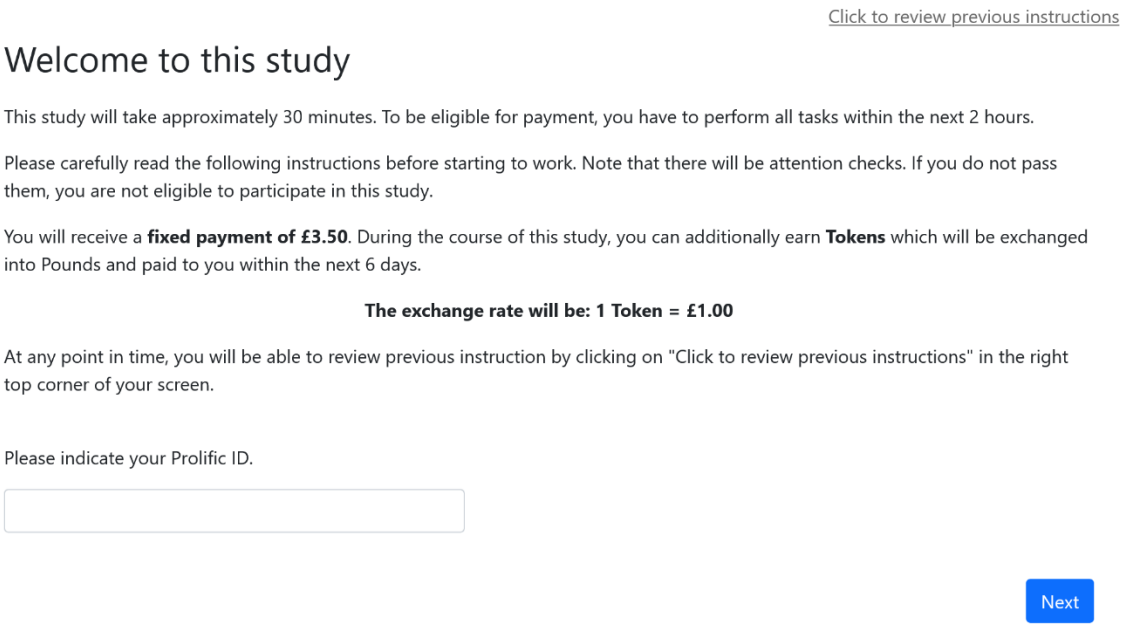


Figure A12: Treatment Innovation Piece Rate: Screen 1 of 14

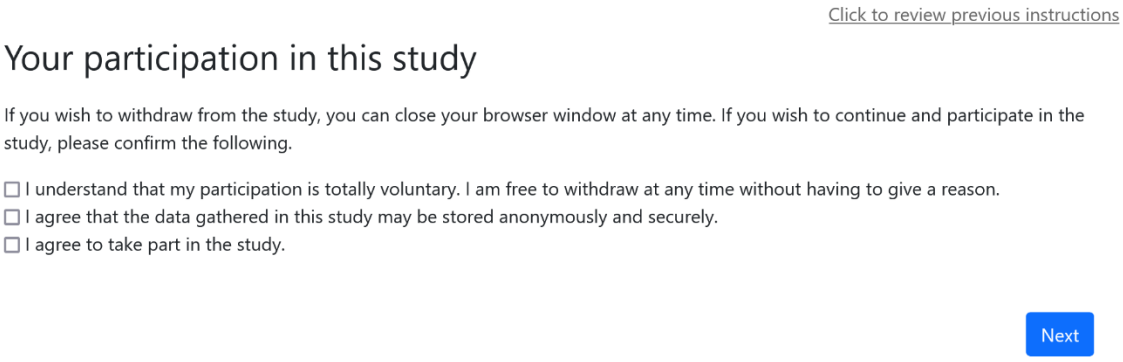


Figure A13: Screen 2 of 14

Click to review previous instructions

Technology

On the next page you will see objects that you can move, resize and rotate in the following ways:

Move: Select the object you want to **move**, **drag it** and **drop it**.

Rotate: Select the object you want to rotate. A control box will appear (see Figure 1). To rotate objects, click on the **green circle** on the top and **rotate it**.

Resize and rescale: Select the object you want to resize. A control box will appear (see Figure 1). **You can resize objects by dragging the blue circles** in the desired direction.

Change layer: **Right-click** on an object to **change its layer**. A menu will appear (see Figure 2), in which you can select between different options. To bring an object to the front of all overlapping elements, click 'Send to Front'. To send an object to the back of all overlapping elements, click 'Send to Back'. To bring an object one step closer to the front, click 'Bring Forward'. To send an object one step toward the back, click 'Send Backward'.

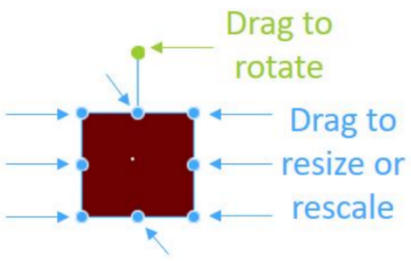


Figure 1: Control box (on click)

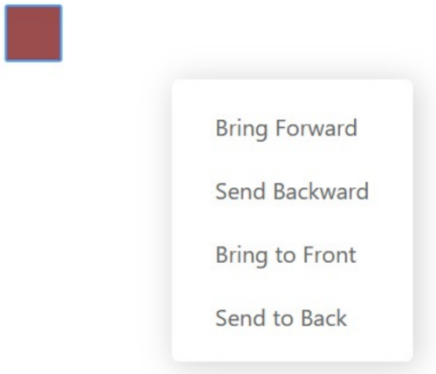


Figure 2: Menu (on right click)

Next

Figure A14: Screen 3 of 14

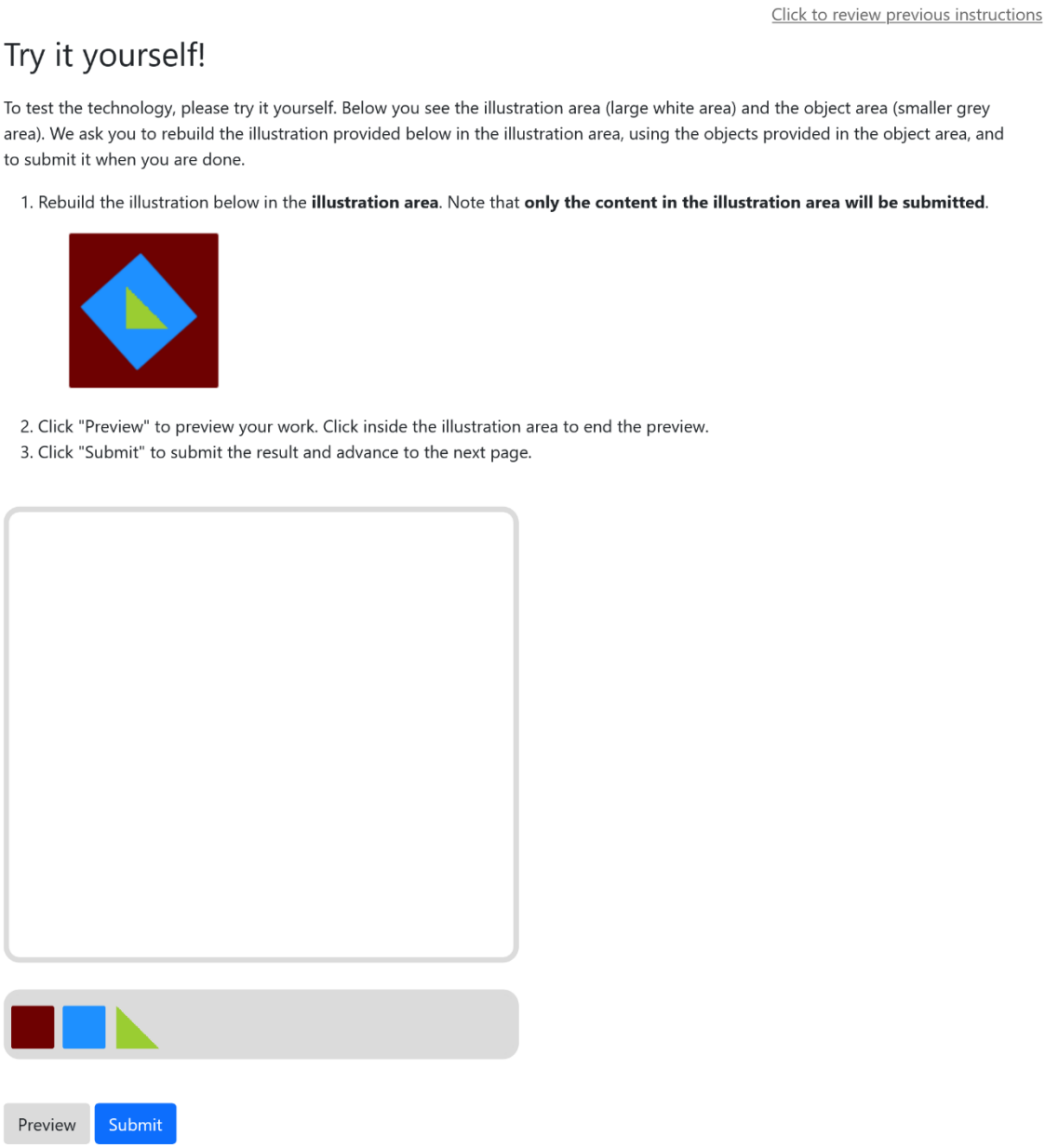


Figure A15: Screen 4 of 14

Click to review previous instructions

Your task

Task
Your task is to think of a word and illustrate it in the illustration area. To do so, you can use all or some of the objects provided in the object area.

The goal is to submit **as many innovative illustrations as possible**.

Innovative illustrations are:

- **original**, that is, the illustrated word is not among a random selection of 100 illustrations by other workers in this task.
- **and** can be **identified** by a relatively large fraction of individuals, who only see the illustration and have to identify the illustrated word. The more individuals identify the illustration, the better the illustration.

Specifically, an illustration is considered **innovative** if it is among the **best 10%** illustrations submitted by workers in this task.

Time
You have a total of **20 minutes** for this task. Only illustrations **submitted** within these 20 minutes will be considered for payment.

Payment
You will be paid **10 Tokens** for this task.

Next

Figure A16: Treatment Fix Pay: Screen 5 of 14

Click to review previous instructions

Your task

Task
Your task is to think of a word and illustrate it in the illustration area. To do so, you can use all or some of the objects provided in the object area.

The goal is to submit **as many innovative illustrations as possible**.

Innovative illustrations are:

- **original**, that is, the illustrated word is not among a random selection of 100 illustrations by other workers in this task.
- **and** can be **identified** by a relatively large fraction of individuals, who only see the illustration and have to identify the illustrated word. The more individuals identify the illustration, the better the illustration.

Specifically, an illustration is considered **innovative** if it is among the **best 10%** illustrations submitted by workers in this task.

Time
You have a total of **20 minutes** for this task. Only illustrations **submitted** within these 20 minutes will be considered for payment.

Payment
You will be paid **10 Tokens** for each submitted illustration. You will receive the payment for each submitted illustration that follows the above specified rules.

Next

Figure A17: Treatment Piece Rate: Screen 5 of 14

Click to review previous instructions

Your task

Task
Your task is to think of a word and illustrate it in the illustration area. To do so, you can use all or some of the objects provided in the object area.

The goal is to submit **as many innovative illustrations as possible**.

Innovative illustrations are:

- **original**, that is, the illustrated word is not among a random selection of 100 illustrations by other workers in this task.
- **and** can be **identified** by a relatively large fraction of individuals, who only see the illustration and have to identify the illustrated word. The more individuals identify the illustration, the better the illustration.

Specifically, an illustration is considered **innovative** if it is among the **best 10%** illustrations submitted by workers in this task.

Time
You have a total of **20 minutes** for this task. Only illustrations **submitted** within these 20 minutes will be considered for payment.

Payment
You will be paid **10 Tokens** for each submitted **innovative** illustration. You will receive the payment for each submitted innovative illustrations that follows the above specified rules.

Next

Figure A18: Treatment Innovation Piece Rate: Screen 5 of 14

Click to review previous instructions

Rules

Rules
When illustrating words, there are some **rules**:

- Work inside the illustration area.
- In the illustration area, you should illustrate words and not write them.
- Your illustrations may not include any symbol that is depicted on the keyboard (for example, illustrations that include "➔", "8", "b", "@", ">" or "+" are not permitted).
- Illustrate only one word at a time. 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 political, offensive, or inappropriate content.
- The word you illustrate should be included in **a standard American English dictionary**.

After this task, we will evaluate if your illustrations followed the rules.

Next

Figure A19: Screen 6 of 14

Click to review previous instructions

Attention check

This is an attention check. If you pass it, you will advance on the next page of the task. If you do not pass it, you cannot participate in this task.

What is your **goal** in this task?

- ○ Submit as many innovative illustrations as possible
- ○ Illustrate provided words in an identifiable way
- ○ Identify words based on others' illustrations
- ○ Submit one illustration that is as innovative as possible

Innovative illustrations ...

- ○ ...illustrate words that are illustrated by a relatively large fraction of other workers in this task
- ○ ...are original and can be identified by a relatively large fraction of individuals
- ○ ...are identifiable but not original
- ○ ...are among the top 10% most colorful illustrations

The illustrated word is **original** if...

- ○ ...at least 10% of other workers from this task illustrated the same word
- ○ ...a large fraction of individuals identify the illustrated word based on the illustration only
- ○ ...the illustrated word is not in a standard American English dictionary
- ○ ...the illustrated word is not among a random selection of 100 illustrations by other workers from this task

What determines how **identifiable** an illustration is?

- ○ The fraction of workers in this task who illustrated the same word
- ○ The fraction of individuals who identify the illustrated word based on the illustration only
- ○ The fraction of individuals who find the illustration aesthetic
- ○ The fraction of individuals who did not like the illustrated word

Your **payment** in the following task is

- ○ 10 Tokens
- ○ 10 Token for each submitted innovative illustration
- ○ 10 Tokens for each submitted illustrations
- ○ There is no payment for the following task

Next

Figure A20: Screen 7 of 14

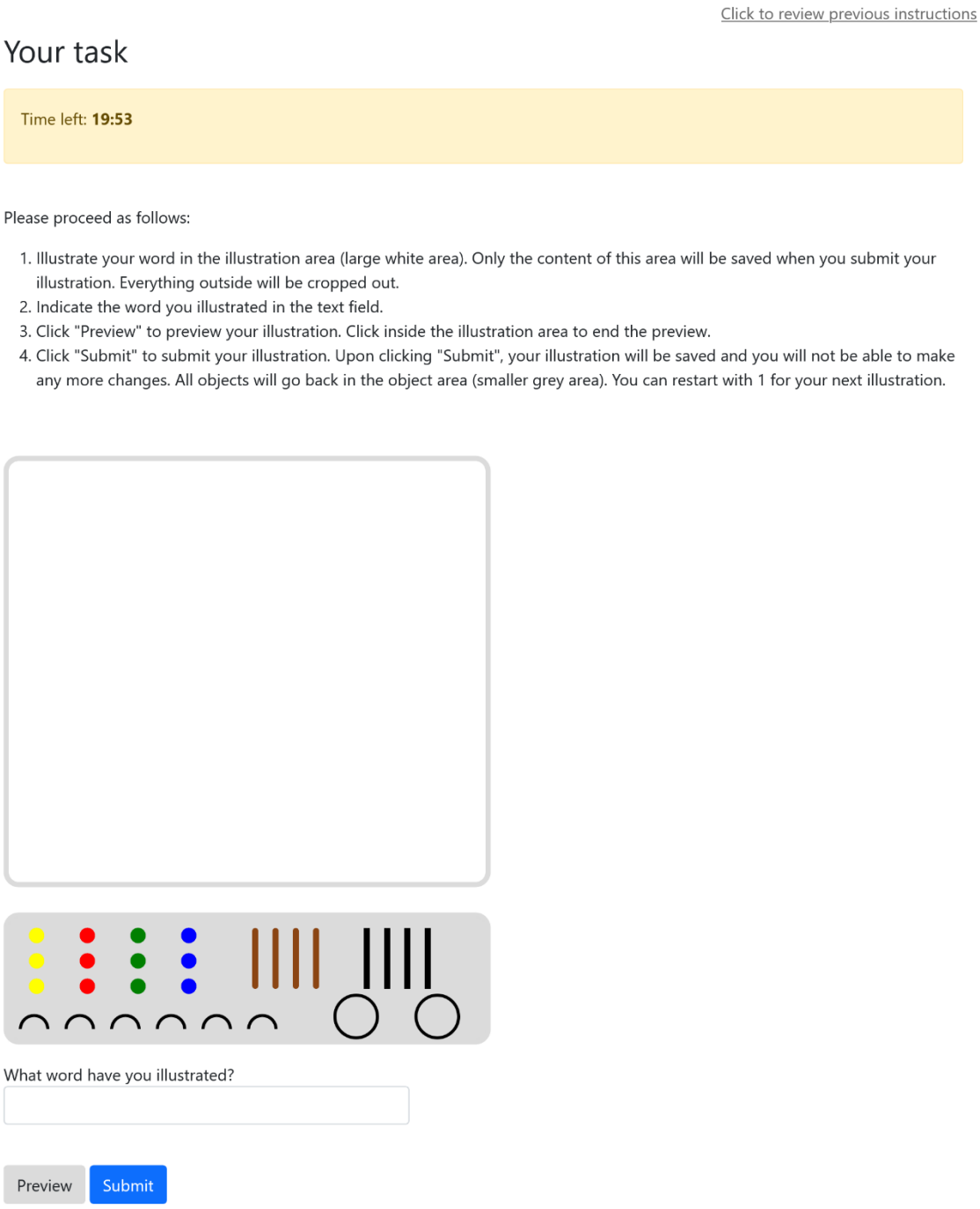


Figure A21: Screen 8 of 14

Click to review previous instructions

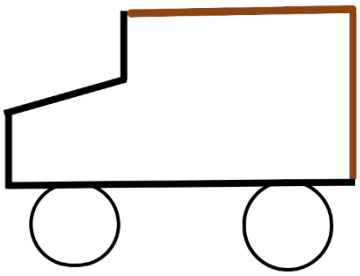
Earn an extra bonus!

Your task
Please select your best innovative illustration.

Your extra bonus payment
Your extra bonus payment may depend on whether your selected illustration is **innovative**.

Remember that **innovative** illustrations are **original** (i.e. the illustrated word is not among a random selection of 100 illustrations by other workers in this task), and can be **identified** by a relatively large fraction of individuals, who only see the illustration and have to identify the illustrated word.

Car



sun

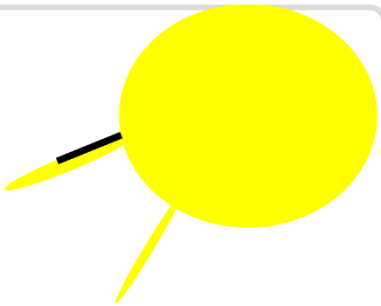


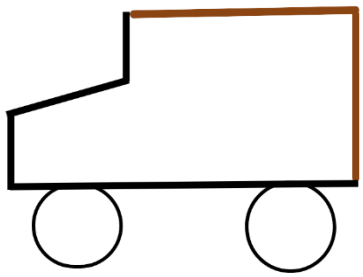
Figure A22: Screen 9 of 14

Click to review previous instructions

Earn an extra bonus!

Are you sure that you want to select the below illustration?

Car



Go Back Confirm

Figure A23: Screen 10 of 14

Click to review previous instructions

Your extra bonus payment

Now you can choose which rule should determine your extra bonus payment.

Under **Rule 1**, your extra bonus payment is:

- **£1.00** if your illustration is **innovative**.
- **£0.00** otherwise.

Under **Rule 2**, your extra bonus payment is:

- **£0.50** regardless of whether your illustration is **innovative** or not.

Please choose:

- ○ Rule 1
- ○ Rule 2

Next

Figure A24: Screen 11 of 14

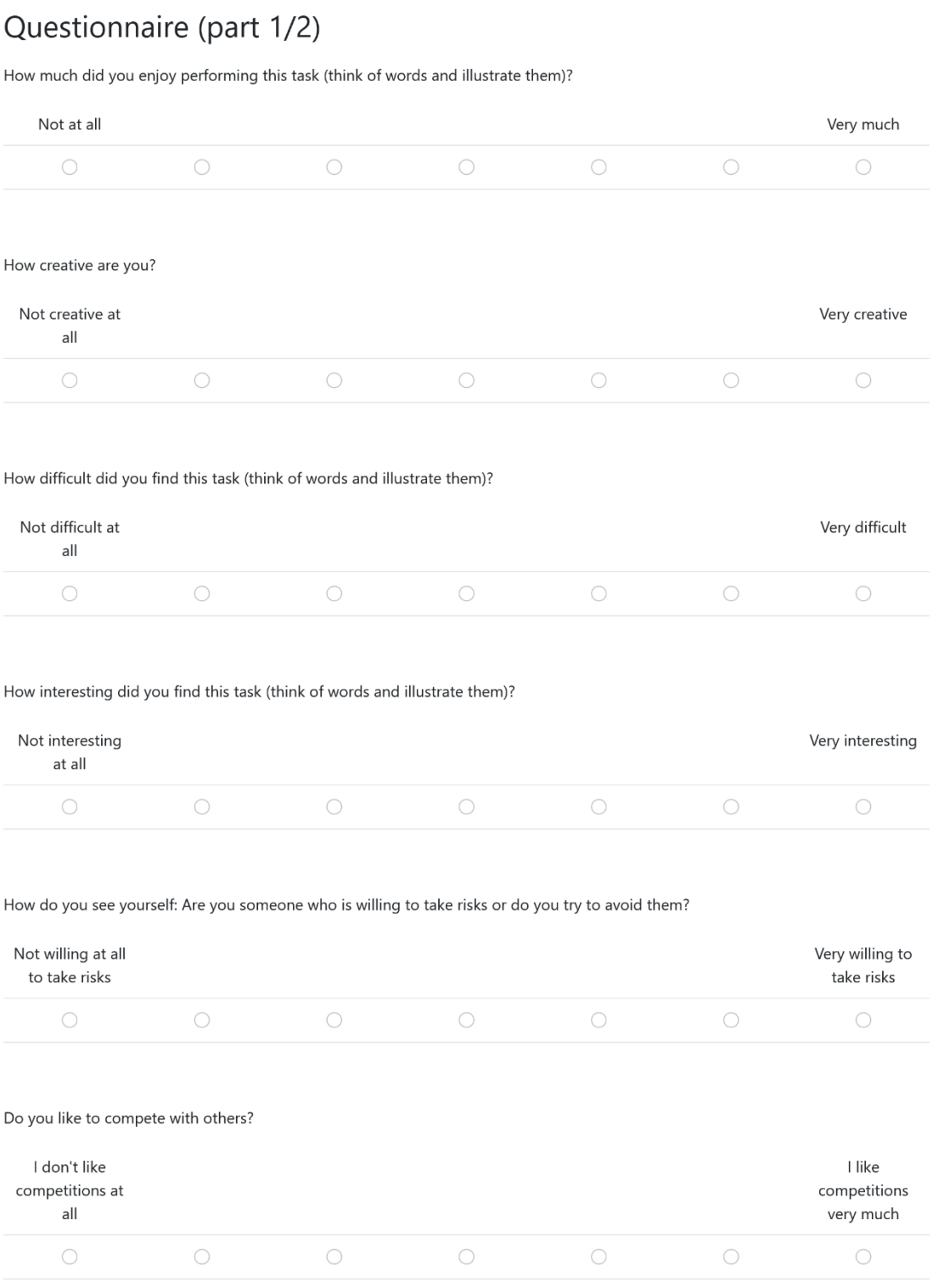


Figure A25: Screen 12 of 14

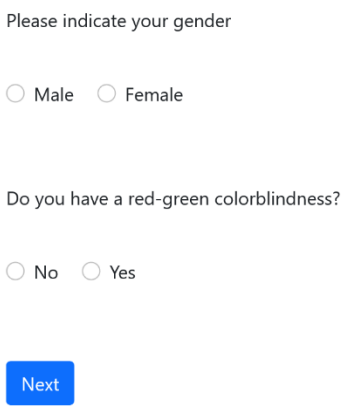


Figure A26: Screen 12 of 14

Questionnaire (part 2/2)

Please use the following 1-5 scale to indicate how often each phrase describes your thinking. Keep in mind that your responses should reflect your own thoughts, rather than how you think you should act.

Please select the answer that is closest to the truth for each item.

	1 Never	2 About once a year	3 About once or twice a month	4 About once or twice a week	5 About every day and sometimes more than once a day
I have ideas for arranging or rearranging the furniture at home.	○	○	○	○	○
I have ideas for making my work easier.	○	○	○	○	○
I think of a new, better, or funny name for something that already has a name.	○	○	○	○	○
I have ideas about what I will be doing in the future.	○	○	○	○	○
I consider alternative careers (or career changes).	○	○	○	○	○
I have trouble sleeping at night, so many ideas keep showing themselves keep me awake.	○	○	○	○	○

	1 Never	2 About once a year	3 About once or twice a month	4 About once or twice a week	5 About every day and sometimes more than once a day
I make plans (e.g., going to a particular restaurant or movie), but something messes it up - yet it is easy for me to find something to do instead.	○	○	○	○	○
I have ideas about a good plot for a movie or TV show.	○	○	○	○	○
I have ideas about a new invention.	○	○	○	○	○
I have ideas for stories or poems.	○	○	○	○	○
I have ideas about a new route between home and school (or work).	○	○	○	○	○
I have ideas for a new business or product.	○	○	○	○	○

Figure A27: Screen 13 of 14

	1 Never	2 About once a year	3 About once or twice a month	4 About once or twice a week	5 About every day and sometimes more than once a day
I see a cloud, shadow, or similar ambiguous figure and have several ideas about what the shape or figure could be.	○	○	○	○	○
I have ideas about what I will be doing 10 years from now.	○	○	○	○	○
I have trouble staying with one topic when writing letters or mails because I think of so many things to say.	○	○	○	○	○
I often see people and think about alternative interpretations of their behavior.	○	○	○	○	○
When reading books or stories I have ideas of better endings.	○	○	○	○	○
When reading the newspaper or a letter that someone wrote, I often have ideas for better wording.	○	○	○	○	○
I hear songs and think of different or better lyrics.	○	○	○	○	○

Next

Figure A28: Screen 13 of 14

You have completed the study

Thank you for participating in this study. Your completion code is CBMSQFOS. You will receive your bonus payment, if applicable, within the next 6 days.

Figure A29: Screen 14 of 14

D.2 Instructions for online survey to assess (materialize) quality

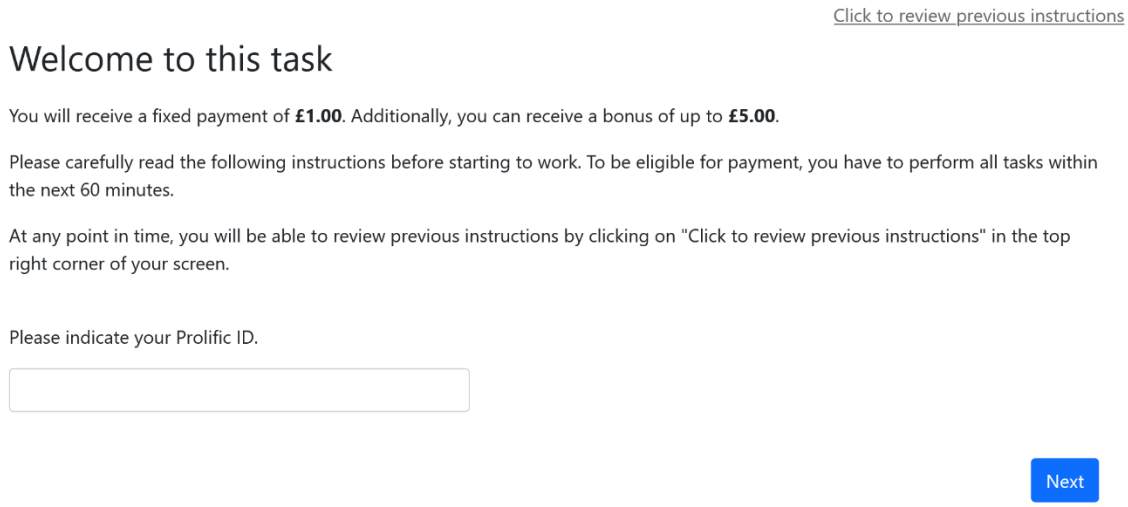


Figure A30: Screen 1 of 54

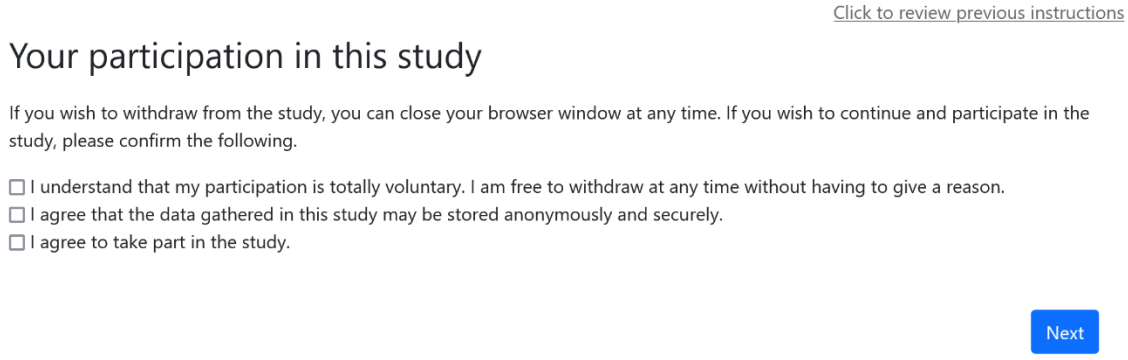


Figure A31: Screen 2 of 54

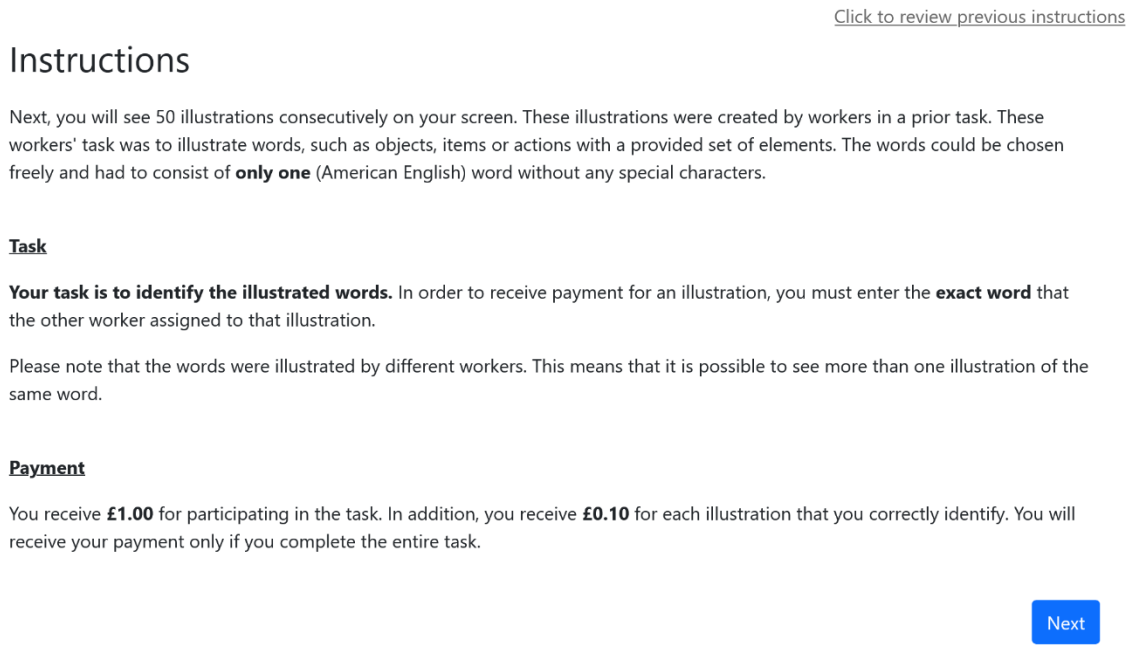


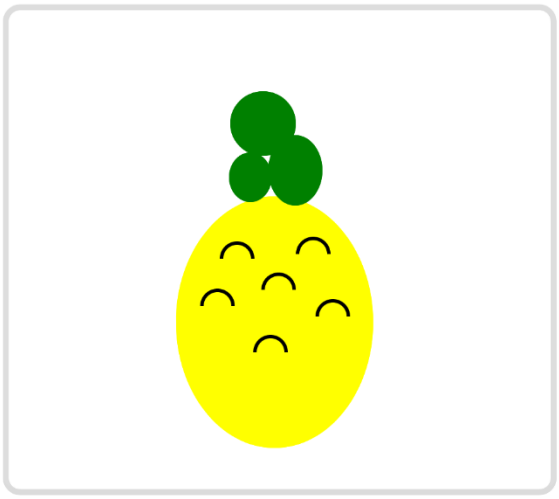
Figure A32: Screen 3 of 54

Click to review previous instructions

Task

You receive **£0.10** for each correctly identified word. A word is correctly identified if you enter the **exact word** that the other worker assigned to that illustration.

Please write down the word that is illustrated below.



Next

Figure A33: Screen 4 – 53 of 54

You have completed the study

Thank you for participating in this study. Your completion code is CBMSQFOS. You will receive your bonus payment, if applicable, within the next 6 days.

Figure A34: Screen 54 of 54