

Blaufus, Kay; Piehl, Kevin; Schröder, Marina

Conference Paper

Input versus Output Incentives in Idea Generation - An Experimental Analysis

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2025: Revival of Industrial Policy

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Blaufus, Kay; Piehl, Kevin; Schröder, Marina (2025) : Input versus Output Incentives in Idea Generation - An Experimental Analysis, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2025: Revival of Industrial Policy, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/325392>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Input versus Output Incentives in Idea Generation

An Experimental Analysis^a

Kay Blaufus ^b

Kevin Piehl ^c

Marina Schröder ^d

February 24, 2025

Abstract

In an experimental idea generation context, we study the effect of input incentives, reward-time invested to generate ideas, and output incentives, rewarding the number of innovative ideas generated. Compared to a fixed wage, we find that both input and output incentives increase the average number of innovative ideas to a similar extent. We can show that the mechanisms behind this increase in innovative ideas varies. The improvement under input incentives arises merely due to an increase in the time spent generating ideas. The improvement under output incentives arises due to both an increase in the time spent and in the number of ideas generated per unit of time. We can show that the increase in the number of ideas generated per unit of time is associated with a shift in the types of ideas generated, i.e., with the generation of less complex ideas.

JEL codes: C90, J33, O31

Keywords: Idea generation, incentives, real-effort experiment, complexity of ideas

^aThe experimental design was approved by the institutional review board of the German Association for Experimental Economic Research (<https://gfew.de/ethik/u4MfwQy5>). The study is preregistered under <https://aspredicted.org/4DM.25D>. We are thankful for extremely helpful comments from Katharina Brütt, Marina Chugunova, E. Glenn Dutcher, Joshua Graff Zivin, Heidrun C. Hoppe-Wewetzer, Felix Kölle, Julia Nafziger, Markus Nagler, participants at the 2022 conference of the German Association for Experimental Economic Research in Salzburg, the 2023 Economic Science Association World Meeting in Lyon, the 2024 Reading Group in Innovation Studies Summer School in Madrid, seminar participants at the Max Planck Institute for Innovation and Competition in Munich, Aarhus University, Ulm University, and workshop participants at the Leibniz University Hanover and the University of Hamburg. This work was supported by the German Research Foundation [project number 492695807].

^bblaufus@steuern.uni-hannover.de; Leibniz University Hanover, Institute of Business Taxation, Königsworther Platz 1, 30167 Hanover, Germany.

^cpiehl@inec.uni-hannover.de; Leibniz University Hanover, Institute of Innovation Economics, Königsworther Platz 1, 30167 Hanover, Germany.

^dschroeder@inec.uni-hannover.de; Leibniz University Hanover, Institute of Innovation Economics, Königsworther Platz 1, 30167 Hanover, Germany and CESifo

1 Introduction

The starting point of any innovation is an idea. Not any idea, but an innovative idea, i.e., an idea that is at the same time good and original. To foster innovation, it is crucial to achieve a better understanding of factors that have an impact on idea generation, e.g., to understand how idea generation can be fostered using monetary incentives. Previous research using both field and lab data provides evidence for a positive effect of output incentives on idea generation (Lam, 2011; Arqu -Castells et al., 2016; Gibbs et al., 2017; Hvide and Jones, 2018; Charness and Grieco, 2019; Kachelmeier et al., 2024; Laske et al., 2024).¹ But are output incentives the best way to incentivize idea generation? There is some evidence suggesting that this may not be the case. Contrary to insights from routine tasks (see e.g., Prendergast, 1999, for an overview), recent evidence suggests that less-aligned incentives (input incentives) may be more effective in fostering idea generation (Manso, 2011; Ederer and Manso, 2013; Amabile and Pratt, 2016; Laske et al., 2024; Blaufus et al., 2025). We seek to provide causal evidence on the effectiveness of input and output incentives and on differences between these two types of incentives in the context of idea generation. In a between-subjects experimental design, we contrast the number of innovative ideas generated under fixed wage, input and output incentives. To provide insights into how incentives affect idea generation, we analyze the effect of incentives on effort provision and the effect of effort provision on resulting innovative ideas. As incentives are shown to also have an impact on the types of ideas generated (Azoulay et al., 2011; Manso, 2011), we additionally explore the effect of input and output incentives on the types of ideas generated.

The experimental context that we implement is a modified version of the Word Illustration Task (WIT) introduced by Laske et al. (2024). In the task, participants, our so-called ideators, use a predefined set of materials to illustrate words. All ideas are rated with respect to their quality (whether or not third-party raters can identify the illustrated word) and their originality (as compared to a previous set of ideas). Only ideas that are both original and of high quality are considered as innovative ideas. Ideators are instructed to create as many innovative ideas as possible. This experimental design has two advantages: First, compared to field data it allows us to observe the first stage of an innovation, i.e. the idea generation stage, which is usually unobservable for researchers in the field. Second, it is possible to quantify the value of ideas with respect to originality and quality.²

¹ Such positive effects arise only when the relevant output is clearly defined and communicated. Output incentives lacking a clear communication and/or definition of the relevant output tended to be ineffective (Erat and Gneezy, 2016; Bradler et al., 2019).

² These two criteria are frequently used when assessing innovative ideas. For example, the U.S. Patent and

In a between-subjects design, we vary the incentives provided to ideators. In the fixed treatment, payment is independent of performance but the level of the fixed payment differs. In the input treatment, payment is contingent on the amount of time spent on the task. In the output treatment, ideators receive a bonus for every innovative idea that they come up with. In all of these treatments, the level of incentives is randomized on the ideator level. To measure the effect of our treatments on effort, we distinguish between effort duration (time spent on the task) and effort intensity (number of ideas per unit of time). The number of innovative ideas generated determines our desirable performance outcome.

In this paper, we explore differences in the types of ideas that are created. Recent evidence suggests that incentives may have an impact on the types of ideas that are generated (Azoulay et al., 2011; Brügger et al., 2018). This may be because of an impact of incentives on individuals' willingness to explore new solutions (Manso, 2011; Ederer and Manso, 2013). Another channel through which incentives may have an impact on the types of ideas generated may be due to differences in the complexity of ideas. Possibly, incentives lead to a focus on less complex ideas that can be realized in less time. For example, when developing research questions researchers can focus on relatively narrowly defined research questions that can be answered using a small subset of available methods and time. On the other hand, researchers can also focus on research questions that are more complex, e.g., because they require a combination of several research methods whose application requires more time. Both cases may result in an important research contribution but the complexity of the underlying research will differ. Incentives may have an impact on the approach taken. To explore such an effect of incentives on the types of ideas generated, we additionally assess and compare the complexity of ideas generated. Ideators in our experiment receive a set of usable materials but are free to choose which and how many of these materials to use. We measure innovation complexity as the number of materials used to illustrate an innovative idea.

For rational agents, we predict that input incentives lead to an increase in effort duration but not in effort intensity as compared to fixed pay. In contrast, we predict that output incentives lead to an increase in both effort intensity and effort duration as both increase the number of innovative ideas and thus, the expected payment. Thus, one central difference between input and output incentives lies in the effect of incentives on effort. Given that input and output incentives have a different impact on how ideators work, we may also assume to see a difference in the types of ideas they generate. Specifically, effort duration and effort intensity may be related to the complexity of ideas generated. Based on previous findings (McCardle et al., 2018; Ganglmair et al., 2019; Laske

Trademark Office (USPTO) requires inventions to be novel, non-obvious, and useful in order to be patentable.

et al., 2024), we conjecture that output incentives may lead to reductions in the complexity of ideas, while input incentives may lead to an increase in the complexity of ideas.

We find evidence for a positive effect of incentives on effort provision. Input and output incentives lead to a substantial increase in effort duration by 40 percent and 25 percent, respectively. As predicted, we find no significant effect of input incentives on effort intensity, i.e., on the number of ideas generated per hour. In contrast, output incentives lead to an increase in effort intensity by 35 percent. Overall, we find a strong effect of incentives on the number of innovative ideas generated. We find that the number of innovative ideas is doubled through the introduction of input incentives, while it increases by 75 percent in the output incentives treatment. Thus, we find evidence for a substantial positive effect of incentives on idea generation.

Comparing input and output incentives, we find that the difference in the number of innovative ideas generated is not statistically significant. However, we find that adjustments in effort provision underlying these improvements vary. Under output incentives, improvements are achieved due to longer working time and more efficient use of time, i.e., due to increases in effort duration and effort intensity. Under input incentives, this increase is achieved merely due to an increase in effort duration. We can show that this has an impact on the timing of the generation of innovative ideas and on the types of ideas generated. Specifically, we find that the time until innovative ideas are generated is larger for input incentives as compared to output incentives. Analyzing differences in the types of ideas, we find that innovative ideas generated under output incentives are significantly less complex as compared to innovative ideas generated under fixed pay or input incentives. Differences arise due to differences in the way ideas are illustrated, i.e., using more materials to illustrate the same word (average word complexity), and due to differences in the underlying ideas, i.e., illustrating words that on average require more complexity in the illustration (relative idea complexity).

Overall, we demonstrate that both input and output incentives are suitable to foster idea generation. However, the optimal contract choice depends on the specific context. If organizations seek to find fast ideas on how to solve pressing problems, output incentives may be more suitable. Due to an increase in effort intensity, ideators under output incentives require less time to generate innovative ideas. If organizations seek to encourage the development of complex ideas, input incentives may be more favorable.

Our insights are highly relevant to organizations seeking to incentivize idea generation. Furthermore, our research contributes to the literature on the effect of incentives on idea generation (Arqué-Castells et al., 2016; Gibbs et al., 2017; Charness and Grieco, 2019; Laske et al., 2024;

Kachelmeier et al., 2024). Finally, we contribute to the literature on interactions between policy interventions and the type of innovations (Azoulay et al., 2011; Manso, 2011).

2 Experimental setup

Ideators are asked to perform an ideation task, the Word Illustration Task (WIT) based on the design by Laske et al. (2024). In this design, ideators receive a set of materials and are asked to use these materials to illustrate words. The task consists of coming up with a word that an ideator wants to illustrate and illustrating the chosen word with the given materials (Laske et al., 2024). Figure 1 provides a screenshot of the experiment illustrating the available materials to work with and the area in which illustrations were provided (see panel 1a of the figure). Using their mouse, ideators could drag and rotate materials, change the size of materials and the sequence in which they were layered. When done illustrating a word, ideators would enter the word that was illustrated (see panel 1b in figure 1) and could then decide whether they wanted to illustrate another word or whether they want to terminate the task.³

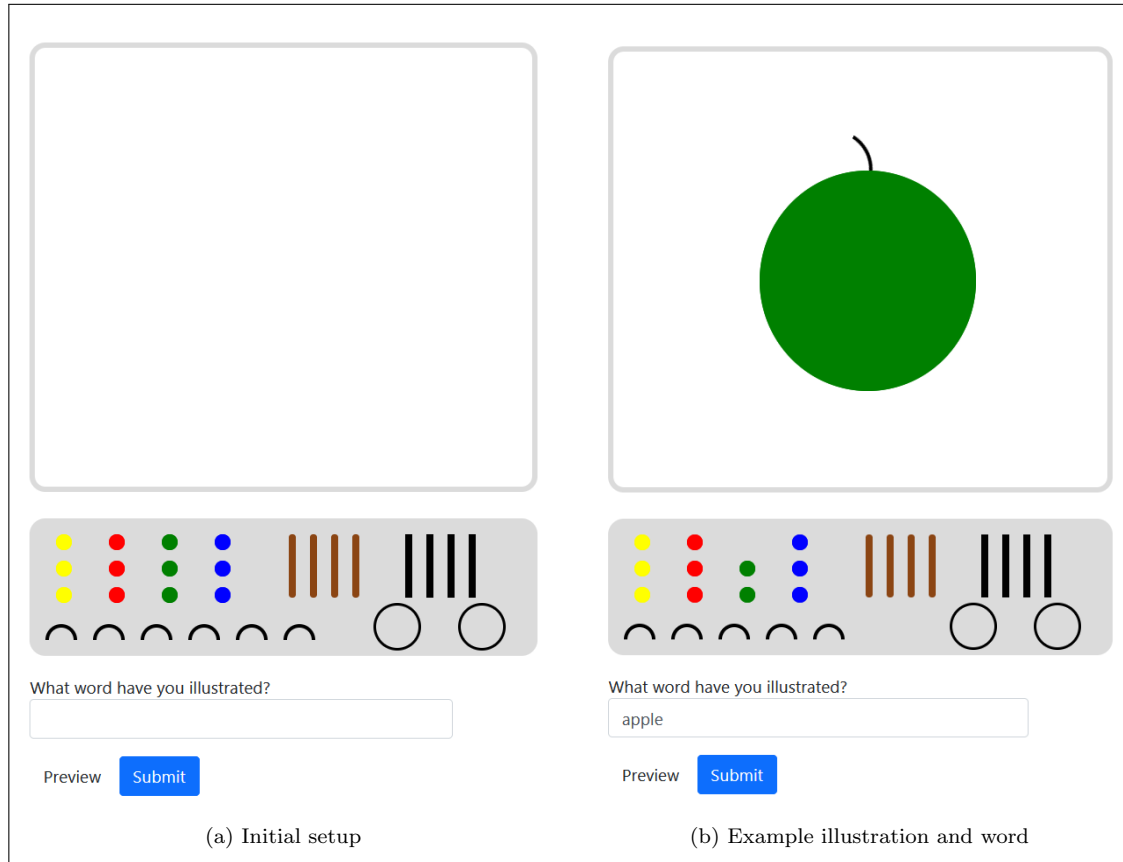


Figure 1: The Word Illustration Task (WIT)

³ Ideators were not allowed to submit the same word twice. Also, ideators could only submit illustrations of terms consisting of a single word that is included in a standard American English dictionary.

NOTE: Screenshots from the experimental task showing the setup in the beginning of the task in panel 1a and an example illustration together with the typed-in word in panel 1b. See section ?? in the supplementary analyses for more details.

Working time in our experiment is endogenous, i.e. ideators are free to decide how long they want to work on the task. Our first measure for effort provision is effort duration, i.e., the time ideators chose to spend performing this task measured in hours. Our second measure for effort provision is the effort intensity, i.e., the number of ideas generated per hour.⁴

All ideators in the experiment are instructed to illustrate as many innovative (high quality and original) ideas as possible. Thus, the desirable outcome in our design is the number of innovative ideas generated. We assessed the originality of ideas through a comparison to 300 ideas generated by different ideators from a previous experiment using a similar procedure. An idea was considered as original if the illustrated word is not among these 300 words and not original otherwise. We assessed quality through a survey among hypothetical customers. On average, each idea was provided to ten customers. These customers received a monetary payoff for identifying the exact word that was illustrated. An idea is classified as high quality if at least half of the customers identify the illustrated word, and as low quality otherwise.

Figure 2 displays examples of ideas which are classified according to quality and originality. The idea depicted in panel (a) is an illustration of the word music. This idea is original but less than half of the customers was able to identify the illustrated word. Thus, it is of low quality. The idea depicted in panel (c) is an illustration of the word cat. In WIT, multiple ideators choose to illustrate cats. Thus, the idea is not original. For this specific illustration of the word cat, only very few customers were able to identify it as a cat. Thus, the idea is of low quality. In panel (d), we provide an illustration of the word house. While the house is not original, the majority of raters identified it as house. Thus, this idea is not original but of high quality. To possibly result in innovation, an idea has to be original and of high quality. The idea in the panel (b) is an example of such an innovative idea generated in our sample. It is an illustration of the word basketball. None of the 300 reference ideas illustrated the word basketball and most of the customers identified it as such.

⁴ To fulfill minimum wage requirements, we had to cap working time at 90 minutes. Ideators were informed about this cap and only roughly 1% of the sample reached this threshold. There are three ideators in the input, and two in the output treatment who reached the 90 minutes threshold. Results are consistent when excluding those ideators who worked 90 minutes from the study. See table ?? in the supplementary analyses. We implemented an inactivity check during the WIT. Excluding those ideators whose task has been automatically ended due to inactivity keeps the results consistent. See table ?? in the supplementary analyses.

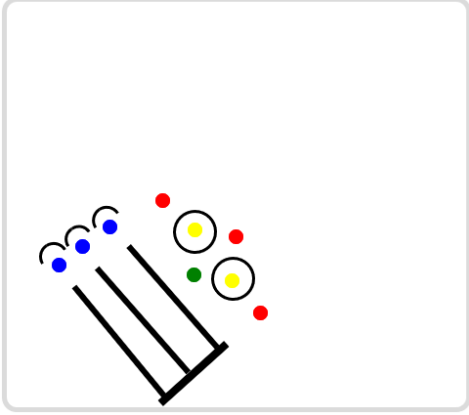
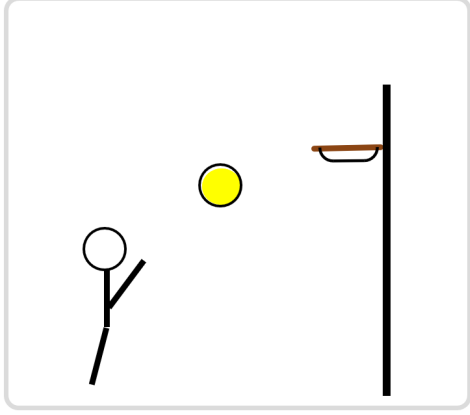
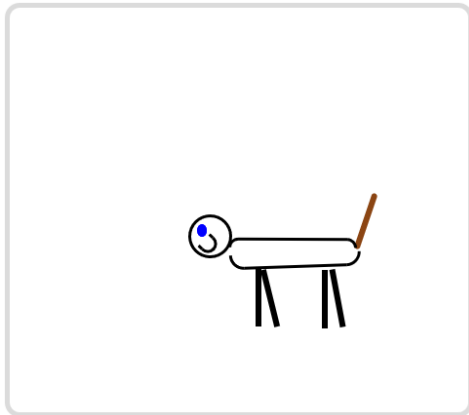
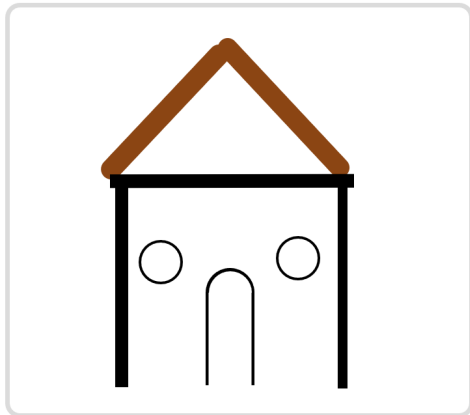
		Quality of ideas	
		Low quality	High quality
Originality of ideas	Original	 <p>(a) MUSIC <i>Not innovative</i></p>	 <p>(b) BASKETBALL Innovative</p>
	Not original	 <p>(c) CAT <i>Not innovative</i></p>	 <p>(d) HOUSE <i>Not innovative</i></p>

Figure 2: Example ideas

NOTE: Example ideas from our experiment. The words illustrated are: (a) music, (b) basketball, (c) cat, and (d) house. (a) and (b) show original words while (b) and (d) are high quality illustrations. Thus, only idea (b) is classified as innovative.

Finally, to derive insights into differences in the types of ideas that are generated, we analyze the complexity of ideas, i.e. the amount of used materials per idea. Ideators receive the given set of materials as presented in panel (a) of figure 1 but are free to choose which and how many of these materials to use for their ideas. Differences in the complexity can arise through two different channels: On the one hand, ideators can choose to illustrate different types of ideas, i.e. words that generally require a different amount of materials. On the other hand, ideators may use more or less materials in order to illustrate the same word. To distinguish between these effects, we consider illustrations of words that are recurring in our sample, i.e. words that multiple ideators illustrated in our experiment. For these recurring ideas, we define average word complexity as the average number of materials used to illustrate the same word. Furthermore, we define the relative idea

complexity as the relative difference in the number of materials used in the relevant idea compared to the average word complexity. Figure 3 provides some examples of ideas indicating the average word complexity and the relative idea complexity. By distinguishing between these two measures, we can identify changes in the complexity which arise due to the choice of the underlying word. For example, illustrating an egg on average involves the use of 2.5 materials (average word complexity = 2.5), while illustrations of a helicopter on average involve the use of 10.3 materials (average word complexity = 10.3). Changes may also arise because the same word is illustrated using a different amount of materials. The helicopter illustrated in panel (b) of figure 3 uses 6 materials, while the helicopter in panel (c) uses 18 materials. We calculate the relative idea complexity as the percentage deviation between the materials used and the average word complexity. Specifically, relative idea complexity is derived as follows: $\text{relative idea complexity} = \frac{\text{materials used} - \text{average word complexity}}{\text{average word complexity}}$. For the two examples in figure 3, the relative idea complexity is equal to -0.42 for the helicopter in panel (b), as it uses 42 percent less materials than the average illustration of a helicopter. The helicopter provided in panel (c) of figure 3 uses 18 materials and thus 75 percent more as compared to the average word complexity.

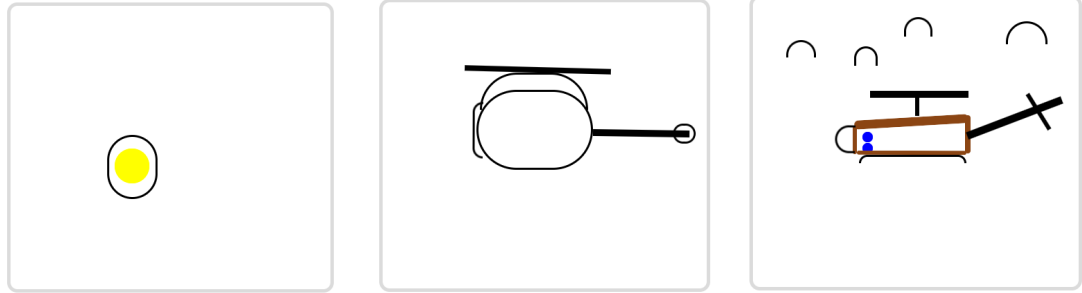
		
(a) EGG (b) HELICOPTER (c) HELICOPTER		
Materials used		
2	6	18
Average word complexity		
2.5	10.3	10.3
Relative idea complexity		
-0.20	-0.42	+0.75

Figure 3: Idea complexity and relative idea complexity

NOTE: Example illustrations with their respective number of materials used, their average word complexity, and their relative idea complexity. Average word complexity is defined as the full-sample average number of materials for a given word. Average word complexity is defined by word. Relative idea complexity is defined as the percental deviation from the full-sample average number of materials for a given word (expressed in decimal numbers). Relative idea complexity is defined by illustration.

Between treatments, we vary the incentives provided. All ideators in our experiment receive a show-up fee of £1.25. In the three main treatments of this experiment, we pay an additional bonus to ideators. In the fixed treatment, ideators receive a bonus independent of performance. In the

input treatment, the bonus is contingent on the time (minutes) spent working on the task. In the output treatment, the bonus is contingent on the number of innovative ideas generated.

In addition to the design of incentives, the level of incentives may have an impact on ideators' behavior. To calibrate incentives, we randomized the nominal bonus size that is communicated to ideators within the same interval of £[1.00; 5.00]. Keeping the interval the same allows us to test for an effect of the communicated nominal bonus size. Additionally, we test for robustness of effects when holding the payment for a given performance constant. To be able to provide this comparison, we conduct an additional treatment, the baseline treatment. In the baseline treatment ideators receive no bonus at all. Performance in this treatment serves as a benchmark to determine the incentives level for a given performance level. Table 1 summarizes the treatments and sample size.

Table 1: Incentive schemes

Payments	Treatment groups			
	Fixed	Input	Output	Baseline
Show-up fee	£1.25	£1.25	£1.25	£1.25
Fixed bonus	£[1.00; 5.00]			
Bonus paid per hour		£[1.00; 5.00]		
Bonus paid per innovative idea			£[1.00; 5.00]	
Average nominal bonus	£2.89	£3.15	£2.96	None
Number of observations	106	101	118	104

NOTE: The show-up fee in the fixed treatment and the additional payments in the input and output treatments are randomly drawn from the denoted intervals between ideators. For the input treatment, the payoff is calculated on a per minute basis, i.e. the additional bonus per minute comes from the interval £[0.02; 0.08]. The denoted payment shows the payoff for working exactly 60 minutes. In the experiment, ideators are provided with both values (see figure ?? in the supplementary analyses).

We programmed the experiment using oTree (Chen et al., 2016). We used a modified version of the code provided by Laske et al. (2024). The main experiment was conducted on Prolific using legal-aged ideators from the United States whose first language is English. The final data collection took place in August 2022. The third-party rating was conducted shortly after the data collection for the main experiment.

After the experiment, we screened all submitted ideas to ensure that they satisfy our rules. In particular, there are 83 out of 1,818 ideas in which ideators did not try to illustrate a word but used the materials to write the word. As pre-registered, these were excluded from the analyses and were not payoff-relevant.⁵ One idea was excluded due to offensive content. Our final sample consists of 1,734 valid ideas.

⁵ The instructions clearly stated that ideators should not use the materials to write (see figure ?? in the appendix). See figure A1 in the appendix for examples of excluded ideas.

To rate the quality of ideas, we conduct an additional survey with customers who were Amazon Mechanical Turk workers. Each customer is provided with fifty consecutive ideas which are randomly drawn from the entire sample of valid ideas. Customers receive a show-up fee of \$0.50 together with a performance-contingent bonus of \$0.10 for every correctly identified word. On average, customers earned \$1.76 and worked on the task for 27 minutes. An idea is correctly identified if a customer types in exactly the same characters as the ideator. On average, an idea was rated by ten customers, where the minimum number of customers is seven and the maximum number is twelve. As noted above, we rate ideas as high quality if at least half of the customers correctly identify the illustrated word.

After completion of the task, ideators proceed to a follow-up questionnaire. For completing the questionnaire, they receive an additional payment of £1.35. In this survey, we include a measure for the general creative potential (RIBS-S) based on 19 questions on a 5-point Likert scale (Runco et al., 2014). We asked ideators for their general willingness to take risks on an 11-point Likert scale ranging from "not at all" to "very much" (Caliendo et al., 2014). Finally, the questionnaire contains questions on the weekly number of hours worked, and the monthly net income.⁶ Age and gender of ideators is provided by Prolific.⁷

3 Results

Table 2 provides summary statistics on effort duration, effort intensity and the number of innovative ideas per treatment. On average, ideators spent half an hour in the task and generated 10 ideas per hour. We find that innovative ideas are rare, so that the average number of innovative ideas per ideator is only 0.30. However, we observe substantial treatment differences in all three measures.⁸

⁶ See figures ?? to ?? in the supplementary analyses for the exact wording of questions and answering scales.

⁷ Different to our own questionnaire, where answering all questions was mandatory, ideators can revoke their consent and Prolific deletes information after certain time intervals assuming they expired. This is why we do not have these data for all ideators.

⁸ See also figure A2 in the appendix for histograms of these variables.

Table 2: Descriptive statistics of our dependent variables

	Obs	Mean	SD	Q25	Q50	Q75	Min	Max
Effort duration (working time in hours):								
Total	325	0.48	0.31	0.26	0.40	0.60	0.09	1.50
Fixed	106	0.40	0.25	0.22	0.33	0.47	0.10	1.26
Input	101	0.56	0.35	0.31	0.50	0.70	0.09	1.50
Output	118	0.50	0.31	0.28	0.42	0.64	0.09	1.50
Effort intensity (number of ideas per hour):								
Total	304	9.71	7.75	4.60	7.18	12.15	0.74	43.81
Fixed	97	8.75	6.50	4.6	6.59	10.86	0.96	40.00
Input	97	8.33	7.29	3.5	5.84	10.14	0.74	34.37
Output	110	11.78	8.72	5.69	9.09	16.23	1.87	43.81
Innovative ideas (number of ideas that are original and of high quality):								
Total	325	0.32	0.79	0	0	0	0	6
Fixed	106	0.20	0.47	0	0	0	0	2
Input	101	0.42	1.06	0	0	0	0	6
Output	118	0.35	0.73	0	0	0	0	4

NOTE: The table reports the number of observations (Obs), the mean, standard deviation (SD), 25th percentile (Q25), median (Q50), 75th percentile (Q75), as well as the minimum (Min) and the maximum value (Max). Parameters are reported both for the full sample as well as separate by treatment group.

Figure 4 illustrates the mean value of effort duration per treatment group. Pairwise Mann-Whitney U tests (MWU) between the three treatment groups show that under both input and output incentives effort duration is significantly larger as compared to fixed incentives (MWU fixed vs. input: $p < 0.001$, fixed vs. output: $p = 0.007$). Compared to the fixed treatment, we find a 40 percent increase in effort duration in the input treatment and a 25 percent increase in effort duration in the output treatment. We observe a slight but insignificant tendency that effort duration is larger in the input as compared to the output treatment (MWU input vs. output: $p = 0.172$).

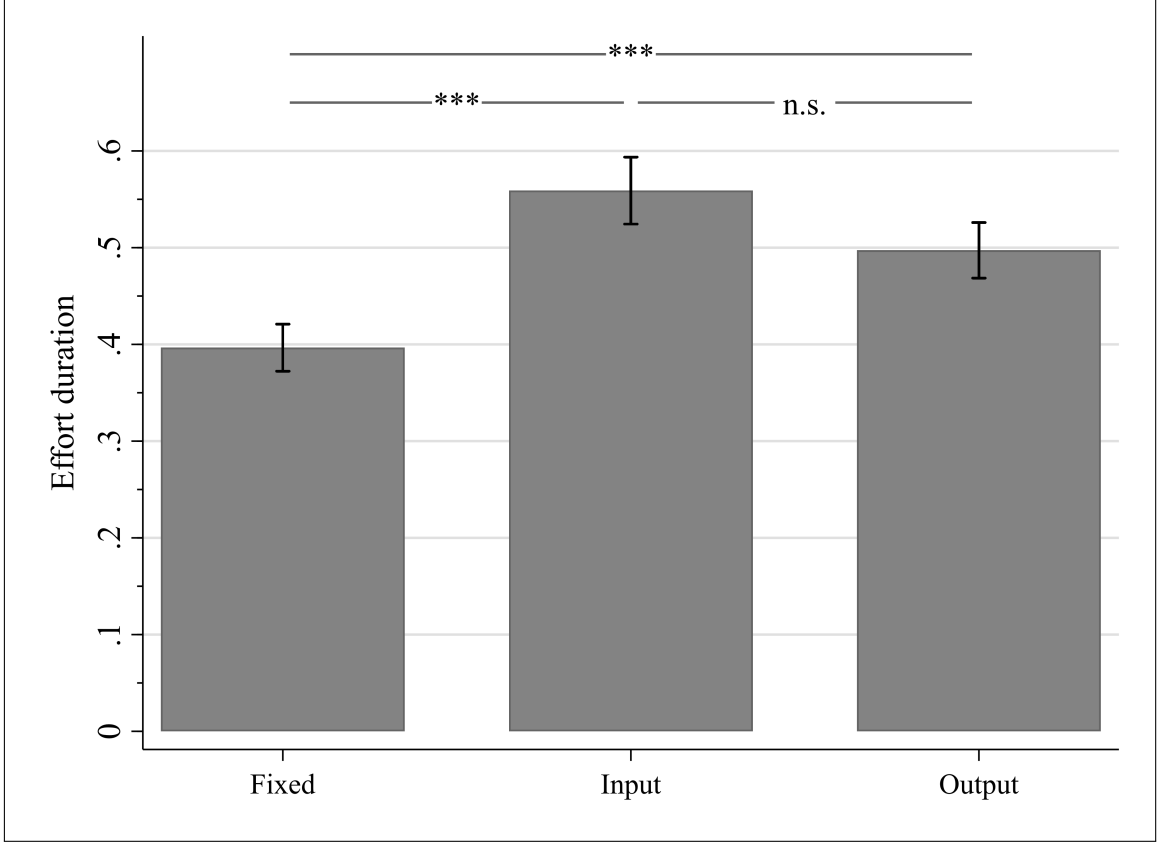


Figure 4: Effort duration

NOTE: Mean working time in hours per ideator per treatment group, $n=325$. Standard errors are shown as vertical lines. Pairwise MWU test results are reported as horizontal lines above the bars, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

We replicate the non-parametric result in regression analyses as shown in table 3. We can show that results are robust to including a set of control variables such as controls for creative potential and demographic characteristics (see columns (1) and (2) of table 3 for details).⁹ A noteworthy difference between the non-parametric and the regression results is that the difference between the input and the output treatments, which was not significant at conventional levels in the non-parametric analysis, is marginally significant (Wald-test: $p = 0.056$) in the regression analysis.

Finding 1: Both input and output incentives lead to an increase in effort duration.

Figure 5 depicts the mean values of effort intensity per treatment group.¹⁰ We find no significant difference between the fixed wage and the input incentives with respect to effort intensity (MWU fixed vs. input: $p = 0.217$). If anything, we observe a slight tendency that effort intensity is lower in the input as compared to the fixed treatment. In contrast, we observe significantly higher effort intensity in the output as compared to the fixed and to the input treatment (MWU fixed vs.

⁹ See subsection A.1 in the appendix for details on control variables.

¹⁰ Note that there are less observations than in figure 4 as effort intensity is only defined for those ideators who create at least one valid idea.

output: $p = 0.006$, input vs. output: $p < 0.001$). Compared to the fixed treatment, we observe an increase in effort intensity by 40 percent. Compared to the input treatment, we observe a 34 percent increase in effort intensity.

Table 3 confirms these findings with OLS regressions in columns (3) and (4). The input treatment dummy is not significantly different from zero. By contrast, the output treatment dummy is significant and positive. Results are robust to including controls.

Finding 2: Output incentives lead to an increase in effort intensity, while input incentives do not have a significant effect on effort intensity.

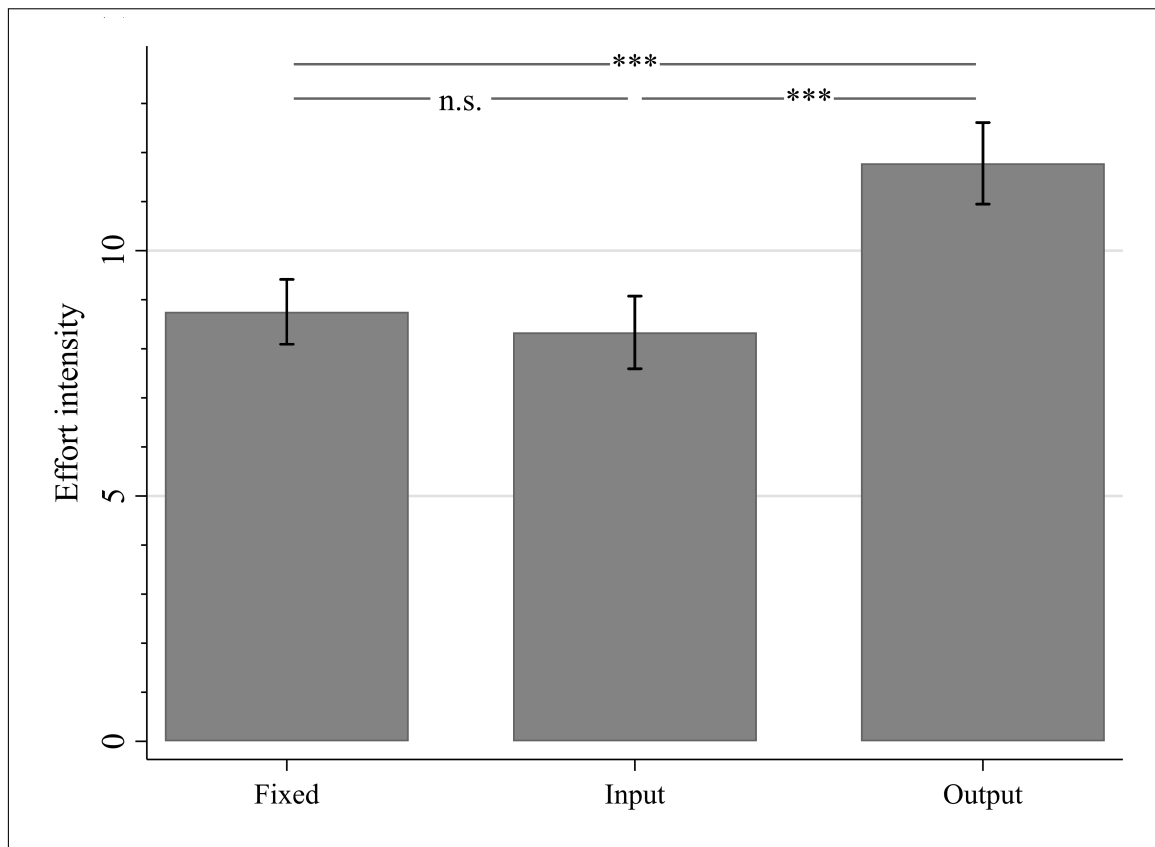


Figure 5: Effort intensity

NOTE: Mean number of ideas divided by the working time in hours per ideator per treatment group, $n=304$. Standard errors are shown as vertical lines. There are less observations than in figure 4 as effort intensity is only defined for those ideators who created at least one valid idea. Pairwise MWU test results are reported as horizontal lines above the bars, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

Table 3: OLS regressions: Effort duration and effort intensity

	(1) Effort duration	(2)	(3) Effort intensity	(4)
Input	0.1624*** (0.0423)	0.1754*** (0.0412)	-0.4201 (0.9917)	-0.4187 (0.9991)
Output	0.1006*** (0.0376)	0.0950*** (0.0358)	3.0259*** (1.0620)	2.9053*** (1.0808)
Constant	0.3966*** (0.0243)	0.4756*** (0.0561)	8.7531*** (0.6602)	8.8220*** (1.3509)
Wald tests on coefficient differences:				
Difference input – output	0.0618	0.0803*	-3.4461***	-3.3240***
Controls		✓		✓
Observations	325	325	304	304
Adjusted R^2	0.0385	0.1306	0.0345	0.0493

NOTE: OLS regression results with effort duration (working time in hours) per ideator as dependent variable in models (1) and (2), and effort intensity (number of ideas divided by the effort duration) as the dependent variable in models (3) and (4). We include treatment dummies with the fixed treatment as reference group. In the models (2) and (4), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Does a change in effort provision result in an increase in the number of innovative ideas generated? Table 4 provides results from a negative binomial regression with the number of innovative ideas as the dependent variable. Both treatment coefficients are positive and significant in the model version of column (1), i.e., without further controls. Thus, we find that the number of innovative ideas increases in both the input and output treatment as compared to fixed. In the models provided in columns (3) to (5) in table 4, we control for effort duration and effort intensity. To facilitate coefficient comparisons, we standardize both effort duration and effort intensity. This provides insights on the relevance of treatment effects on effort duration and effort intensity in explaining the effect of incentives on the number of innovative ideas.

When adding a control for standardized effort duration (column (3)), we find that both treatment coefficients are lower as compared to a setting without this control (χ^2 tests model (2) vs. model (3) input: $p < 0.001$, output: $p = 0.109$). and no longer statistically significant. By contrast, controlling for effort intensity (column (4)) decreases the effect of output incentives (χ^2 tests model (2) vs. model (4) output dummies: $p = 0.007$) but not the effect of input incentives (χ^2 tests model (2) vs. model (4) input dummies: $p = 0.302$). Controlling for both standardized effort duration and standardized effort intensity (column (5)), we observe a significant decrease in the treatment coefficients (χ^2 tests model (2) vs. model (5) input dummies: $p = 0.002$, output

dummies: $p = 0.002$) and no longer observe significant treatment effects.

Table 4: Negative Binomial regressions: Innovative ideas

	(1)	(2)	(3)	(4)	(5)
Input	0.7415** (0.3405)	0.7202** (0.3351)	0.2211 (0.3248)	0.5905* (0.3239)	0.0343 (0.3200)
Output	0.5618* (0.2990)	0.5713* (0.2965)	0.3799 (0.2941)	0.1991 (0.3180)	-0.0589 (0.3055)
Standardized effort duration			0.5271*** (0.0846)		0.6197*** (0.0777)
Standardized effort intensity				0.4933*** (0.0863)	0.5748*** (0.0818)
Constant	-1.6189*** (0.2279)	-1.0532*** (0.4037)	-1.2526*** (0.3701)	-1.1349*** (0.4335)	-1.4428*** (0.3831)
Wald tests on coefficient differences:					
Input – output	0.1797	0.1488	-0.1588	0.3914	0.0932
Controls		✓	✓	✓	✓
Observations	325	325	325	304	304
Pseudo R^2	0.0100	0.0350	0.0907	0.0882	0.1814

NOTE: Negative Binomial regression results with number of innovative ideas per ideator as dependent variable. We include treatment dummies with the fixed treatment as reference group. In the models (3) and (5), we control for standardized effort duration, and in models (4) and (5), we control for standardized effort intensity. In the models (2) to (5), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. The standardized effort intensity measure is not available for the 21 participants who submitted no illustration thereby reducing the number of observations in columns (4) and (5). Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Overall, we find that both input and output incentives lead to an increase in the number of innovative ideas that is of similar magnitude. However, the positive impact of input incentives on the number of innovative ideas arises merely due to an increase in effort duration. The positive effect of output incentives arises due to both an increase in effort duration and an increase in effort intensity. To visualize the relevance of effort duration and effort intensity in explaining the increase in the number of innovative ideas, we construct a generalized structural equation model, where the input and output treatment dummies are used as exogenous variables, the standardized effort duration and standardized effort intensity are our mediators and the number of innovative ideas is the endogenous variable. The estimation results are shown in table A4 in the appendix and are consistent with the results discussed above. Figure 6 plots the estimated indirect effects by multiplying the first-order derivatives of the respective effects' paths as proposed by Geldhof et al. (2018). The direct effects are not depicted as they are not significantly different from zero, i.e. the treatment effects are fully mediated through effort duration and effort intensity.

Finding 3: Input and output incentives lead to an increase in the number of innovative ideas. This increase is of similar magnitude but stems from different adaptations in effort provision. The increase in the output treatment stems from both an increase in effort duration and an increase in effort intensity. In contrast, the increase in innovative ideas observed in the input treatment is solely due to an increase in effort duration.

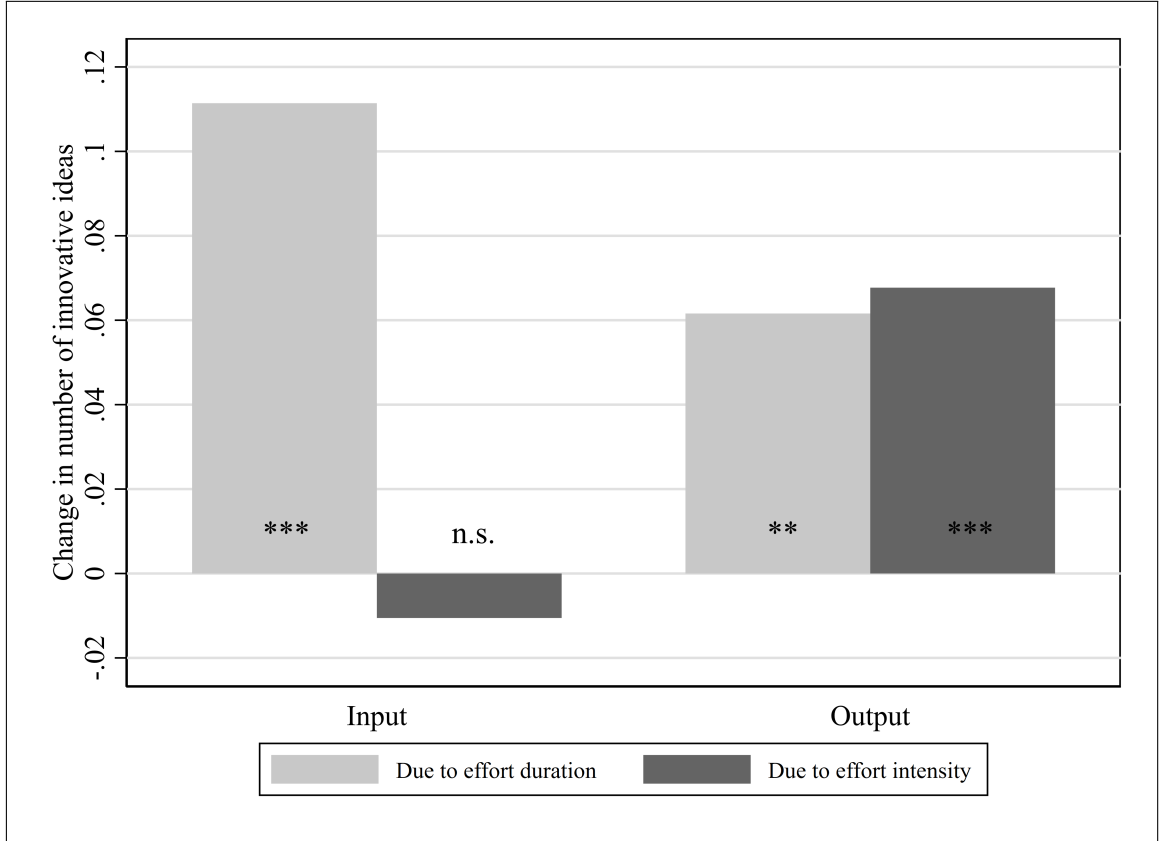


Figure 6: Decomposition of the treatment effects

NOTE: Estimated coefficients via a generalized structural equation model (GSEM) as shown in columns (1) to (3) of table A4 in the appendix. Input and output are defined as exogenous variables, standardized effort duration and effort intensity are the mediators, and the number of innovative ideas is the endogenous variable. The effect "due to effort duration" is the indirect effort duration effect, and "due to effort intensity" is the indirect effort intensity effect, respectively. Asterisks report significance levels of the respective coefficients calculated via bootstrapping with 500 repetitions, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

To further understand the positive effect of incentives, table 5, provides regression results for the effect of incentives on the number of ideas (columns (1) and (2)), the share of original ideas (columns (3) and (4)), and the share of high quality ideas. We find that both input and output incentives have a positive impact on the number of ideas generated. In the context with endogenous choice of working time, we additionally find that output incentives have a slight positive effect on the share of original ideas. We find no significant effect of input or output incentives on the fraction of high quality ideas. Thus, we do find some evidence that incentives shape the types of ideas generated.

Table 5: OLS regressions: Number of ideas and share of ideas by dimension

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of ideas		Share of original ideas		Share of high quality ideas	
Input	1.1381** (0.5456)	1.2460** (0.5682)	-0.0174 (0.0505)	-0.0183 (0.0497)	0.0281 (0.0414)	0.0296 (0.0410)
Output	2.1620*** (0.5770)	2.1136*** (0.5874)	0.0789* (0.0469)	0.0590 (0.0475)	-0.0227 (0.0347)	-0.0124 (0.0340)
Constant	2.9906*** (0.3090)	3.6326*** (0.7587)	0.6102*** (0.0377)	0.6497*** (0.0641)	0.1988*** (0.0276)	0.2738*** (0.0519)
Wald tests on coefficient differences:						
Input – output	-1.0238	-.8676	-0.0964**	-0.0773*	.0508	.0419
Controls		✓		✓		✓
Observations	325	325	304	304	304	304
Adjusted R^2	0.0333	0.0640	0.0099	0.0250	-0.0004	0.0105

NOTE: OLS regression results with total number of ideas per ideator as dependent variable in models (1) and (2), share of original ideas out of all ideas per ideator as dependent variable in models (3) and (4), and share of high quality ideas out of all ideas per ideator as dependent variable in models (5) and (6). We include treatment dummies with the fixed treatment as reference group. In the models (2), (4), and (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Do we also find differences in relevant characteristics of innovative ideas? The right panel of figure 7 illustrates the innovation complexity, i.e., the mean number of materials used for the illustration of innovative ideas. We find that innovative ideas provided in the output treatment are significantly less complex as compared to those provided in the fixed or the input treatments (MWU fixed vs. output: $p = 0.005$, input vs. output: $p = 0.006$). Note that we find no difference between fixed and input (MWU: $p = 0.831$). Analyzing recurring ideas, we can show that this difference is due to both lower average word complexity and due to lower relative idea complexity (see tables A6 and A7 in the appendix for details). Both of these effects are mainly explained by the difference in effort intensity, i.e. a high effort intensity goes hand in hand with a shift towards less complex ideas.

We show in the appendix that the average word complexity does not systematically change neither its originality nor its average quality. That is, speaking in terms of figure 3, the decision between illustrating an egg or illustrating a helicopter does not systematically change whether the idea is original or of high quality. However, the relative idea complexity is associated with an increase in the quality of an idea. That is, using more materials for a given word makes it more likely that the respective idea becomes a high quality idea. This effect, however, is rather weak so that there is no effect detectable on the innovativeness likelihood (see the appendix for more details). This

shows that an increase in the relative idea complexity is not just a wastage of materials but that it can positively impact the quality of an idea.

We can show that the pattern regarding changes in the complexity further holds for the types of materials that are used (e.g., red circles, brown sticks, and so on). Taking the average number of used material types per ideator shows again that ideators from the output treatment use significantly less than both other groups (MWU: fixed vs. output: $p = 0.004$, input vs. output: $p = 0.001$), but there is no difference between fixed and input (MWU fixed vs. input: $p = 0.721$). Aggregating this measure even further by looking at the shapes that are used (e.g., any circle, any stick, and so on) comes again to the same result: Ideators in the output treatment use less shapes than both other groups (MWU fixed vs. input: $p = 0.516$, fixed vs. output: $p = 0.031$, input vs. output: $p = 0.005$). Thus, the output treatment's tendency for simplification seems to be a fairly general pattern.

Finally, we consider possible delays in innovation. Analyzing the time to innovation in the left panel of figure 7, we find that it takes more time for innovative ideas to arise under input incentives as compared to both other treatment groups (MWU fixed vs. input: $p = 0.095$, fixed vs. output: $p = 0.840$, input vs. output: $p = 0.026$). Remember that ideators under input incentives increase their effort duration but keep their effort intensity constant. This is why the average submission time of (innovative) ideas increases. By contrast, ideators under output incentives also increase their effort intensity so that the submission time per innovative idea stays on a similar level as for ideators under the fixed treatment.

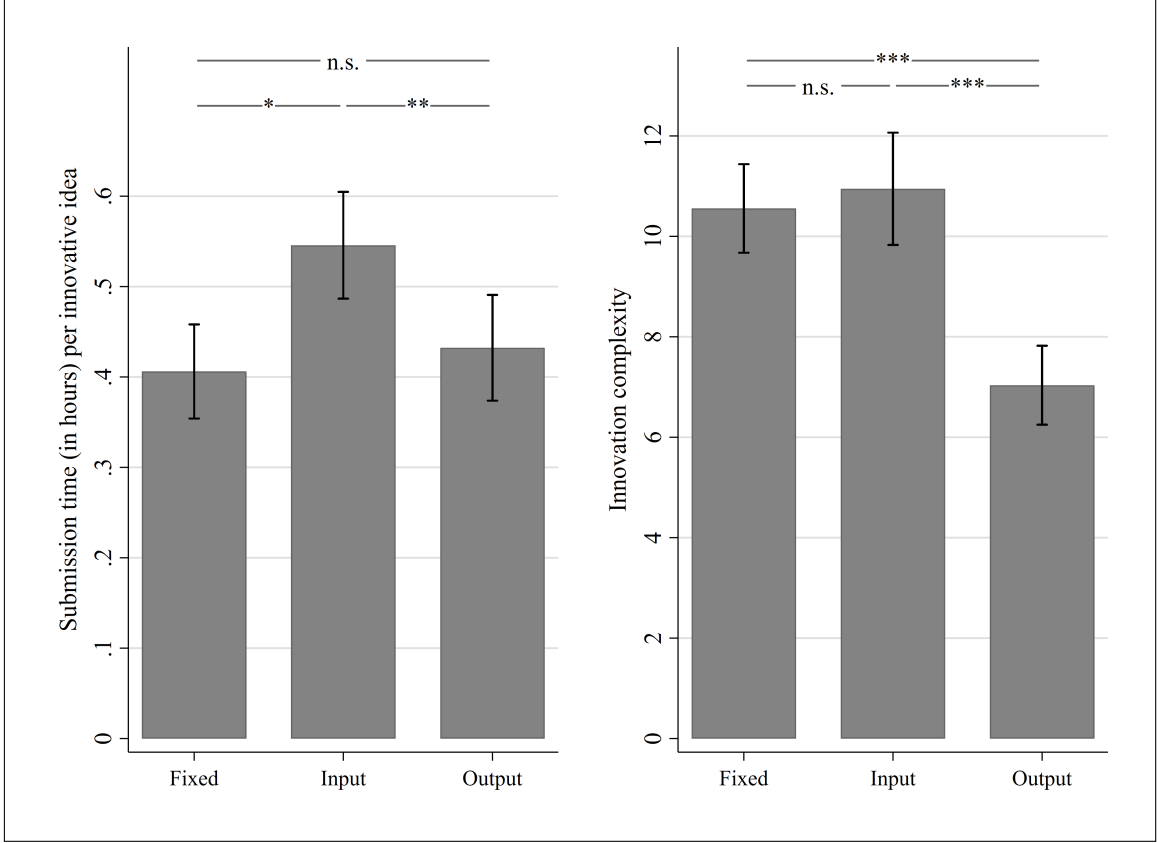


Figure 7: Submission time and types of innovative ideas

NOTE: Mean measure per ideator per treatment group. The left panel shows the mean submission time of innovative ideas per ideator per treatment group ($n=65$). The submission time is the working time that has passed until the submission of the respective innovative idea (in hours). The right panel shows the innovation complexity, i.e. the mean number of used materials for innovative ideas per ideator per treatment group ($n=65$). Standard errors are shown as vertical lines. Pairwise MWU test results are reported as horizontal lines above the bars, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

Finding 4: Input incentives lead to a delay in the generation of innovative ideas, while output incentives lead to a shift towards less complex innovative ideas.

We find that the main results of the paper are robust to controlling for the bonus size in various ways as shown in tables A10 to A12 in the appendix. First, we include the nominal bonus size, i.e. the drawn number that is communicated to ideators and drawn from the interval of $[1.00, 5.00]$. We can replicate our main findings from tables 3 and 4 when including this control. For instance, looking at the number of innovative ideas the treatment coefficients are 0.7240 (input) and 0.5436 (output) in the model version of column (5) of table A10, respectively. Comparing them to the coefficients of 0.7415 (input) and 0.5618 (output) from the model of column (1) of table 4 and considering the standard errors of about 0.3 for each of the four coefficients shows that the estimates change only incrementally.

Possibly, it is not the nominal bonus size that matters but the realized bonus controlling for performance matters. To control for this possibility, we use the baseline group as our performance

reference and construct a control for the real bonus that is payoff-equivalent given the average performance in this baseline group. Again, the results are qualitatively similar to the results absent of this control. In this case, the coefficients increase to 0.8717 (input) and 0.7221 (output) as shown in column (5) of table A11. While the coefficient for the input treatment remains significant, the coefficient for the output treatment does not meet conventional significance levels ($p = 0.129$). Thus, it seems that differences in the real bonus do not seem to be of major importance when explaining our treatment effects.

Confirming this interpretation, we do not find any differences between the fixed treatment and the baseline group. That is, there is no difference neither in effort duration nor in effort intensity, and thus, also not in the number of innovative ideas between these two groups. Hence, we conclude that it is mainly the type of incentives and not the size that creates the observed treatment effects. See the appendix for more details.

We provide additional robustness checks in the supplementary analyses. For example, we can show that results are qualitatively robust to adaptations of the sample included, and in the thresholds for quality and originality in determining innovative ideas. Furthermore, we show that results are robust to different statistical approaches.

4 Conclusion

This paper presents experimental evidence on the effect of incentives on idea generation. By varying the type of incentives, we show that both input and output incentives can effectively increase the number of innovative ideas generated compared to a fixed bonus. However, the mechanism leading to this increase in the number of innovative ideas varies. The increase in the number of innovative ideas generated in the input treatment stems merely from an increase in effort duration. The increase in the number of innovative ideas in the output treatment arises both due to an increase in effort duration and an increase in effort intensity. We further provide evidence for side-effects of incentives. Under output incentives, ideators provide less complex ideas as compared to input incentives or the fixed bonus. Input incentives tend to delay the generation of innovative ideas. Thus, the optimal incentive contract is context-specific.

Our results expand the existing literature on incentives in the context of innovation and help to provide clear recommendations how incentives should be designed. With effective incentive schemes in their hands, employers are enabled to increase the innovation potential of their workforce and thus, their economic prospects in terms of efficiency gains and comparative advantages (Geroski

et al., 1993; Blundell et al., 1999; Schulze and Hoegl, 2008; Anderson et al., 2014). In addition, our results are informative for policy makers in the discussion regarding the design of tax incentives to promote innovation on a broader scope. Broadly speaking, tax incentives can be categorized as either input- or output-based incentives. The former are for instance given by tax credits that are contingent on the R&D expenditures while for the latter there is the well-known example of patents in general and patent boxes as a more country specific intervention (Bloom et al., 2019). Importantly, our results provide causal evidence for the effect of such incentives on early stages of the innovation process, i.e., on idea generation. However, tax incentives are shown to also lead to global relocation ((Gaessler et al., 2021; Knoll et al., 2021; Schwab and Todtenhaupt, 2021). The experimental setup presented in this study provides valuable insights on the incentive effects, however it does not give consideration to the complex global inter-dependencies that are relevant in the context of innovation policy.

Data availability

Data will be made available on request.

References

- Amabile, T. M. and Pratt, M. G. (2016). The Dynamic Componential Model of Creativity and Innovation in Organizations: Making Progress, Making Meaning. *Research in Organizational Behavior*, 36:157–183. [link].
- Anderson, N., Potočník, K., and Zhou, J. (2014). Innovation and Creativity in Organizations: A State-of-the-Science Review, Prospective Commentary, and Guiding Framework. *Journal of Management*, 40(5):1297–1333. [link].
- Arqué-Castells, P., Cartaxo, R. M., García-Quevedo, J., and Godinho, M. M. (2016). Royalty Sharing, Effort and Invention in Universities: Evidence from Portugal and Spain. *Research Policy*, 45(2016):1858–1872. [link].
- Azoulay, P., Zivin, J. S. G., and Manso, G. (2011). Incentives and Creativity: Evidence from the Academic Life Sciences. *The RAND Journal of Economics*, 42(3):527–554. [link].
- Blaufus, K., Milde, M., Schaefer, M., and Schröder, M. (2025). Time to Innovate: The Role of Input Incentives in Creative Production. [link].

- Bloom, N., van Reenen, J., and Williams, H. (2019). A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives*, 33(3):163–184. [link].
- Blundell, R., Griffith, R., and van Reenen, J. (1999). Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms. *The Review of Economic Studies*, 66(3):529–554. [link].
- Bradler, C., Neckermann, S., and Warnke, A. J. (2019). Incentivizing Creativity: A Large-Scale Experiment with Performance Bonuses and Gifts. *Journal of Labor Economics*, 37(3):793–851. [link].
- Brüggen, A., Feichter, C., and Williamson, M. G. (2018). The Effect of Input and Output Targets for Routine Tasks on Creative Task Performance. *The Accounting Review*, 93(1):29–43. [link].
- Caliendo, M., Fossen, F., and Kritikos, A. S. (2014). Personality Characteristics and the Decisions to Become and Stay Self-Employed. *Small Business Economics*, 42(4):787–814. [link].
- Charness, G. and Grieco, D. (2019). Creativity and Incentives. *Journal of the European Economic Association*, 17(2):454–496. [link].
- Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97. [link].
- Ederer, F. and Manso, G. (2013). Is Pay for Performance Detrimental to Innovation? *Management Science*, 59(7):1496–1513. [link].
- Erat, S. and Gneezy, U. (2016). Incentives for Creativity. *Experimental Economics*, 19(2):269–280. [link].
- Gaessler, F., Hall, B. H., and Harhoff, D. (2021). Should There Be Lower Taxes on Patent Income? *Research Policy*, 50(1):104–129. [link].
- Ganglmair, B., Simcoe, T., and Tarantino, E. (2019). Learning When to Quit: An Empirical Model of Experimentation in Standards Development. *NBER Working Paper Series No. 24358*. [link].
- Geldhof, G. J., Anthony, K. P., Selig, J. P., and Mendez-Luck, C. A. (2018). Accommodating Binary and Count Variables in Mediation: A Case for Conditional Indirect Effects. *International Journal of Behavioral Development*, 42(2):300–308. [link].

- Geroski, P., Machin, S., and van Reenen, J. (1993). The Profitability of Innovating Firms. *The RAND Journal of Economics*, 24(2):198–211. [link].
- Gibbs, M., Neckermann, S., and Siemroth, C. (2017). A Field Experiment in Motivating Employee Ideas. *The Review of Economics and Statistics*, 99(4):577–590. [link].
- Hvide, H. K. and Jones, B. F. (2018). University Innovation and the Professor’s Privilege. *The American Economic Review*, 108(7):1860–1898. [link].
- Kachelmeier, S. J., Webb, R. A., and Williamson, M. G. (2024). Do Performance-Contingent Incentives Help or Hinder Divergent Thinking? *The Accounting Review*, 99(2):229–248. [link].
- Knoll, B., Riedel, N., Schwab, T., Todtenhaupt, M., and Voget, J. (2021). Cross-Border Effects of R&D Tax Incentives. *Research Policy*, 50(9):104326. [link].
- Lam, A. (2011). What Motivates Academic Scientists to Engage in Research Commercialization: ‘Gold’, ‘Ribbon’ or ‘Puzzle’? *Research Policy*, 40(2011):1354–1368. [link].
- Laske, K., Römer, N., and Schröder, M. (2024). Piece-Rate Incentives and Idea Generation — An Experimental Analysis. *CESifo Working Papers*, 11594. [link].
- Manso, G. (2011). Motivating Innovation. *The Journal of Finance*, 66(5):1823–1860. [link].
- McCardle, K. F., Tsetlin, I., and Winkler, R. L. (2018). When to Abandon a Research Project and Search for a New One. *Operations Research*, 66(3):799–813. [link].
- Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1):7–63. [link].
- Runco, M. A., Walczyk, J. J., Acar, S., Cowger, E. L., Simundson, M., and Tripp, S. (2014). The Incremental Validity of a Short Form of the Ideational Behavior Scale and Usefulness of Distractor, Contraindicative, and Lie Scales. *The Journal of Creative Behavior*, 48(3):185–197. [link].
- Schulze, A. and Hoegl, M. (2008). Organizational Knowledge Creation and the Generation of New Product Ideas: A Behavioral Approach. *Research Policy*, 37(10):1742–1750. [link].
- Schwab, T. and Todtenhaupt, M. (2021). Thinking Outside the Box: The Cross-Border Effect of Tax Cuts on R&D. *Journal of Public Economics*, 204:104536. [link].

A Appendix

A.1 Experimental setup

As discussed in section 2, ideators were supposed to illustrate words by using the provided materials. Some ideators, however, misunderstood the task and used the materials instead to provide writings of the respective word. Figure A1 shows three examples of such writing illustrations. These ideas were excluded from our analyses.

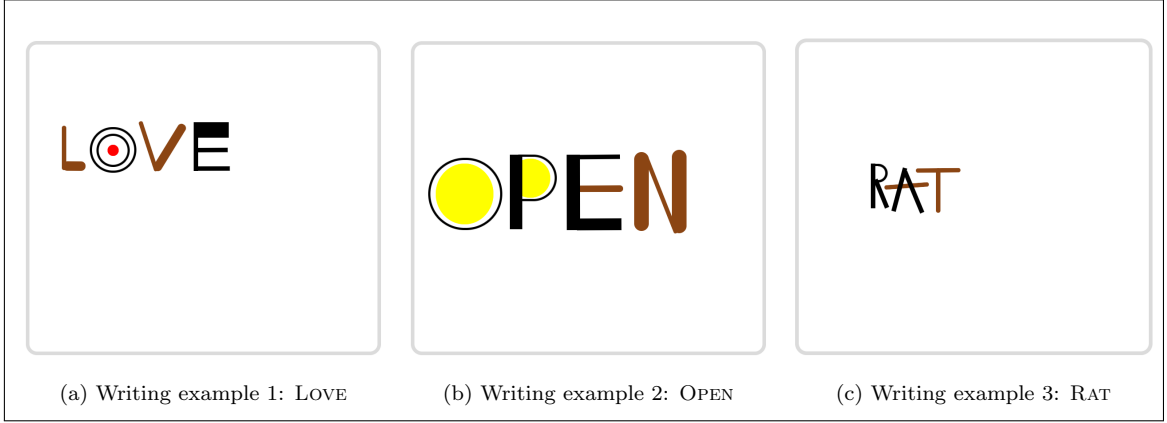


Figure A1: Exemplary writing illustrations

NOTE: Three example ideas that were excluded from our analyses as ideators used the materials to write words, not to illustrate them.

Table A1 shows some summary statistics of our control variables. The nominal bonus is the proposed nominal bonus assigned to ideators that translates into the realized payoff given an ideator's performance. The inactivity dummy is defined as one if the respective ideator has seen at least one warning message due to an inactivity of at least three minutes. The age dummy is defined as one if the ideators reported any age and if this age is above the median of the whole sample, i.e. a reported age of 35 years. The gender dummy is defined as one if the ideators reported any gender and if this gender is male. The risk aversion dummy is defined as one if the general risk level is not bigger than the median, i.e. five, on our eleven-point scale. The weekly work time is given in hours whereas the net income is reported categorically in £500 bins. The low (high) potential dummy is defined as one if the ideators reported a creative potential that is lower than the first tercile (higher than the second tercile), i.e. 27 (39) on our 73-point aggregate measure.

Table A1: Summary statistics of control variables

	Obs.	All	Fixed	Input	Output	Diff.
Nominal bonus ^c	325.	2.99 (1.20)	2.89 (1.19)	3.15 (1.15)	2.96 (1.24)	
Inactivity (dummy) ^a	325.	0.09 (0.29)	0.08 (0.28)	0.07 (0.26)	0.11 (0.31)	
Age (dummy) ^d	325.	0.43 (0.50)	0.42 (0.50)	0.41 (0.49)	0.47 (0.50)	
Gender (dummy) ^d	325.	0.50 (0.50)	0.51 (0.50)	0.42 (0.50)	0.57 (0.50)	*
Risk aversion (dummy) ^b	325.	0.57 (0.50)	0.50 (0.50)	0.61 (0.49)	0.58 (0.49)	
Weekly work time ^b	325.	30.41 (15.97)	29. (15.81)	30.96 (15.54)	31.19 (16.52)	
Monthly net income (categorical) ^b	325.	5.28 (3.63)	4.97 (3.59)	5.51 (3.62)	5.36 (3.69)	
Low potential (dummy) ^b	325.	0.30 (0.46)	0.33 (0.47)	0.37 (0.48)	0.21 (0.41)	**
High potential (dummy) ^b	325.	0.33 (0.47)	0.27 (0.45)	0.34 (0.47)	0.38 (0.49)	

NOTE: The table reports the mean of the fixed, input, and output treatment groups together, and for each treatment group separately. Standard deviations are reported in parentheses. ^a Variables have been collected via our Word Illustration Task (WIT). See figures ?? to ?? in the supplementary analyses for the exact wording of the different questionnaire measures. ^b Variable has been randomly assigned by us. ^c Variable has been randomly assigned by us. ^d Variables have been collected by Prolific. We coded all missing values as zero (young, female). The first column shows the number of observations. There are only 304 observations for our object number measure as there is no observation from participants who did not submit any illustration. The last column reports the significance level from Kruskal-Wallis tests (for categorical and continuous variables) and χ^2 tests (for dummy variables), respectively, on treatment group differences, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.2 Results

Figure A2 shows the histograms of our three main dependent variables, i.e. effort duration, effort intensity, and the number of innovative ideas. The figure shows that the number of innovative ideas has a strongly skewed distribution rendering OLS inappropriate. This is why we use a Negative Binomial model in table 4 instead.

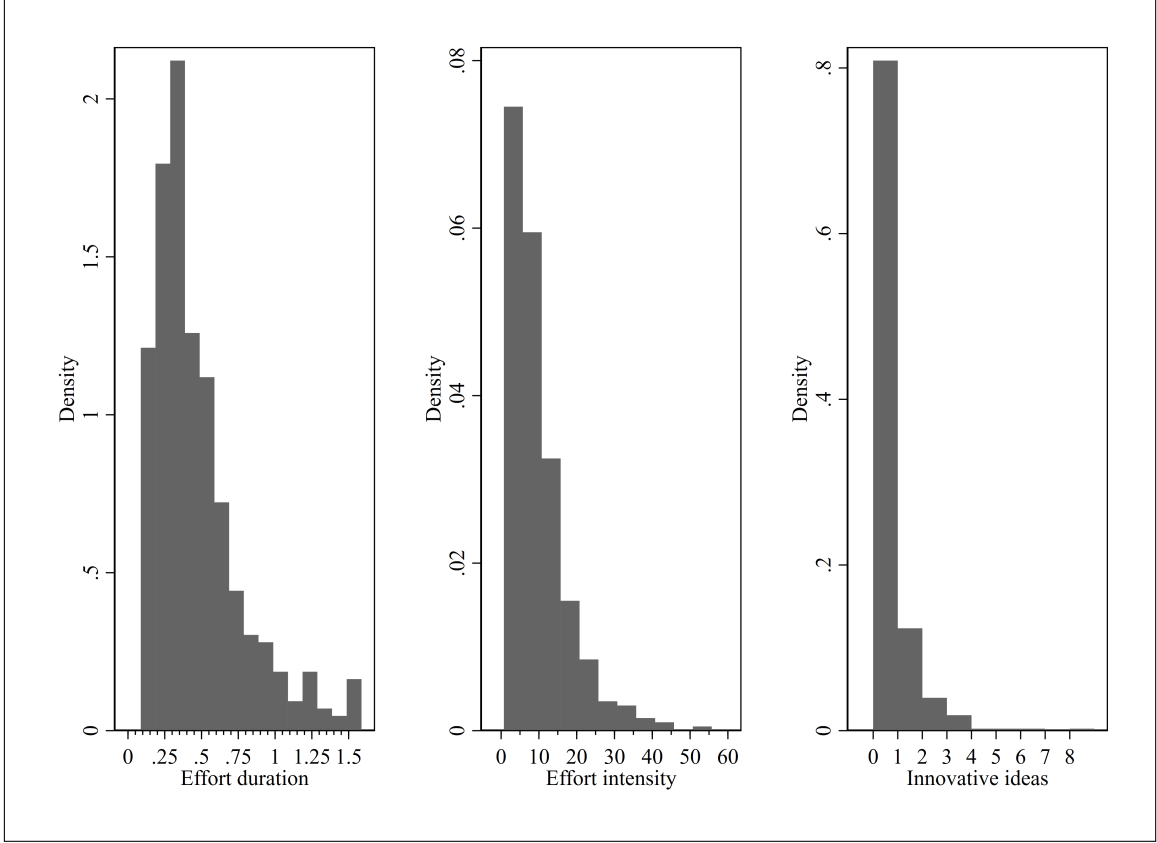


Figure A2: Histograms of the main dependent variables

NOTE: Histograms for the total working time (in hours) per ideator (effort duration, $n=429$), the number of ideas per hour working time per ideator (effort intensity, $n=400$), and the number of innovative ideas (innovative ideas, $n=429$) per ideator.

Table A2 shows the estimation results for the models in columns (1) and (2) of table 4 but when splitting the results by both dimensions separately. That is, the models in columns (1) and (2) of table A2 use the number of original ideas as the dependent variable while the models in columns (3) and (4) take the high quality ideas. One can see that the positive treatment effect of output incentives is mainly driven by the originality dimension. That is, ideators in the output treatment strongly increase the number of original ideas. While keeping the number of high quality ideas rather constant, they can finally increase the number of innovative ideas as shown in section 3. By contrast, ideators in the input treatment concentrate less on one of the two dimensions but increase the number of both, original ideas and high quality ideas. This increase is of identical size. However, the increase in original ideas is smaller for the input treatment than for the output treatment.

Table A2: Negative Binomial regressions: Original and high quality ideas

	(1)	(2)	(3)	(4)
	Original ideas		High quality ideas	
Input	0.4023** (0.1683)	0.4213** (0.1648)	0.4181* (0.2352)	0.4256* (0.2292)
Output	0.7214*** (0.1533)	0.6334*** (0.1480)	0.2716 (0.2064)	0.3322 (0.2020)
Constant	0.5623*** (0.1066)	0.6637*** (0.2002)	−0.3327** (0.1601)	0.0041 (0.2571)
Wald tests on coefficient differences:				
Input – output	−0.3191*	−0.2121	0.1465	0.0935
Controls		✓		✓
Observations	325	325	325	325
Pseudo R^2	0.0167	0.0333	0.0044	0.0224

NOTE: Negative Binomial regression results with number of original ideas per ideator as dependent variable in models (1) and (2), and number of high quality ideas as dependent variable in models (3) and (4). We include treatment dummies with the fixed treatment as reference group. In the models (2) and (4), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In table 5, we have shown the treatment effects on the number of ideas and the share of original, and high quality ideas, respectively. For completeness, table A3 below shows the share of innovative ideas. The results point to the same direction as before, i.e. the treatment effects mostly come from a quantity effect and lesser through an increase in the share of innovative ideas. As the results of table A3 show, there is no difference between the three treatment groups detectable.

Table A3: OLS regressions: Number of ideas

	(1)	(2)
Input	−0.0051 (0.0210)	−0.0031 (0.0208)
Output	0.0006 (0.0205)	0.0016 (0.0207)
Constant	0.0581*** (0.0157)	0.1116*** (0.0296)
Wald tests on coefficient differences:		
Difference input − output	−0.0057	−0.0047
Controls		✓
Observations	304	304
Adjusted R^2	−0.0063	0.0151

NOTE: OLS regression results with share of innovative ideas out of all ideas per ideator as the dependent variable. We include treatment dummies with the fixed treatment as reference group. In model (2), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4 shows the estimation results from a Generalized Structural Equation Model (GSEM). We use this model to estimate the part of our found treatment effects that is mediated through our measures of effort duration and effort intensity. Thus, we include these two as our mediator variables whereas the treatment dummies are the exogenous variables, and the number of innovative ideas is the endogenous variable. We show in columns (2) and (3) that our dummies are correlated with the two effort measures and in column (1) that these effort measures are correlated with the number of innovative ideas. As the dummies are not significant in column (1), we see that it is a full mediation of the treatment effects through effort duration, and effort intensity. The same finding also holds when introducing our usual set of control variables as in columns (4) to (6).

Table A4: GSEM: Decomposition of treatment effects

	(1)	(2)	(3)	(4)	(5)	(6)
Exogenous variables:	Innovative ideas	Effort duration	Effort intensity	Innovative ideas	Effort duration	Effort intensity
Input	0.0683 (0.1024)	0.6480*** (0.1688)	-0.0646 (0.1673)	0.0710 (0.1028)	0.6736*** (0.1606)	-0.0829 (0.1654)
Output	-0.0452 (0.0993)	0.4016** (0.1624)	0.4653*** (0.1623)	-0.0321 (0.0999)	0.3744** (0.1543)	0.4533*** (0.1615)
Mediator variables:						
Standardized effort duration	0.2549*** (0.0332)			0.2688*** (0.0350)		
Standardized effort intensity	0.2536*** (0.0347)			0.2509*** (0.0352)		
Constant	0.2057*** (0.0708)	0.0000 (0.1179)	0.0000 (0.1183)	0.3755** (0.1579)	0.0498 (0.2502)	-0.2544 (0.2601)
Controls				✓	✓	✓
Observations	325	325	325	325	325	325

NOTE: Generalized Structural Equation Model (GSEM) estimation results with number of innovative ideas as endogenous variable. We include treatment dummies as exogenous variables with the fixed treatment as reference group. In the model of column (4) to (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Mediator variables are the standardized effort duration and standardized effort intensity. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We can also create a similar graph like in figure 6 for the findings regarding the material usage for innovative ideas that is reported in figure 7. Figure A3 shows the estimated coefficients for the indirect treatment effects together with the bootstrapped p-values as before and table A5 shows the results from the underlying GSEM. As can be seen, the material reduction of the output treatment stems from the indirect effort intensity effect but different to figure 6, there is now also a direct output effect. That is, the material reduction effect of the output treatment is partly mediated through the observed change in effort intensity but partly also driven by further aspects not explained by our two effort measures.

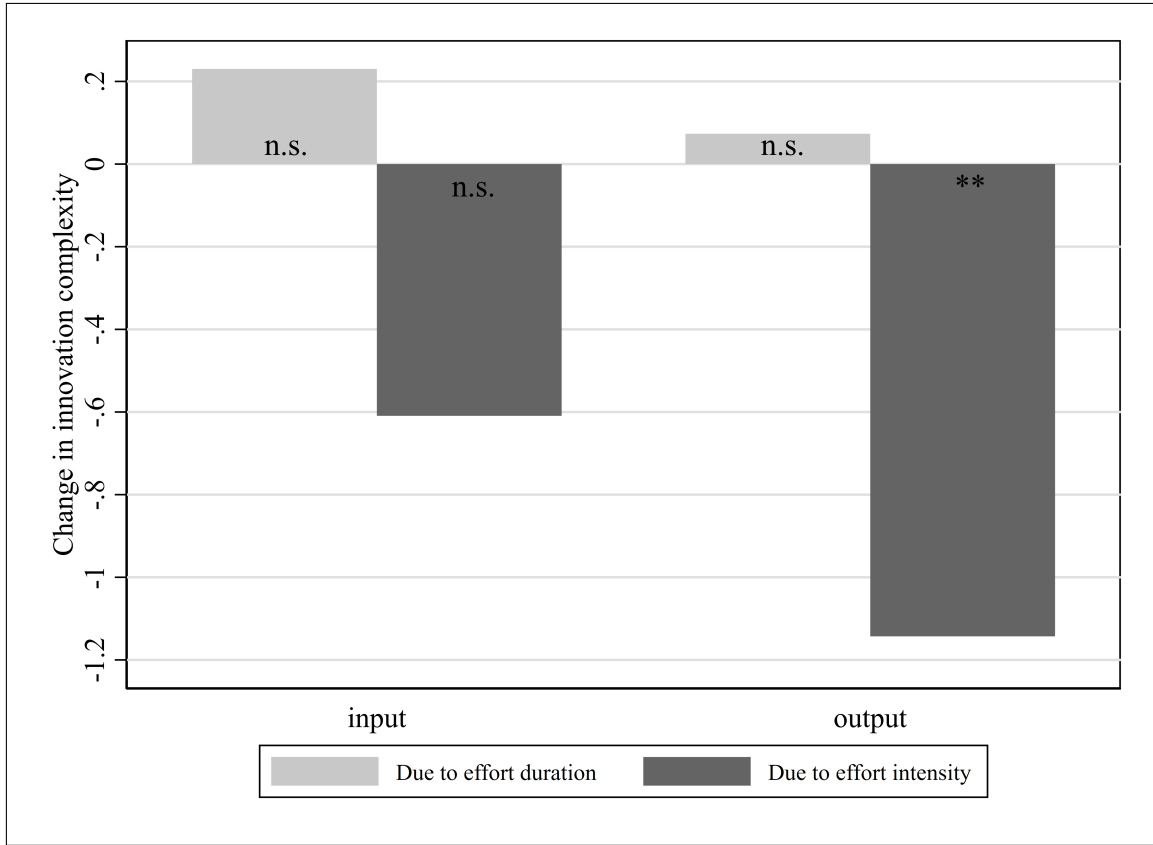


Figure A3: Decomposition of the treatment effects (innovation complexity)

NOTE: Estimated coefficients via a generalized structural equation model (GSEM) as shown in columns (1) to (3) of table A5. Input and output are defined as exogenous variables, standardized effort duration and effort intensity are the mediators, and the number of used materials per innovative idea is the endogenous variable. The effect "due to effort duration" is the indirect effort duration effect, and "due to effort intensity" is the indirect effort intensity effect, respectively. Asterisks report significance levels of the respective coefficients calculated via bootstrapping with 500 repetitions, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

Table A5: GSEM: Decomposition of treatment effects (innovation complexity)

	(1)	(2)	(3)	(4)	(5)	(6)
Exogenous variables:	Innovation complexity	Standardized effort duration	Standardized effort intensity	Innovation complexity	Standardized effort duration	Standardized effort intensity
Input	0.7709 (1.3314)	1.0879** (0.4400)	0.4873 (0.4422)	0.9955 (1.2774)	0.9381** (0.4118)	0.3973 (0.4273)
Output	-2.4503** (1.2052)	0.3467 (0.4041)	0.9142** (0.4062)	-2.3109* (1.1802)	0.2880 (0.3819)	0.9391** (0.3922)
Mediator variables:						
Standardized effort duration	0.2116 (0.3650)			0.1793 (0.3775)		
Standardized effort intensity	-1.2503*** (0.3632)			-1.3322*** (0.3589)		
Constant	10.8930*** (0.9262)	0.5106 (0.3153)	0.3563 (0.3169)	11.6489*** (1.5165)	0.4767 (0.5175)	-1.0500* (0.6222)
Controls				✓	✓	✓
Observations	65	65	65	65	65	65

NOTE: Generalized Structural Equation Model (GSEM) estimation results with the mean innovation complexity per ideator as endogenous variable. We include treatment dummies as exogenous variables with the fixed treatment as reference group. In the model of column (4) to (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Mediator variables are the standardized effort duration and standardized effort intensity. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.3 Complexity

Our finding reported in section 3 can be separated in two distinct changes: a change in the average complexity of the words that are chosen to be illustrated, and a change in the relative idea complexity when holding the word constant. Table A6 shows that ideators in the output treatment systematically choose to illustrate ideas with a lower average word complexity than ideators in the input treatment. This effect is mostly explained by the observed difference in effort intensity. Importantly, the effect is not explainable by a quality change that could be associated with the average word complexity change as the estimates remain qualitatively unchanged when including our quality control. This quality control takes the average quality of a word across all ideators. That is, it takes the average quality by word and averages this on the ideator level. Thus, this variable captures the average quality of the words that ideators choose to illustrate.

Second, there is also a change in the relative idea complexity. Table A7 shows that ideators in the output treatment also show a more negative relative idea complexity than those in the input treatment. Again, this effect cannot be explained by quality changes but mostly by changes in effort intensity. As the relative idea complexity is defined not only by word but by idea, the quality control is simply the average quality per ideator across all own ideas.

Table A6: OLS regressions: Average word complexity

	(1)	(2)	(3)	(4)	(5)
Input	0.2584 (0.4443)	0.2953 (0.4471)	0.2142 (0.4712)	0.2659 (0.4026)	0.2831 (0.4208)
Output	-0.5302 (0.4182)	-0.5533 (0.4279)	-0.5977 (0.4398)	-0.0491 (0.3944)	-0.0377 (0.4108)
Standardized effort duration			0.1243 (0.1582)		-0.0266 (0.1444)
Standardized effort intensity				-0.9957*** (0.1035)	-0.9995*** (0.1047)
Average word quality		1.8006** (0.7825)	1.7284** (0.7815)	1.7534** (0.7447)	1.7687** (0.7500)
Constant	9.3903*** (0.3273)	9.0630*** (0.6014)	9.0618*** (0.6035)	9.2341*** (0.5641)	9.2350*** (0.5647)
Wald tests on coefficient differences:					
Difference input – output	0.7886**	0.8486**	0.8119**	0.3149	0.3208
Controls		✓	✓	✓	✓
Observations	270	270	270	270	270
Adjusted R^2	0.0068	0.0113	0.0101	0.1856	0.1826

NOTE: OLS regression results with average word complexity per ideator as dependent variable. Average word complexity is defined as the full-sample average number of objects for a given word. We include treatment dummies with the fixed treatment as reference group. In the models (3) and (5), we control for standardized effort duration, in models (4) and (5), we control for standardized effort intensity, and in models (2) to (5), we control for average word quality. In the models (2) to (5), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. The average word complexity measure is not available for the 21 ideators who submitted no illustration nor for the 34 ideators who only submitted illustrations of words that nobody else illustrated thereby reducing the number of observations as compared to tables 3 and 4. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: OLS regressions: Relative idea complexity

	(1)	(2)	(3)	(4)	(5)
Input	0.0085 (0.0539)	−0.0232 (0.0605)	−0.0766 (0.0570)	−0.0313 (0.0561)	−0.0775 (0.0521)
Output	−0.1299*** (0.0476)	−0.1601*** (0.0546)	−0.1884*** (0.0546)	−0.1034** (0.0500)	−0.1311*** (0.0497)
Standardized effort duration			0.0825*** (0.0251)		0.0720*** (0.0225)
Standardized effort intensity				−0.1132*** (0.0157)	−0.1073*** (0.0146)
Average idea quality		0.1407* (0.0847)	0.1038 (0.0851)	0.1371* (0.0786)	0.1051 (0.0782)
Constant	0.1173*** (0.0386)	0.1256 (0.0859)	0.1438* (0.0842)	0.1406* (0.0807)	0.1557** (0.0786)
Wald tests on coefficient differences:					
Difference input − output	0.1384***	0.1368**	0.1118**	0.0721	0.0536
Controls		✓	✓	✓	✓
Observations	270	270	270	270	270
Adjusted R^2	0.0307	0.0335	0.0965	0.1822	0.2296

NOTE: OLS regression results with average relative idea complexity per ideator as dependent variable. Relative idea complexity is defined as the percental deviation from the full-sample average number of objects for a given word (expressed in decimal numbers). We include treatment dummies with the fixed treatment as reference group. In the models (3) and (5), we control for standardized effort duration, in models (4) and (5), we control for standardized effort intensity, and in models (2) to (5), we control for average word quality. In the models (2) to (5), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. The average word complexity measure is not available for the 21 ideators who submitted no illustration nor for the 34 ideators who only submitted illustrations of words that nobody else illustrated thereby reducing the number of observations as compared to tables 3 and 4. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A4 shows that our findings of the right panel of figure 7 are not limited to the set of innovative ideas but that this pattern holds also for the recurring ideas, i.e. those ideas that are used for our complexity measures. As figure A4 shows it is again the output incentive that makes ideators use significantly less materials per (recurring) idea as compared to any other group (MWU fixed vs. output: $p = 0.040$, input vs. output: $p < 0.001$). The other two groups show no difference, however (MWU fixed vs. input: $p = 0.352$).

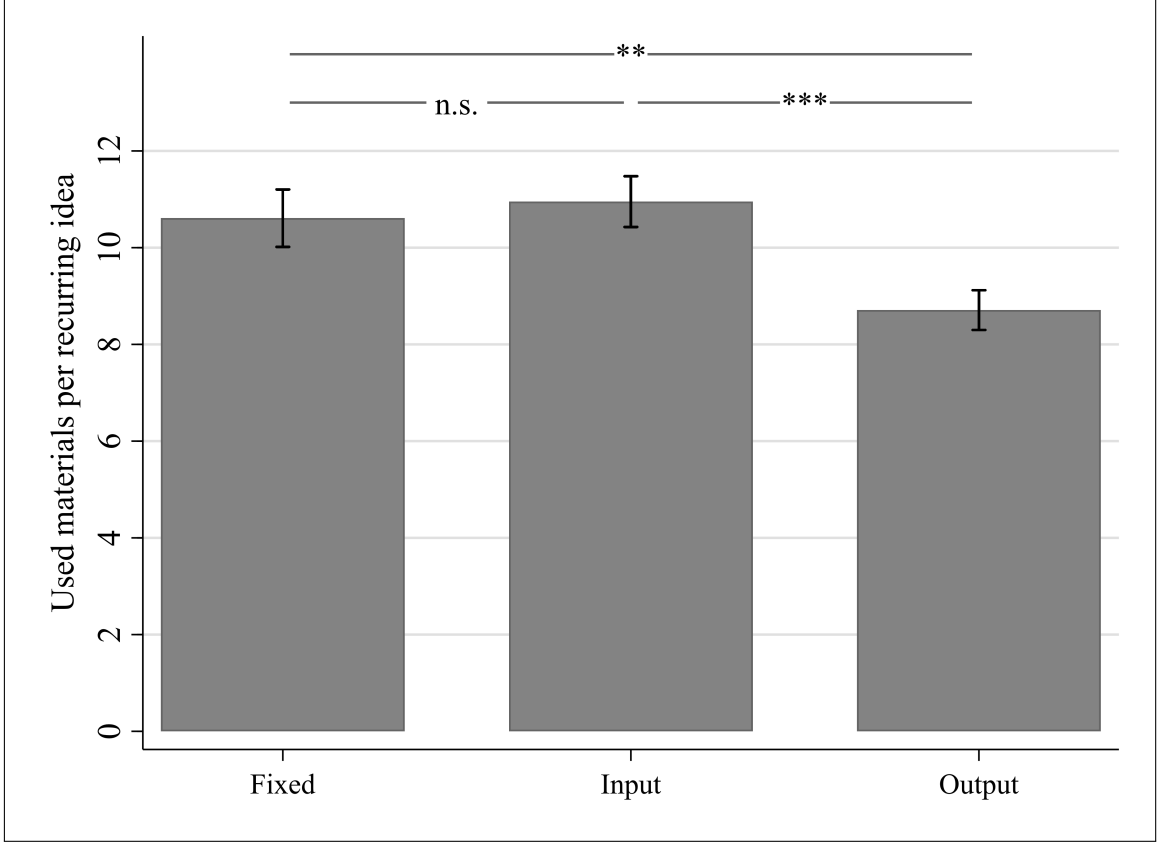


Figure A4: Types of recurring ideas

NOTE: Mean number of used materials per recurring idea (n=270). Standard errors are shown as vertical lines. Pairwise MWU test results are reported as horizontal lines above the bars, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p \geq 0.1$.

We have shown above that the output treatment makes ideators reducing their average word complexity. The question is whether this is simply an additional feature or whether this reduction in complexity has also an economic impact. To investigate this aspect, we employ again the benefit of our experimental design with which we can clearly differentiate the two dimensions of originality and quality. We turn to the idea level to investigate whether the average word complexity or the relative idea complexity systematically changes whether an idea is original and/or of high quality and/or innovative. First, we look at the average word complexity. This measure is defined on the word level. Hence, our Logit regressions in table A8 include each word as one observation. While the originality is also set on the word level, the quality and thus, the innovativeness differ from idea to idea. We use the average high quality dummy and the average innovation dummy per word to account for this. As table A8 shows, there is no systematic difference associated with a change in the average word complexity. That is, both simple and more complex words may, e.g., be original or not, but it is not the case that complex words are generally more or less likely to be original. The same holds also for the likelihood of being high quality, and innovative.

Table A8: Logit regressions: Idea dimensions per recurring word by average word complexity

	(1) Average idea is innovative	(2) Average idea is original	(3) Average idea is of high quality
Average word complexity	−0.0167 (0.0699)	−0.0321 (0.0556)	0.0261 (0.0557)
Constant	−1.0707* (0.6150)	1.1466** (0.5007)	−0.6622 (0.4975)
Observations	227	227	227
Pseudo R^2	0.0003	0.0011	0.0007

NOTE: Logit regression results with average innovation dummy per word as the dependent variable in model (1), average original dummy per word as the dependent variable in model (2), and average high quality dummy per word as the dependent variable in model (3). We include average word complexity in all models. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Second, table A9 looks at the relative idea complexity. This variable is defined per idea so that we can include each idea as a separate observation. However, there can be systematic differences with respect to the words. We employ a random effects model with the random effects being calculated on the word level to account for this. Again, we do not find a significant effect for the originality of an idea. This is intuitive as the originality is set by word and should not differ with relative idea complexity per word. However, the results in column (3) show that ideas with a higher relative idea complexity are significantly more likely to be a high quality idea. This effect is not strong enough to be detectable on the innovativeness of ideas as well but it shows that a higher relative idea complexity is not only a wastage of materials but that it can also positively impact the quality of an idea.

Table A9: Logit RE regressions: Idea dimensions by relative idea complexity

	(1) Idea is innovative	(2) Idea is original	(3) Idea is of high quality
Relative idea complexity	0.7952 (0.6353)	−0.1249 (0.2131)	1.3551*** (0.4182)
Constant	−4.0962*** (0.4441)	−3.9449 (0.0000)	−2.4343*** (0.2856)
Observations	970	970	970
Word groups	227	227	227

NOTE: Logit regression results with innovation dummy per idea as the dependent variable in model (1), original dummy per idea as the dependent variable in model (2), and high quality dummy per idea as the dependent variable in model (3). We include relative idea complexity in all models. Random effects are calculated on the word level. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.4 Robustness checks on payment size effects

In the main analyses, we have investigated the average treatment effect of different incentive schemes. Additionally, we can also investigate whether this effect is actually driven by the type of incentive or also its payment size as we did not only vary the type of incentive but also its nominal bonus size. As a first exercise, we include the actually drawn nominal bonus as our control for the size, i.e. the number from the interval $[1, 5]$ as shown in table 1. Table A10 replicates the regression results from tables 3 and 4 including the drawn nominal bonus per ideator. As the results show, our results remain qualitatively unchanged. That is, it is the type of incentive scheme and not its assigned nominal bonus size that creates our treatment effects.

Table A10: OLS and Negative Binomial regressions: Nominal bonus control

	(1) Effort duration	(2) Effort duration	(3) Effort intensity	(4) Effort intensity	(5) Innovative ideas	(6) Innovative ideas
Input	0.1547*** (0.0426)	0.1688*** (0.0414)	−0.5567 (0.9744)	−0.5393 (0.9913)	0.7240** (0.3513)	0.6951** (0.3396)
Output	0.0986*** (0.0372)	0.0938*** (0.0354)	3.0518*** (1.0644)	2.9479*** (1.0852)	0.5436* (0.2959)	0.5514* (0.2967)
Nominal bonus	0.0294** (0.0130)	0.0251** (0.0127)	0.7527** (0.3517)	0.6450* (0.3490)	0.0853 (0.1199)	0.1027 (0.1069)
Constant	0.3117*** (0.0449)	0.4091*** (0.0666)	6.5293*** (1.1793)	7.0984*** (1.7815)	−1.8674*** (0.4174)	−1.3026** (0.5119)
Wald tests on coefficient differences:						
Input − output	0.0561	0.0750*	−3.6085***	−3.4872***	0.1804	0.1437
Controls		✓		✓		✓
Observations	325	325	304	304	325	325
Adjusted R^2	0.0483	0.1370	0.0447	0.0558		
Pseudo R^2					0.0111	0.0366

NOTE: OLS regression results with effort duration (working time in hours) per ideator as dependent variable in models (1) and (2), and effort intensity (number of ideas divided by effort duration) as dependent variable in models (3) and (4). Negative Binomial regression results with number of innovative ideas per ideator as dependent variable in models (5) and (6). We include treatment dummies with the fixed treatment as reference group and the nominal drawn bonus per ideator. In the models (2), (4) and (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Despite the two pretests that we used to calibrate our treatments, the nominal bonuses created different average payments that could be realized. Ideators in the fixed treatment earned on average £4.14, those in the input treatment earned £3.02, and those in the output treatment earned £2.40, respectively (including the show-up fee of £1.25 each). Thus, a nominal fixed pay bonus of, e.g., £3.00 is perceived as an expected payment of the same size, while a nominal input or output bonus

of that size could have been perceived to be worth less. To rule out that this drives our results, we calculate hypothetical, real bonuses that account for these realized payoffs. To do so, we use the average working time and average number of innovative ideas from the baseline group for reference and deflate the nominal input and output bonus accordingly. That is, we create hypothetical, real bonuses that would be payoff-equivalent given the behavior of ideators in the baseline group. For instance, the average ideator in the baseline group worked 0.42 hours. We multiply the nominal input bonus, i.e. the number between 1.00 and 5.00, with 0.42 so that the new payoff-equivalent real bonus ranges between 0.42 and 2.10 (and similarly with the output bonus). By contrast, we keep the fixed bonus identical. Thus, a fixed real bonus of £1.00 is payoff-equivalent to an input real bonus of £0.42 given the performance in the baseline group. Table A11 repeats our main regression analyses but uses this real bonus as an alternative control variable. Comparing them to the results of tables 3 and 4, we find the same or even stronger effects.

Table A11: OLS and Negative Binomial regressions: Real bonus control

	(1) Effort duration	(2) Effort duration	(3) Effort intensity	(4) Effort intensity	(5) Innovative ideas	(6) Innovative ideas
Input	0.2061*** (0.0521)	0.2204*** (0.0510)	1.1770 (1.2849)	0.8872 (1.2385)	0.8717** (0.4089)	0.9284** (0.4099)
Output	0.1564*** (0.0546)	0.1525*** (0.0514)	5.0485*** (1.4421)	4.5773*** (1.4159)	0.7221 (0.4755)	0.8324* (0.4295)
Real bonus	0.0278 (0.0190)	0.0284 (0.0188)	0.9705* (0.4948)	0.7899* (0.4774)	0.0810 (0.1787)	0.1317 (0.1596)
Constant	0.3164*** (0.0602)	0.3983*** (0.0782)	5.8857*** (1.5001)	6.6453*** (1.9558)	-1.8546*** (0.5671)	-1.4069** (0.6131)
Wald tests on coefficient differences:						
Input – output	0.0497	0.0679	-3.8715***	-3.6901***	0.1497	0.0960
Controls		✓		✓		✓
Observations	325	325	304	304	325	325
Adjusted R^2	0.0401	0.1326	0.0401	0.0518		
Pseudo R^2					0.0103	0.0359

NOTE: OLS regression results with effort duration (working time in hours) per ideator as dependent variable in models (1) and (2), and effort intensity (number of ideas divided by effort duration) as dependent variable in models (3) and (4). Negative Binomial regression results with number of innovative ideas per ideator as dependent variable in models (5) and (6). We include treatment dummies with the fixed treatment as reference group and the real payoff-equivalent bonus per ideator. In the models (2), (4) and (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Finally, we can show that the bonus size does not even matter when the results could be interpreted as an example of gift exchange. We compare the fixed with the baseline group for this exercise, i.e. those two groups that received a fixed wage only but where the size differed. The baseline group

received the fixed wage that has been advertised on Prolific for our study (see figure ??) while the fixed treatment has received an additional fixed wage on top of this. Thus, ideators in the fixed treatment could show a positive effect to reciprocate this gift. Table A12 shows the results. It shows again no effect of the nominal bonus size on any of our variables. Thus, we can conclude in line with previous research (e.g., Bradler et al., 2019) that gift exchange cannot be observed in the context of idea generation tasks.

Table A12: OLS and Negative Binomial regressions: Comparison Baseline and Fixed

	(1)	(2)	(3)	(4)	(5)	(6)
	Effort duration		Effort intensity		Innovative ideas	
Fixed	-0.0208 (0.0352)	-0.0093 (0.0359)	-0.6577 (1.0429)	-0.6732 (1.0036)	-0.4085 (0.4058)	-0.3843 (0.3695)
Constant	25.0421*** (1.5269)	31.4253*** (3.3078)	9.4107*** (0.8073)	9.2829*** (1.8361)	-1.2104*** (0.3357)	-0.6469 (0.5492)
Controls		✓		✓		✓
Observations	210	210	193	193	210	210
Adjusted R^2	-0.0031	0.0657	-0.0031	-0.0071		
Pseudo R^2					0.0046	0.0414

NOTE: OLS regression results with effort duration (working time in hours) per ideator as dependent variable in models (1) and (2), and effort intensity (number of ideas divided by effort duration) as dependent variable in models (3) and (4). Negative Binomial regression results with number of innovative ideas per ideator as dependent variable in models (5) and (6). We include a fixed pay treatment dummy with the baseline treatment as reference group. In the models (2), (4) and (6), we further control for the creative potential, age, gender, risk preferences, hours of work, income, and inactivity. See appendix for details on control variables. Coefficients of the control variables are reported in table ?? in the supplementary analyses. Robust standard errors are reported in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.