



Gutenberg School of Management and Economics
& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

***Incentives and Creativity in Groups —
Experimental Evidence on Creative
Processes and Dimensions***

Erik Sarrazin

15th January 2026

Discussion paper number 2601

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
<https://wiwi.uni-mainz.de/>

Contact details

Erik Sarrazin
Chair of Public and Behavioral Economics
Johannes Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz
Germany
sarrazin@uni-mainz.de

Incentives and Creativity in Groups— Experimental Evidence on Creative Processes and Dimensions*

Erik Sarrazin[†]

January 15, 2026

Abstract

Creativity and teamwork are essential in today’s rapidly evolving labor market, yet little is known about how incentives shape creative group processes across multiple creativity dimensions, specifically quantity, quality, and originality. I introduce a novel verbal creative coordination task to incentivize and objectively measure these dimensions of creativity and to generate insights into the full creative group process. Thereby, this paper investigates how idea generation, evaluation, and selection can be effectively incentivized in groups and which incentive scheme maximizes innovation—ideas that are both high in quality and originality. In a laboratory experiment with 640 participants, groups are randomly assigned to a control or one of three group-level relative performance pay treatments targeting either quantity, quality, or originality. Results show that quantity incentives lead to broader exploration, increasing both the number and average originality of ideas, and outperform all other treatments on the combined indicator of innovative ideas. Quality incentives improve idea quality, while originality incentives fail to boost originality. Incentives mainly operate through higher individual effort rather than group dynamics. Across all conditions, the evaluation and selection phases act as a bottleneck, with highly original ideas systematically discarded. These findings highlight that maximizing creative output requires incentives that promote exploration and structures that preserve original ideas.

Keywords: creativity · innovation · incentives · teamwork · laboratory experiment

JEL classification: O31 · M52 · C92 · D02

*Pre-registration: AsPredicted #208,106. IRB Approval by the Joint Ethics Committee in Economics at the University of Frankfurt and the University of Mainz on November 14, 2023. I thank Peter Andre, Volker Benndorf, David Blum, Raphael Brade, Tor Eriksson, Anders Frederiksen, Katharina Hartinger, Larry Hedges, Holger Herz, Ariel Kalil, Lavinia Kinne, Fabian Kosse, Maria Krempf, Anna Ressi, Daniel Schunk, Niklas Witzig, Isabell Zipperle, and participants of the PhD Workshop of the University of Chicago Harris School of Public Policy, the Swiss Leading House course program, the GSEFM course program, and the IPP Ideas Crunch for helpful comments and suggestions. I gratefully acknowledge funding from the interdisciplinary research unit IPP at the University of Mainz. I also thank João Andrade, Sebastian Auer, Ramon Berg, Annika Göttert, and Nellie Kremer for excellent research assistance. All errors are my own.

[†]Johannes Gutenberg University Mainz, sarrazin@uni-mainz.de, ORCID-ID: 0009-0005-7480-3804

1 Introduction

Creativity, defined as "the production of novel and useful ideas" (Amabile et al., 1996), is an essential skill for both successful job performance and everyday problem-solving. At the micro-level, it serves as the foundation for inventions, which in turn lead to innovations (Arthur, 2007). At the macro-level, innovation drives productivity and long-term economic growth (Romer, 1990; Grossman and Helpman, 1994; Sampson, 2023). The relevance of creativity has been increasing for decades: as automation takes over routine tasks, non-routine tasks demanding creativity and flexibility have grown in importance (Autor et al., 2003; Autor and Price, 2013). Ongoing technological change, the transition to a sustainable economy, shifting geoeconomic conditions, and demographic change are rapidly reshaping today's labor markets, altering the tasks and skills required of workers. Amid these transformations, creativity stands out as a key skill of growing future importance as employers consider it among the skills least likely to be substituted by generative AI (Leopold et al., 2025).

At the same time, work has become increasingly collaborative, with employees spending roughly 50 percent more time on teamwork than two decades ago (Cross et al., 2016). Reflecting this trend, employers identify problem-solving and teamwork as the most critical yet scarce skills in new hires (Deming and Silliman, 2024). As organizations rely more on teams to generate ideas and solve problems, their effectiveness in creative tasks becomes increasingly consequential. Yet, despite the potential of groups to leverage diverse perspectives and complementary skills, groups often underperform in creative tasks relative to the combined output of individuals working independently (e.g., Gneezy et al., 2022). This underperformance arises from cognitive and social frictions inherent in group interaction, including production blocking, evaluation apprehension, and free-riding (Paulus et al., 2002). Even when groups generate creative ideas, they frequently fail to recognize and select highly original contributions, with originality perceived as too risky, infeasible, or uncertain (Rietzschel et al., 2019).

Despite these insights, existing research has focused predominantly on the incentivization of *individual* creativity, leaving creative group processes comparatively understudied. Specifically, little is known about how groups generate, evaluate, refine, and select ideas, and how incentive structures shape these stages across multiple dimensions of creativity. While most prior studies concentrate on idea generation, the subsequent stages of evaluation and selection, which are critical for translating creative ideas into implementable solutions, remain underexplored. This paper addresses these gaps by investigating how idea generation and selection can be effectively incentivized in groups, and which incentive scheme maximizes overall innovation, defined as ideas that are both high in quality and originality. It further explores the mechanisms through which incentives shape group-level creative outcomes and examines how groups navigate trade-offs between quantity, quality, and originality.

To do so, I introduce a novel verbal creative coordination task that captures the full creative process, including idea generation, evaluation, revision, and selection. The task allows for the objective measurement of multiple dimensions of individual and group creative per-

formance: quantity, quality (usefulness), and originality (novelty). This enables a deeper understanding of creativity in all its facets and the design of incentives that effectively stimulate all aspects of creative performance. Unlike many classical creativity tasks in psychology, this task explicitly considers the usefulness of ideas, while also mirroring common organizational brainstorming practices. Using a controlled laboratory design, this paper provides the first evidence on incentivizing and measuring different dimensions of group creative performance.

The novel task centers around verbal creativity and requires groups to generate as many clue pairs (i.e., two words) as possible that are both original and useful for a randomly assigned assessor to identify a target word. At the start of each round, participants are shown the target word and three taboo words that are not permitted to appear in any clue pair. The task unfolds in four sequential phases: (1) *Individual Ideation* – participants generate candidate clue pairs independently; (2) *Group Evaluation* – participants share their clues and provide feedback and ratings to one another; (3) *Individual Revision* – participants revise clue pairs based on the received feedback; (4) *Group Selection* – the group coordinates via chat and selects a final clue pair for assessment.

Participants are randomly assigned to one of four experimental conditions: a Control group (flat payment) or one of three relative performance pay group-level incentive conditions targeting either Originality, Quality, or Quantity. Relative performance pay is frequently used in the context of innovation because it can generate efficient outcomes while requiring information only on relative rankings rather than absolute performance (Lazear and Rosen, 1981). The design isolates the causal effect of rewarding a specific creative dimension while keeping all other elements of the task constant. Participants in the three Treatment groups can earn a performance-based bonus, determined by their group’s rank relative to the other four groups in the session.

The laboratory experiment, conducted between January and June 2025, includes 640 participants: 480 ideators organized into 160 groups of three, and 160 quality assessors. The dataset comprises 2,880 individual-level observations (480 ideators \times 6 rounds) and 960 group-level observations (160 groups \times 6 rounds). *Originality* is operationalized as the semantic distance between a clue pair and the target word, computed using pre-trained fastText word vectors (Grave et al., 2018) to capture fine-grained conceptual similarity and divergence. Higher semantic distance indicates greater cognitive flexibility and associative divergence (Olson et al., 2021). *Quality* captures how effectively the final clue pair enables a quality assessor to correctly identify the target word, measured by the number of attempts required. *Quantity* is measured as the total number of clue pairs generated by all group members per target word. The measures of *Innovation* combine quality and originality of the final clue pair, as a standardized average index and as a binary indicator for top-percentile ideas.

Consistent with classic microeconomic labor supply theory, I hypothesize that incentives increase effort, thereby improving performance in the incentivized dimensions. At the same time, incentives may produce spillover effects on non-incentivized dimensions, which could be either positive or negative. I also examine whether groups exhibit a systematic bias against

originality during evaluation and selection, and whether such bias can be mitigated through targeted incentives.

Results show that quantity-based incentives are the most robust and consistent driver of innovation. Incentivizing quantity substantially increases both the number of ideas and, indirectly, average originality, consistently outperforming all other payment schemes on the binary indicator of the most innovative ideas. Mediation results indicate that this effect operates primarily through greater effort during idea generation rather than shifts in group processes. Quantity incentives foster sustained and broader exploration of the idea space, thereby increasing the likelihood of producing more original and high-quality ideas. By contrast, Quality incentives tend to promote the intensive development of a limited set of ideas, improving quality but limiting exploratory breadth. Direct incentives for originality fail to increase originality, likely reflecting uncertainty about what constitutes originality. Originality emerges more reliably as a by-product of broad exploration than as a direct response to incentives targeting it. In contrast, quality responds positively to incentives targeting any creativity dimension, suggesting that it is a fundamental and consistently valued aspect across all treatment groups. Beyond incentives, greater effort in idea generation increases quantity and originality, while intensive evaluation (measured by suggestions per minute) enhances quality, underscoring inherent trade-offs in collaborative creative work.

Evaluation and selection introduce a key bottleneck in creative group processes. Groups systematically discard highly original ideas in revision and selection. The persistence of this pattern across all incentive conditions suggests that motivational alignment alone cannot overcome structural biases against unconventional solutions. Even under the explicit Originality incentive, groups do not retain the most original ideas, either because they are unable to recognize the most original ideas or because implicit bias against originality discourages selecting them. This decline in originality from generation to refinement and selection highlights a central challenge for creativity in groups: incentives can increase the supply of original ideas, but they do not ensure that such ideas are retained and selected.

With these findings, this paper contributes to three strands of literature. First, it builds on research on group-level incentives in interdependent production and work processes (e.g., [Friebel et al., 2017](#); [Englmaier et al., 2023, 2024](#); [Frederiksen et al., 2024](#); [Goette and Senn, 2024](#)) by examining whether incentives in creative production—which likewise depends on interdependent group processes—also affect not just effort but also coordination, leadership, and task allocation. Second, the paper contributes to the literature on incentive design for creative tasks. Two studies are particularly relevant: [Charness and Grieco \(2023\)](#) examine group-level relative performance pay in a setting where creative tasks are performed individually but embedded in a group context. While group members may support one another and reallocate tasks, they do not jointly solve a single task and consider multiple dimensions of creativity, a gap that this paper addresses. They find that group-level incentives and ranking against out-group members increase individual creativity by fostering stronger pro-sociality among group members and establishing a social norm of high effort. While I do not find that

incentives operate through increased pro-social attitudes, groups with higher mean pro-social attitudes exhibit better performance in my setting. Consistent with [Laske et al. \(2024\)](#), who show that individual piece-rate incentives increase innovative idea production, I demonstrate that group-level relative performance incentives also stimulate exploration, thereby indirectly enhancing creative performance in a task that encompasses both idea generation and evaluation/selection processes. Third, the paper adds to psychological and organizational research on group processes in creative tasks by providing evidence on the persistence of bias against originality in incentivized group settings during the understudied phases of idea evaluation and selection (e.g., [Mueller et al., 2011](#); [Rietzschel et al., 2019](#)). Overall, this paper advances our understanding of how group-level relative performance pay shapes multiple dimensions of creativity across the full creative process. The findings offer direct implications for organizations and policymakers seeking to optimize group processes and design effective incentive structures for creative tasks.

The remainder of the paper is structured as follows: Section 2 reviews the related literature on incentives, group performance, and creativity. Section 3 details the experimental design, including the novel task and the incentive conditions. Section 4 describes the data, measurement, and empirical strategy. Section 5 presents the main results regarding the effects of incentives on the three creativity dimensions and overall innovation. Section 6 investigates the heterogeneity and mechanisms underlying these effects. Section 7 summarizes and discusses all findings.

2 Related Literature

This paper relates to three literature strands: (i) the effectiveness of group-level incentives in interdependent production processes, (ii) optimal incentive design for creative tasks, and (iii) psychological and organizational research on creative group processes.

2.1 Group-level Incentives in Interdependent Production

A large literature shows that monetary incentives can substantially increase individual productivity in routine tasks (e.g., [Lazear, 2000](#)). Extending this insight to groups introduces additional complexity: when performance is observable only at the group level, group-based incentives can create both peer pressure and free-riding ([Kandel and Lazear, 1992](#)). Nonetheless, field and laboratory evidence demonstrates that group-based piece-rate incentives can be effective in interdependent production and work processes ([Friebel et al., 2017](#); [Frederiksen et al., 2024](#); [Goette and Senn, 2024](#)). Recent work by [Englmaier et al. \(2023, 2024\)](#) provides particularly nuanced insights. In a series of large-scale field experiments in escape rooms, they find that tournament incentives substantially improve team performance in non-routine analytical tasks, driven by social image concerns from public rankings, motivating the top performers, and by monetary prizes, which boost outcomes across the entire performance

distribution. Incentives also reshape team organization by increasing the demand for leadership, which a follow-up experiment confirms as a causal driver of performance improvements (Englmaier et al., 2025).

Across these studies, motivational crowding out appears minimal (cf. Deci et al., 1999). This literature strand shows that group-level incentives can enhance performance not only in routine tasks but also in interdependent and non-routine tasks, affecting not just effort but also coordination, leadership, and task allocation. This paper investigates whether similar motivational and organizational responses arise in creative production when teams jointly generate and develop ideas.

2.2 Creativity and Incentives

The literature on creativity and incentives demonstrates that the effects of financial rewards are highly context-dependent. Experimental evidence is mixed, reflecting differences in payment schemes, reward magnitude, task type, group interaction, and creativity assessment (Attanasi et al., 2021).

Incentives tend to boost creativity in closed or constrained tasks. Charness and Grieco (2019) distinguish between closed and open creative tasks, finding that monetary rewards enhance performance in constrained tasks but are less effective for open-ended exploration. Bradler et al. (2019) document large positive effects of tournament incentives in the Alternate Uses Task (Guilford, 1967), which can be classified as a middle point between closed and open tasks, described as “open with constraints” (cf. Attanasi et al., 2021). The task in this paper falls into the same category.

In group settings, incentives and between-group competition can foster creativity. Charness and Grieco (2023) show that ranking against out-group members and group-level incentives positively influence individual creativity. They show that competing against an out-group fosters a stronger pro-social attitude among group members and establishes a social norm of high effort. This is consistent with Chen et al. (2012), who find that between-group tournaments increase cohesion and collaborative effort, leading to more creative group solutions compared to group piece-rate pay. Collaboration itself can enhance creativity: in the complex Candle Problem (Duncker, 1945), groups outperform individuals (Ramm et al., 2013), though this does not generalize to simpler tasks (Gneezy et al., 2022).

The effectiveness of incentives depends strongly on their design. Laske et al. (2024) introduce the word illustration task and investigate individual-level incentives tied to specific dimensions of creativity. They find that piece-rate incentives increase the number of innovative ideas (ideas that are both high in quality and original). Interestingly, unweighted incentives, granted for any idea, perform at least as well as weighted incentives that additionally depend on the quality and/or originality of ideas. Kachelmeier et al. (2008), working with small samples, find that combining quantity and creativity incentives reduces total productivity without increasing creativity relative to quantity incentives alone. In a field experiment

at a large technology company, where employees are encouraged to submit ideas on process and product improvements, offering rewards increased idea quality and overall participation but reduced ideas per participant, leaving total idea quantity unchanged (Gibbs et al., 2017). Blaufus et al. (2025) find that both input and output incentives raise the number of innovative ideas, but output incentives lead to less complex ideas. More nuanced schemes that allow early failure while rewarding long-term achievement encourage exploration vis-à-vis exploitation and increase, in turn, the probability of highly creative ideas (Azoulay et al., 2011; Ederer and Manso, 2013).

By contrast, several studies find null or even negative effects of incentives on creativity. Kleine (2021) documents negative incentive effects on individual performance in the “Candle Problem”, regardless of gain or loss framing. Eckartz et al. (2012) observe no significant impact of flat, linear, or tournament pay in a word-based creativity task, and conclude that individual skills are the primary determinant of creative performance. Similarly, Erat and Gneezy (2016) find that neither piece-rate nor tournament incentives enhance creativity, even though they increase effort. Competitive incentives can even have a detrimental effect, an outcome largely driven by the women in their sample, consistent with the “choking under pressure” hypothesis of Ariely et al. (2009), who show that very high stakes reduce performance. Based on relatively small samples, Kachelmeier et al. (2020) replicate the negative effects in single-solution tasks, but show that incentives can foster creativity once the same problems allow for multiple solutions. Competition intensity is also an important moderator: Gross (2020) provides evidence from a field experiment showing that the most original creative output occurs under moderate competition, compared to either no competition or intense competition. Similarly, Baer et al. (2010) show that creativity increases from low to intermediate levels of intergroup competition but remains flat for intermediate to high levels, with these effects being mediated by within-group collaboration.

Taken together, three key moderators explain the heterogeneity of effects: (i) task characteristics, with closed, multi-solution, or complex tasks benefiting more from incentives, while open-ended or single-solution tasks are more pressure-sensitive; (ii) competition intensity, which follows a pattern whereby moderate competition increases creativity more than low-intensity competition, while high-intensity competition produces mixed results; and (iii) group interaction, where groups can outperform individuals when incentives foster cooperation. However, these studies primarily focus on the incentivization of individual creativity and leave downstream stages of idea evaluation and selection largely unexplored.

2.3 Creative Group Processes

A large body of research highlights that group creativity depends on social and cognitive processes rather than the sum of individual abilities. Two well-documented mechanisms explain productivity losses in brainstorming: production blocking, when individuals must take turns sharing their ideas, which disrupts cognitive flow and reduces both the organization and

flexibility of idea generation (Nijstad et al., 2003), and evaluation apprehension, the fear of being negatively judged by others, which can inhibit the expression of novel or unconventional ideas. Together, these factors account for the frequently observed performance gap between interactive groups and nominal groups (groups of individuals working independently), as formalized in the social and cognitive influence model of group brainstorming (Paulus et al., 2002). Building on these insights, Korde and Paulus (2017) show that alternating individual and group sessions generate more ideas than either mode alone. To mitigate production blocking and evaluation apprehension, I therefore introduce alternating phases of individual and group work in the novel task.

Rietzschel et al. (2019) emphasize that group creativity extends far beyond the mere generation of ideas. Creative ideas must be recognized, valued, and finally selected to have an impact. However, two critical stages in the creative process, specifically idea evaluation and selection, remain understudied. A central obstacle in effective evaluation and selection is the tension between originality and feasibility: highly original ideas are often disliked or rejected because they are perceived as risky. Mueller et al. (2011) demonstrate that even when individuals explicitly claim to value creativity, they display an implicit bias against originality that intensifies under conditions of uncertainty and hampers their ability to identify creative ideas. This tendency may be exacerbated in group settings, where individual-level biases are compounded. This paper provides evidence on the prevalence of the bias against originality in incentivized group settings. Epistemic and pro-social motivation further shape group creativity: De Dreu et al. (2011) find that higher epistemic motivation (the extent to which group members systematically process and exchange information) increases originality and idea quantity when combined with pro-social motivation (the tendency to pursue collective rather than individual gain), which can be fostered through group-level incentives.

This paper contributes to all three literature strands. First, the paper introduces a novel task and provides the first evidence on how group-level financial incentives affect multiple dimensions of creativity in a controlled laboratory setting. Second, by recording the entire creative process, it identifies how incentives influence idea generation, elaboration, and selection, filling a gap in the literature on evaluation and selection. Third, by comparing dimension-specific group incentives, the paper reveals trade-offs and complementarities between creativity dimensions. This approach bridges the literature on incentives, group dynamics, and creativity and generates insights for organizations seeking to foster innovative teamwork.

3 Experimental Design

3.1 Procedure

The experiment was conducted between January and June 2025 at the Mainz Behavioral and Experimental Laboratory (MABELLA) and the Frankfurt Laboratory for Experimental Eco-

conomic Research (FLEX). Participants, recruited via an open call through ORSEE (Greiner, 2015), were randomly assigned to one of four experimental conditions using cluster randomization stratified by treatment, with the session as the unit of randomization. The design was between-subjects, with all participants in a given session exposed to the same treatment condition. They were unaware in advance that the study focused on creativity. The experiment received ethics approval from the Joint Ethics Committee in Economics at Johannes Gutenberg University Mainz and Goethe University Frankfurt. I closely follow the pre-registered experimental design and analysis plan (AsPredicted #208,106).

I collected data from 640 participants: 480 ideators organized into 160 groups of three and 160 quality assessors. A total of 32 sessions were conducted (eight per treatment), each with 20 participants and lasting around 75 minutes. To ensure exactly 20 participants per session, 24 sessions took place at MABELLA and eight at FLEX; sessions are balanced across the different locations with respect to treatment assignment. Within each session, five groups of three participants are randomly formed as ideators, and the remaining five participants serve as quality assessors. The experiment was programmed and run on individual computers using oTree in a controlled laboratory environment (Chen et al., 2016). Participants interacted anonymously, with no verbal communication allowed. They received a 5 EUR participation fee and treatment-dependent additional payments, resulting in average earnings of 13.50 EUR (based on the hourly wage of student assistants without a degree).

After the main task (see section 3.2), the ideators play one round of a dictator game to measure pro-social behavior. Each decides how much of a 10 EUR endowment to transfer to a recipient, who was randomly described as either a group member or a member of a competing group, to identify whether pro-social behavior works through group identity mechanisms. Fifteen allocation decisions are collected per session, and one was randomly selected and paid out to the allocator and a randomly chosen recipient according to the chosen distribution (cf. Charness et al., 2016). Ideators then complete two questionnaires: one concerning the main task (Table A1) and another collecting demographics and general attitudes toward groups and creativity (Table A2). Finally, they perform the Alternate Uses Task (AUT) to assess their divergent thinking abilities (Guilford, 1967). In this task, participants think of as many alternative uses as possible for two everyday objects. Quality assessors perform the Compound Remote Associates Task (CRAT) to measure their convergent thinking abilities (Mednick, 1962). In this task, they find a fourth word that forms a meaningful combination with three seemingly unrelated words. Each assessor solves five problems of increasing difficulty and then completes two questionnaires analogous to the ideators (Tables A3, A4).

3.2 Task

The novel task measures performance in three core dimensions of creativity—quantity, quality, and originality. Participants collaborate to generate as many clue pairs as possible that are both original and enable a quality assessor to identify a target word. Each group completes six

rounds, one per target word¹. At the beginning of each round, ideators are shown the target word and three taboo words; these taboo words are among the most frequent associations for a given target word in the publicly available dataset of the online game *Just One* (collected from [Board Game Arena](#)) and are not permitted to appear in a clue pair². Each clue pair consists of exactly two single words. The translated instructions are provided in *Appendix C: Instructions*. The task proceeds in four phases:

1. **Individual ideation:** Participants independently generate candidate clue pairs without communication.
2. **Group evaluation:** Participants share and discuss their generated clues, providing feedback and ratings on them.
3. **Individual revision:** Participants revise their clue pairs based on the received feedback.
4. **Group selection:** The group coordinates via chat and selects the final clue pair through a voting, which is used for originality and quality assessment.

This structure combines divergent and convergent thinking, enabling broad idea generation followed by focused selection of the most creative ideas. After submission, a quality assessor receives the final clue pair and has up to three attempts to identify the target word. After each round, groups are informed of their performance across the three creativity dimensions, while between-group rankings are revealed only at the end. A second assessor reviews all pairs to verify that each can be meaningfully connected to the target word (which was the case for all pairs), excluding invalid pairs. Additional automated validity checks are implemented³. Quality assessors rotate between groups across rounds.

3.3 Treatment Conditions

The experimental manipulation varies only in the incentives provided to the ideators and the display of a session-level group ranking at the end. All other elements of the experiment

¹ After the first session (control condition), one round was eliminated to reduce overall session length (yielding six rounds instead of seven).

² The target and taboo words are: **wind:** *blow, breeze, squall*; **meal:** *food, brunch, breakfast*; **grain:** *wheat, barley, sand*; **science:** *physics, biology, chemistry*; **emergency:** *ambulance, urgent, alarm*; and **farm:** *agriculture, tractor, harvest* (translated from German). The online dataset was collected between September 2023 and November 2024, and comprises 25,664 rounds of *Just One*, recording group number, round number, submitted clues, guesses, target word, game mode, speed, and language (German, English, French). The six target words are selected from a pool of 794 words based on the following criteria: each word was played more than 25 times (>75th percentile), received over 125 clues (>75th percentile), had an average originality between 54.33 (5th percentile) and 82.33 (95th percentile), exhibited a standard deviation of originality above 14.6 (50th percentile), and was correctly guessed in fewer than 75% of rounds.

³ Pairs are automatically classified as invalid if they contain the target word itself, words sharing the same linguistic root, direct translations or alternate spellings, made-up words, or items with special characters.

are identical across treatments. Hence, the design isolates the causal effect of rewarding a particular creative dimension. All participants are informed beforehand about the three dimensions of creativity and how they are assessed. The four experimental conditions are as follows:

1. **Control Condition:** Groups receive a flat payment of 13 EUR per participant independently of their performance in the task.
2. **Relative Performance Pay 1 – Originality Incentive:** Payments depend on the originality of the final clue pair selected by each group. Originality is measured as the semantic distance to the target word and displayed on a 1–10 scale.
3. **Relative Performance Pay 2 – Quality Incentive:** Payments are based on the quality of the submitted clue pair, measured by whether a quality assessor correctly identifies the target word:
 - First try: 3 points
 - Second try: 2 points
 - Third try: 1 point
4. **Relative Performance Pay 3 – Quantity Incentive:** Payments depend on the number of valid clue pairs generated by the group, with one point awarded per valid pair each round.

In all three relative performance treatments, groups are ranked from first to fifth based on the sum of points earned across the six rounds. Payoffs comprise the participation fee of 5 EUR plus a bonus of up to 15 EUR awarded according to the final ranking: first place receives a 15 EUR bonus (total payoff 20 EUR), second 10 EUR (total 15 EUR), third 7 EUR (total 12 EUR), fourth 5 EUR (total 10 EUR), and fifth 3 EUR (total 8 EUR)⁴. The payment scheme is identical in each treatment. Payments are determined at the group level, so all members of a group receive the same bonus.

4 Data and Estimation Strategy

I collected data from 640 participants, comprising 480 ideators and 160 quality assessors. This results in 2,880 individual-level observations, corresponding to 480 ideators completing six rounds each, and 960 group-level observations, corresponding to 160 groups completing

⁴ In case of a tie, the groups sharing the same position receive the average of the respective bonuses. For example, if two groups tie for second, each receives the mean of the second- and third-place bonuses, i.e., 8.50 EUR per participant.

six rounds each⁵. Each treatment condition includes 40 groups (120 participants) and 40 quality assessors. Table A5 in the Appendix presents descriptive statistics of the ideators and balance checks across the four treatment conditions. Randomization was successful, with no statistically significant differences for most variables. Table A6 reports additional descriptive statistics for the sample of quality assessors.

I measure four primary outcome variables—originality, quantity, quality, and innovation—across six rounds, each corresponding to a specific target word. Originality and quantity are measured at both the individual and group level, before and after feedback and revision, whereas quality and innovation are measured only at the group level for the final clue pair.

Originality is measured by the semantic distance between a clue pair and the target word. This approach builds on a growing body of research showing that semantic distance provides a reliable, valid, and efficient proxy for human judgments of creative originality (Beaty and Johnson, 2021; Olson et al., 2021; Beaty et al., 2023). I compute the semantic distance using pre-trained fastText word vectors (Grave et al., 2018), which represent words as high-dimensional embeddings based on their contextual co-occurrence patterns across vast multilingual text corpora (Wikipedia and Common Crawl)⁶. Such embeddings capture fine-grained semantic relationships, allowing for precise quantification of conceptual similarity and divergence. I calculate the cosine distance between the vectors of each word in the clue pair and the target word, average the two distances, and multiply by 100 to facilitate interpretability. A score of 0 indicates no distance (identical words), while the theoretical maximum of 200 occurs when the words are maximally different. Originality scores in this study typically range between 40 and 80, with values below 50 indicating relatively conventional associations and those above 70 indicating semantically remote and novel associations. Compared to the divergent association task (DAT) (Olson et al., 2021), the scores are slightly lower. This difference likely reflects that all ideas are generated with respect to a target word, whereas the DAT allows participants to generate any unrelated words.

At the individual level, average originality is calculated as the mean semantic distance across all clue pairs generated by a participant for a given target word. At the group level, average originality is defined as the mean of the individual mean originality scores of the three group members. The pre-feedback average originality in this dataset is 61.81 (SD = 5.65, Min = 39.29, Max = 81.35), while the post-feedback average originality is 60.83 (SD = 6.05, Min

⁵ In one session (quality condition), a brief network interruption temporarily affected the quality assessors, resulting in two groups having limited quality measures for one round (2/960 observations).

⁶ In the pre-registration, I had indicated that I would also use the German GloVe pre-trained word vectors from deepset. However, these embeddings are outdated and no longer available online. An important advantage of using fastText instead is that it can generate vectors for unknown words, as it is based on subword components. This means that fastText will produce a vector for every word, ensuring that even rare or original terms are not excluded. In addition, German GloVe embeddings are trained using only lowercase text, which results in the loss of distinction between nouns and verbs in German.

= 39.29, Max = 81.35). The final originality of the group’s selected clue pair has a mean of 59.39 (SD = 9.88, Min = 30.50, Max = 88.57).

The semantic distance measure is theoretically grounded in associative and network-based models of creativity, which posit that creative individuals can flexibly traverse distant regions of semantic space (Beaty et al., 2014; Olson et al., 2021; Baader et al., 2025). Higher semantic distance reflects the ability to connect concepts that are not typically co-activated in semantic memory. This view aligns with recent evidence showing that semantic knowledge plays a central role in guiding exploratory thought and enabling cumulative cultural innovation (Yaman et al., 2025). Large-scale validations show that semantic distance measures not only correlate positively with human-rated creativity but also exhibit meaningful associations with broader indicators of creative potential, including openness to experience, creative achievement, and self-perceived creativity (Beaty et al., 2023). Furthermore, semantic distance metrics retain validity across linguistic and cultural boundaries (Beaty et al., 2023). The fastText models are trained for 157 languages and validated for ten languages, including German (Grave et al., 2018). Thus, the operationalization of originality via semantic distance represents a theoretically informed, empirically validated, and cross-linguistically generalizable approach that advances beyond traditional rater-based methods.

Quality captures how effectively clue pairs enable a quality assessor to recognize the target word. Recognizability reflects quality and usefulness; clue pairs are more valuable when they make the target words easier to identify. This approach builds on the concept of quality measurement in creative tasks, as established in the word illustration task (Laske et al., 2024). I score the total number of attempts a quality assessor requires to correctly guess the target word using a group’s final clue pair: 3 points for a correct first attempt, 2 points for a correct second attempt, 1 point for a correct third attempt, and 0 points if unrecognized. Quality has a mean of 1.07 (SD = 1.30, Min = 0, Max = 3). Quality assessors are incentivized under the same payment scheme as the ideators. Each quality assessor accumulates points over six rounds, which determine their relative performance ranking among the five assessors within a session. Accordingly, their total payoff consists of a 5 EUR participation fee plus a performance-based bonus of up to 15 EUR.

Quantity is defined as the total number of clue pairs generated per target word. At the individual level, this reflects the sum of clue pairs contributed by a participant in a round; at the group level, it reflects the sum contributed by all group members. Before feedback and revision, the mean quantity is 5.88 (SD = 2.83, Min = 1, Max = 29), and after feedback and revision, it is 5.61 (SD = 2.78, Min = 1, Max = 28).

Innovation is measured both as a binary indicator and a standardized average index. In line with the definition of creativity (Amabile et al., 1996), a clue pair is innovative if it is both original and of high quality (useful). The binary innovation measure equals 1 if a final clue pair is both highly original (\geq 75th percentile) and achieves a quality score \geq 1 (i.e., is recognizable). Robustness checks use thresholds at the 80th, 85th, and 90th percentiles. The binary Innovation 75 measure has a mean of 0.08 (SD = 0.27, Min = 0, Max = 1). The

Innovation Index is the standardized mean of final originality and quality, with a mean of 0 (SD = 1, Min = -2.85, Max = 3.17). Descriptive statistics for the primary outcomes are shown in Table A7 (by treatment) and Table A8 (by round) in the Appendix.

Figure A1 and Table A9 in the Appendix report pairwise correlations among the primary performance measures. Quality is weakly and negatively correlated with quantity ($r = -0.057$, $p < 0.1$), suggesting a modest trade-off between the amount of output and its assessed quality. The negative relationship between quality and average originality is more pronounced ($r = -0.161$, $p < 0.01$), indicating a clearer trade-off whereby more original contributions tend to receive lower quality evaluations. In contrast, quantity and originality are positively associated ($r = 0.192$, $p < 0.01$), implying that groups generating a higher volume of output also tend to exhibit greater originality.

In addition to the primary outcomes, I assess several secondary outcomes to better understand the creative process. First, I measure groups' self-assessment of their clue pairs with respect to quality and originality on a 0-10 scale, capturing confidence in their submissions. Quality assessors also evaluate the clue pairs in terms of quality, originality, and overall creativity on a 0-10 scale. Assessors always evaluate the clue pairs of a different group than the one whose pairs they previously used to guess the target word. I further measure semantic exploration using two computational indicators. Flexibility captures the semantic variation between all ideas produced by a participant in a single round and is defined as the mean pairwise cosine distance between all ideas. Idea divergence measures the spread of words around their centroid, calculated as the mean cosine distance between each word and the average vector of all words produced by the participant in that round. Higher values of flexibility or divergence indicate greater semantic variation and broader coverage of the idea space.

I also collect a wide range of additional variables as potential moderators or mediators. These include demographic characteristics (age, gender, field of study, and native speaker status), individual divergent thinking ability for ideators (measured by the AUT), and individual convergent thinking ability for quality assessors (measured by the CRAT). Further measures include creative self-efficacy and identity, creative style, effort (based on page times) and motivation, teamwork (feedback exchange and survey measures), risk tolerance, coordination (voting behavior, chat data, and survey measures), and pro-sociality (dictator game and survey measures). A detailed description of all moderator and mechanism variables is provided in *Appendix D: Detailed Variable Descriptions*.

Empirical Strategy: I analyze the effect of treatment conditions on creative output using regression models. For each outcome $y_i \in \{Quality, Quantity, Originality, Innovation\}$, I estimate:

$$y_i = \beta_0 + \beta_1 \text{OriginalityPRP}_i + \beta_2 \text{QualityPRP}_i + \beta_3 \text{QuantityPRP}_i + \gamma \mathbf{X}'_i + \varepsilon_i$$

where OriginalityPRP, QualityPRP, and QuantityPRP indicate whether group i is assigned to a performance-related pay (PRP) condition. \mathbf{X}'_i includes all control variables: demographics

(age, gender, field of study, native speaker), risk tolerance, pro-sociality, AUT performance, survey-based measures of creative self-assessment, identity, and style, as well as group work attitudes, conformity, and competitiveness. Depending on the outcome, I use OLS, logit, probit, and ordered logit specifications. In some specifications, I also include round fixed effects and control for other creativity dimensions. Beyond group-level analyses, I investigate individual-level outcomes, specifically average originality and quantity, comparing each Treatment group to the Control group. Standard errors are clustered at the group level for group-level outcomes and at the individual level for individual-level outcomes.

5 Main Results

The results are presented in three parts: first, I examine the effects of the incentive treatments on the three creativity dimensions—quantity, quality, and originality; second, I assess how these incentives shape overall group innovation performance; and third, I analyze the evolution of originality throughout the creative process, from initial idea generation to final selection.

5.1 Effect of Incentives on Creativity Dimensions

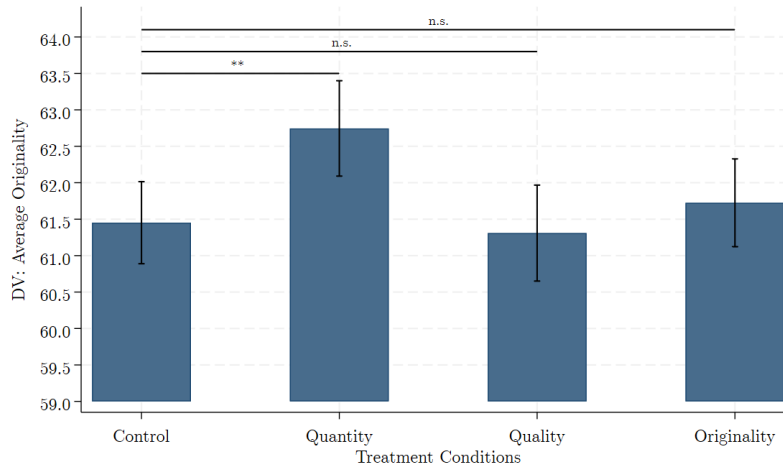
I begin by evaluating how the various incentive schemes impact the quantity, quality, and originality of ideas produced at the group level. Figure 1 reports mean values by treatment condition (Control, Quantity, Quality, Originality), with 90% confidence intervals and pairwise comparison p-values relative to the Control group. Tables 1, 2, and 3 present the corresponding OLS and ordered logit regression results.

Starting with the **originality dimension** (Figure 1a), I first assess whether directly incentivizing originality enhances group performance, in line with my hypothesis that incentives increase effort and thereby performance in the incentivized dimension. Surprisingly, the Originality incentive does not significantly enhance average originality relative to the Control group (see Table 1). Effort measures reported in Table A10 in the Appendix also show no significant differences between the Originality incentive and the Control group. So, the incentive even fails to increase effort. Several factors may explain why the Originality incentive does not increase originality: first, generating original ideas is strongly intrinsically motivated, and financial incentives may crowd out that motivation (cf. [Deci et al., 1999](#)). Second, although the measurement of originality is known to all participants, the underlying concept remains abstract. Uncertainty about what constitutes originality may limit the effectiveness of direct incentives. The production function linking effort to outcomes is opaque. This interpretation aligns with the mechanisms behind the effectiveness of the Quantity incentive, where effort is more directly mapped into output, and perceived self-efficacy plays a pronounced role (see section 6.2).

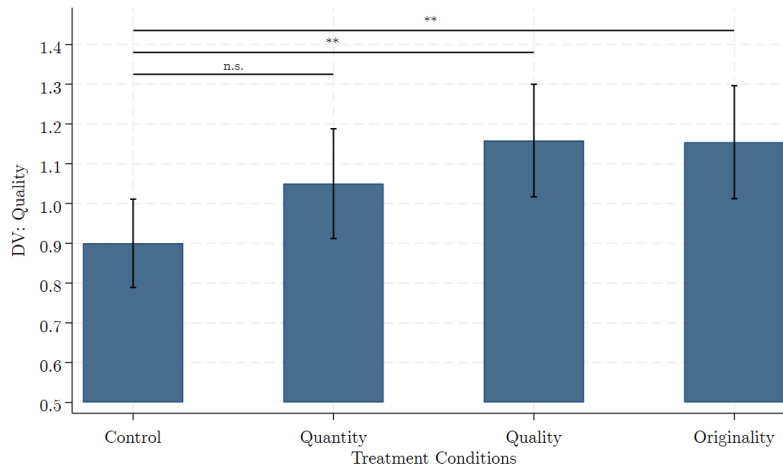
Next, I examine potential spillover effects of other incentive types on originality. Groups in the Quantity incentive treatment achieve significantly higher average originality scores than

Figure 1: Creativity Dimensions

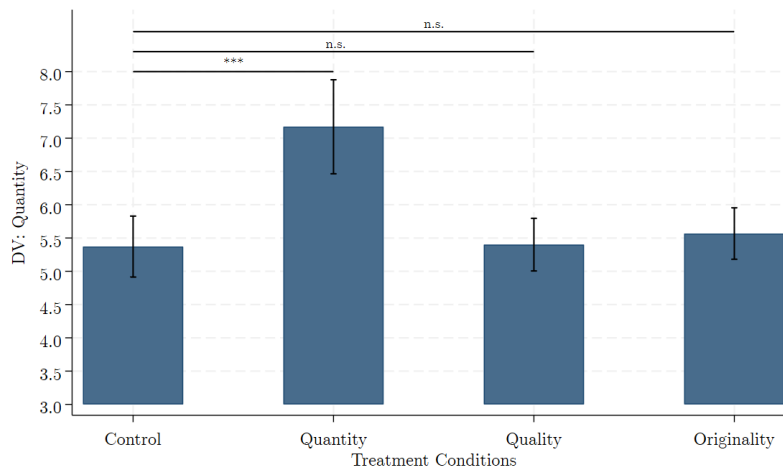
(a) Average Originality



(b) Quality



(c) Quantity



Notes: Significance stars are based on pairwise t-tests (Quantity, Originality) as well as Mann-Whitney U tests (Quality) comparing the Control group to the respective Treatment groups. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Error bars indicate 90% confidence intervals.

Table 1: Group Originality Average

	before feedback and revision				after feedback and revision			
	(1) Originality \emptyset	(2) Originality \emptyset	(3) Originality \emptyset	(4) Originality \emptyset	(5) Originality \emptyset FB	(6) Originality \emptyset FB	(7) Originality \emptyset FB	(8) Originality \emptyset FB
Incentive Originality	0.273 (0.497)	0.221 (0.435)	0.221 (0.436)	0.258 (0.429)	0.396 (0.522)	0.218 (0.442)	0.218 (0.443)	0.246 (0.416)
Incentive Quality	-0.143 (0.522)	0.155 (0.501)	0.155 (0.502)	0.253 (0.459)	-0.289 (0.579)	-0.058 (0.556)	-0.058 (0.558)	0.072 (0.517)
Incentive Quantity	1.293** (0.520)	1.572*** (0.503)	1.572*** (0.505)	0.945** (0.475)	1.323** (0.540)	1.556*** (0.540)	1.556*** (0.541)	0.815 (0.495)
Quality				-0.408 (0.300)				-0.205 (0.332)
Quantity				0.349*** (0.087)				
Quality \times Quantity				-0.006 (0.043)				
Quantity FB								0.462*** (0.088)
Quality \times Quantity FB								-0.045 (0.047)
Constant	61.452*** (0.340)	62.712*** (3.215)	60.240*** (3.243)	61.029*** (3.027)	60.475*** (0.390)	61.178*** (3.368)	58.842*** (3.362)	59.867*** (3.184)
Group Controls		✓	✓	✓		✓	✓	✓
Round FE								
Obs.	960	960	960	960	960	960	960	960
R2	0.010	0.044	0.366	0.398	0.010	0.039	0.355	0.393

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

those in the control condition ($p < 0.05$). Table 1 confirms this pattern: the Quantity incentive effect remains robust after including controls and round fixed effects (columns 2-3, 1.572, $p < 0.01$) and is equivalent to a Cohen’s d of 0.29 that can be interpreted as a small to medium effect. Columns 1–4 refer to the pre-feedback phase, and columns 5–8 to the post-feedback phase. After feedback and revision, the Quantity incentive effect continues to be strong (columns 6–7, 1.556, $p < 0.01$), suggesting that receiving feedback and making revisions does not diminish its positive impact on originality. The quantile regressions presented in Table A11 in the Appendix demonstrate that the Quantity incentive consistently exerts a significant positive effect across the 50th, 75th, and 90th quantiles. These results indicate that quantity-based incentives indirectly foster originality, potentially by encouraging a broader exploration of the idea space. Importantly, when controlling for the number of ideas generated (columns 4 and 8), part of the originality boost from the Quantity incentive appears mediated by increased idea production. In contrast, the Quality incentive does not generate significant spillover effects on originality, indicating that rewarding quality does not substantially affect performance in this non-incentivized dimension.

Table A12 in the Appendix reports results for the originality of the final clue pair selected by each group. Although no treatment effects reach statistical significance, the coefficients for the Quantity incentive remain positive and in the expected direction, suggesting that groups rewarded for idea quantity also tended to choose more original final outputs. The lack of significance is likely due to higher variance and lower power when analyzing only the single

final clue pair per group, compared to the average originality across all generated clue pairs. Complementary evidence using the semantic distance to the taboo words as an alternative measure for originality is provided in Table A13. Here, the Quantity incentive also shows positive and significant effects on average originality ($p < 0.01$), reinforcing the robustness of the previous results.

Table 2: Group Quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quality	Quality	Quality	Quality	Quality	Quality	Quality
	OLS	OLS	OLS	OLS	OLogit	OLogit	OLogit	OLogit
Incentive Originality	0.254** (0.109)	0.261** (0.110)	0.261** (0.111)	0.273** (0.108)	0.361** (0.164)	0.386** (0.167)	0.395** (0.179)	0.411** (0.177)
Incentive Quality	0.258** (0.108)	0.360*** (0.111)	0.360*** (0.111)	0.365*** (0.110)	0.388** (0.161)	0.552*** (0.168)	0.552*** (0.179)	0.561*** (0.179)
Incentive Quantity	0.150 (0.107)	0.197* (0.102)	0.197* (0.102)	0.279*** (0.102)	0.218 (0.162)	0.302* (0.157)	0.331* (0.170)	0.443** (0.175)
Originality				-0.029*** (0.009)				-0.050*** (0.016)
Quantity FB				-0.131 (0.082)				-0.220 (0.137)
Originality \times Quantity FB				0.002 (0.001)				0.003 (0.002)
Constant	0.900*** (0.067)	-0.381 (0.636)	0.216 (0.640)	1.707** (0.831)				
Group Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.007	0.028	0.102	0.121				
Pseudo R2					0.003	0.012	0.047	0.056

Note: The table reports coefficients of OLS and ordered logit regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Ordered logit coefficients are reported as logits, not as odds ratios or marginal effects. Cutpoints (thresholds) exist in the model but are omitted from this table for clarity. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Turning to the **quality dimension** (Figure 1b), I first examine the effect of directly incentivizing quality. Groups in the Quality incentive treatment exhibit significantly higher final idea quality than those in the Control group ($p < 0.05$). Table 2 shows that the Quality incentive increases quality by 0.258–0.365 points ($p < 0.01$) across OLS (columns 1–4) and ordered logit specifications (columns 5–8) with group controls and round fixed effects. This effect corresponds to a Cohen’s d of 0.20–0.28, which can be interpreted as a small to medium effect size. These results indicate that directing incentives toward qualitative aspects of creativity improves performance along that dimension.

I further explore how incentives aimed at different dimensions affect quality. Interestingly, the Originality incentive also increases quality, with gains of 0.254–0.273 points ($p < 0.05$). Although quality is not directly incentivized, participants receive feedback on all performance dimensions after each round. This feedback may induce spillovers across dimensions, motivating participants to improve their ideas in the quality dimension even when monetary

incentives target originality. A similar, though weaker, pattern emerges for the Quantity incentive. While no significant effect on quality is detected in the baseline specification, modest positive effects appear once controlling for performance in the other creativity dimensions (columns 4 and 8). These results indicate that quality responds positively to incentives aimed at all creativity dimensions, in contrast to originality, which proves resistant to direct incentivization. Quality appears to be a fundamental consideration in all treatment groups, likely because all participants generate words in relation to the respective target word, ensuring a baseline level of quality⁷.

Table 3: Group Quantity

	before feedback and revision				after feedback and revision			
	(1) Quantity	(2) Quantity	(3) Quantity	(4) Quantity	(5) Quantity FB	(6) Quantity FB	(7) Quantity FB	(8) Quantity FB
Incentive Originality	0.196 (0.361)	0.226 (0.336)	0.226 (0.337)	0.222 (0.330)	0.183 (0.353)	0.210 (0.323)	0.210 (0.324)	0.210 (0.318)
Incentive Quality	0.029 (0.365)	0.178 (0.358)	0.178 (0.359)	0.172 (0.341)	-0.067 (0.345)	0.100 (0.342)	0.100 (0.343)	0.098 (0.327)
Incentive Quantity	1.800*** (0.508)	2.080*** (0.451)	2.080*** (0.453)	1.944*** (0.440)	1.775*** (0.496)	2.015*** (0.438)	2.015*** (0.439)	1.896*** (0.426)
Quality				-0.824 (0.681)				-0.842 (0.658)
Originality \emptyset				0.085*** (0.028)				0.075*** (0.027)
Quality \times Originality \emptyset				0.012 (0.011)				0.013 (0.011)
Constant	5.371*** (0.276)	-0.343 (2.348)	-2.027 (2.364)	-7.001*** (2.602)	5.137*** (0.271)	-0.711 (2.236)	-2.272 (2.250)	-6.652*** (2.486)
Group Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.070	0.156	0.293	0.319	0.074	0.155	0.277	0.300

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Finally, for the **quantity dimension** (Figure 1c), groups exposed to the Quantity incentive treatment produce substantially more ideas than those in the control condition ($p < 0.01$). Neither the Quality nor Originality treatments affect the number of ideas generated. Table 3 confirms that only incentives explicitly targeting output volume successfully increase idea production. This is valid both before (1.800–2.080, $p < 0.01$) and after feedback and revision (1.775–2.015, $p < 0.01$) and robust to the inclusion of group controls, round fixed effects, and controlling for other creativity dimensions. The effect is large and equals a Cohen’s d

⁷ Table A14 in the Appendix further examines the determinants of quality and shows specifically that the quality assessors’ perceived difficulty in forming associations from the clue pairs is negatively related to quality ($p < 0.01$). Table A15 corroborates previous findings by showing that the Quality incentive substantially lowers the perceived difficulty of choosing a final guess ($p < 0.01$), indicating that clue pairs generated under this condition lead more clearly to a specific target word and are therefore more useful.

of up to 0.68. Table A16 in the Appendix further corroborates these results, showing robust positive effects of the Quantity incentive across the entire quantity distribution. Moreover, there is a positive association between quantity and originality across pre- and post-feedback rounds (0.075–0.085, $p < 0.01$), supporting the idea that increased idea generation partially mediates the effect of Quantity incentives on originality. In the Appendix, Tables A17 and A18 report individual-level results on originality and idea quantity that are in line with the findings presented here.

Overall, these results align with theories of exploration–exploitation trade-offs in creativity. Quantity incentives appear to promote a more explorative search process, generating a broader range of ideas, which increases the chance of discovering highly original ideas (Ederer and Manso, 2013). In contrast, Quality incentives tend to promote the intensive development of a limited set of ideas, improving quality but limiting exploratory breadth. Originality incentives fail to stimulate exploration and to increase originality. The mediating role of quantity in translating Quantity incentives into originality gains highlights a potential key mechanism by which Quantity incentives foster overall innovation. Mechanisms are further examined in sections 5.3 and 6.2.

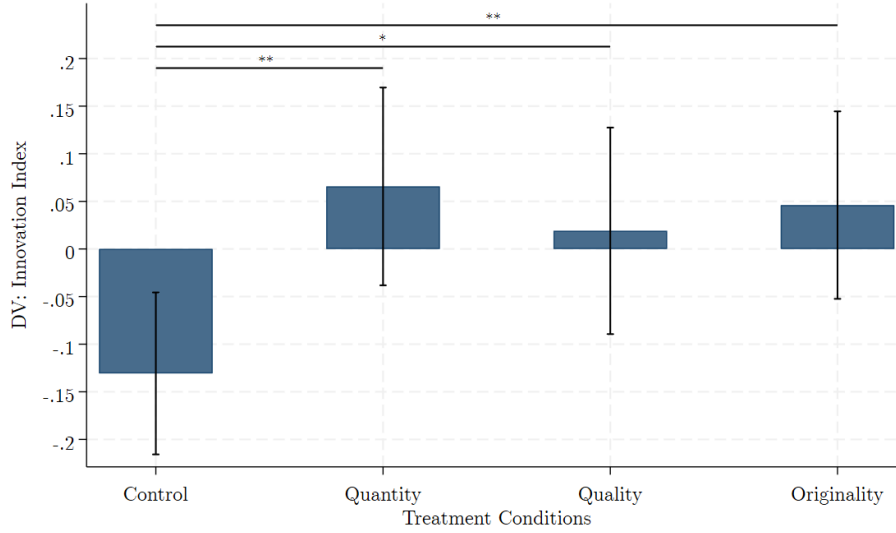
5.2 Effect of Incentives on Innovation

Next, I investigate how the different incentive treatments (Quantity, Quality, and Originality) affect overall group innovation performance. Innovation is captured both as a standardized continuous index—the Innovation Index (averaging standardized final originality and quality scores)—and as a binary outcome, indicating whether a group’s final clue pair simultaneously meets high originality (≥ 75 th percentile) and above-average quality (≥ 1) criteria. The latter measure is complemented by robustness checks using stricter thresholds (80th, 85th, 90th percentiles). Figure 2 depicts mean values of the standardized Innovation Index across treatment conditions, along with 90% confidence intervals and pairwise p-values from t-tests relative to the Control group. Tables 4 and 5 report corresponding OLS and limited dependent variable regressions using both the continuous Innovation Index and binary innovation measures.

Starting with Figure 2, I observe that all three incentive treatments lead to higher mean Innovation Index scores relative to the control condition. Consistent with the patterns observed for creativity dimensions, the Quantity treatment exhibits the largest positive effect (mean ≈ 0.06 , $p < 0.05$), followed by the Originality treatment ($p < 0.05$), whereas the Quality incentive yields a smaller and only marginally significant improvement ($p < 0.1$). Control group performance is notably lower, with a negative mean index value of around -0.1 . Taken together, these descriptive differences suggest that all incentive treatments enhance group innovation relative to flat pay, with quantity-based incentives producing the most robust and consistent effect.

Table 4 corroborates these descriptive patterns. Across specifications (columns 1–3), all three incentive treatments improve group-level innovation relative to flat pay. The Quantity

Figure 2: Innovation Index



Notes: Significance stars are based on pairwise t-tests comparing the Control group to the respective Treatment groups. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Error bars indicate 90% confidence intervals.

incentive displays the largest and most robust effect (0.218, $p < 0.01$), followed by the Quality (0.201, $p < 0.05$) and Originality (0.141, $p < 0.1$) treatments. The inclusion of group controls and round fixed effects increases explanatory power, yet the treatment effects remain stable in magnitude and significance. The quantile regressions reported in Table A19 in the Appendix provide an additional nuance: while all three incentive types exert positive effects throughout the distribution of the Innovation Index, the Quantity and Quality incentives become particularly pronounced at the upper quantiles (75th and 90th), suggesting that these treatments are especially effective at boosting the performance of the most innovative groups.

Table 4: Group Innovation Index

	(1)	(2)	(3)
	Inno Index	Inno Index	Inno Index
Incentive Originality	0.177** (0.078)	0.141* (0.077)	0.141* (0.078)
Incentive Quality	0.150* (0.083)	0.201** (0.085)	0.201** (0.085)
Incentive Quantity	0.196** (0.081)	0.218*** (0.075)	0.218*** (0.075)
Constant	-0.131** (0.051)	-1.479*** (0.481)	-1.232** (0.485)
Group Controls		✓	✓
Round FE			✓
Obs.	960	960	960
R2	0.006	0.032	0.164

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table 5 provides complementary evidence using binary innovation measures, coded as one if the group’s final clue pair meets both the originality and quality thresholds. Columns (1)–(3) use OLS specifications, while columns (4)–(8) employ logit and probit models and vary the originality thresholds from the 75th to the 90th percentile. Consistent with the index results, the Quantity incentive again shows the strongest and most consistent positive effects across all model specifications. For the baseline Innovation 75 threshold, quantity-based rewards increase the probability that groups achieve a high-innovation outcome by 7.4 percentage points ($p < 0.01$) compared to the Control group; a meaningful increase given the baseline probability of 4.6% (Cohen’s $d = 0.28$). This effect is robust to the inclusion of group controls and round effects, and remains statistically significant under both logit ($p < 0.01$) and probit ($p < 0.01$) specifications. The Quality incentive also yields a positive and marginally significant effect (around 4–5 percentage points, $p < 0.1$), whereas the Originality incentive fails to reach significance across all specifications, suggesting that directly rewarding originality does not reliably translate into higher joint originality–quality outcomes. Robustness checks using stricter originality thresholds (Innovation 80, 85, 90) show that the Quantity treatment effect gradually attenuates but remains significant, whereas other treatments become statistically insignificant. These patterns suggest that Quantity incentives not only increase the likelihood of achieving moderately high innovation final results but also push groups toward the top of the innovation distribution.

Table 5: Group Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Innovation 75	Innovation 75	Innovation 75	Innovation 75	Innovation 75	Innovation 80	Innovation 85	Innovation 90
	OLS	OLS	OLS	Logit	Probit	OLS	OLS	OLS
Incentive Originality	0.025 (0.024)	0.024 (0.027)	0.024 (0.027)	0.407 (0.465)	0.188 (0.208)	0.011 (0.018)	0.016 (0.016)	0.020 (0.016)
Incentive Quality	0.042* (0.023)	0.046* (0.025)	0.046* (0.025)	0.708* (0.418)	0.324* (0.190)	0.034* (0.019)	0.017 (0.018)	0.018 (0.017)
Incentive Quantity	0.067** (0.027)	0.074*** (0.028)	0.074*** (0.028)	1.079*** (0.418)	0.527*** (0.197)	0.065** (0.025)	0.046** (0.022)	0.036* (0.021)
Constant	0.046*** (0.013)	-0.010 (0.188)	-0.046 (0.189)	-3.947 (2.836)	-2.001 (1.356)	0.010 (0.145)	-0.110 (0.126)	-0.221* (0.121)
Group Controls		✓	✓	✓	✓	✓	✓	✓
Round FE			✓					
Obs.	960	960	960	960	960	960	960	960
R2	0.008	0.024	0.038	0.045	0.044	0.019	0.018	0.023
Pseudo R2								

Note: The table reports coefficients of OLS, logit, and probit regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Logit and probit coefficients are reported on the logit/probit scale, not as odds ratios or marginal effects. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

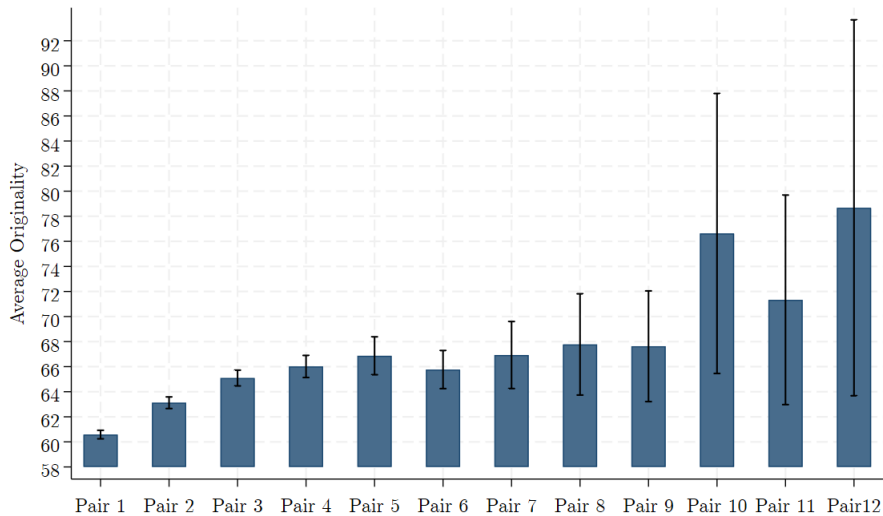
Taken together, the evidence from Figure 2 and Tables 4 and 5 points to a consistent pattern: Quantity-based incentives not only increase idea production (as shown earlier) but also create spillovers on both quality and originality, and enhance the aggregate innovativeness of final group outputs, as reflected in both continuous and binary innovation measures. Quality incentives appear to yield smaller but directionally similar gains, while directly rewarding originality fails to produce measurable improvements on the binary innovation indicator.

These findings suggest that encouraging the generation of a larger number of ideas fosters innovative solutions, potentially by broadening the exploration of the idea space and increasing the probability of identifying high-quality and original combinations⁸.

5.3 From Idea Generation to Selection: The Bias Against Originality

Finally, I examine how originality evolves throughout the group creative process, from initial idea generation to final idea selection. Figure 3 depicts the average originality scores across sequentially generated clue pairs (Pair 1 to Pair 12), capturing the dynamics of the idea generation process. Following this, Figure 4 and Table 6 shed light on how subsequent stages—feedback, revision, and selection—further influence average originality levels.

Figure 3: Originality and Quantity



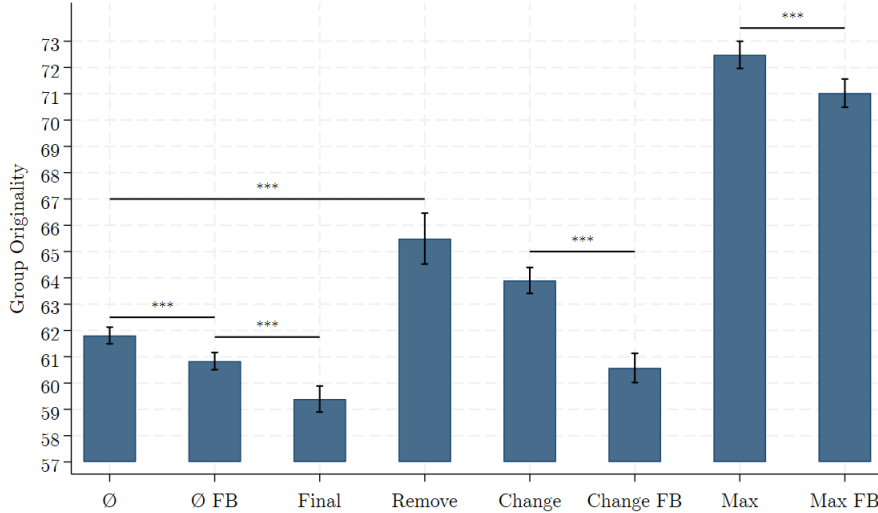
Note: The bar chart reports the mean originality per sequentially generated pair (Pair 1–12), averaged across rounds and across all ideators. Error bars indicate 90% confidence intervals.

Starting with idea generation, Figure 3 shows that as participants generate additional clue pairs for the same target word, the originality of their ideas steadily increases. Later clue pairs are, on average, substantially more original than earlier ones. This pattern is robust despite increasing variance in later stages (as indicated by widening confidence intervals), suggesting that continued idea production helps participants move beyond conventional solutions toward more novel and original ones. Table A21 in the Appendix shows that originality increases on average by 1.5 points with each additional clue pair. This aligns with the results in Table 1,

⁸ Table A20 in the Appendix reports pooled performance-pay results across all incentive treatments. Performance pay significantly increases quality, quantity, and both innovation measures ($p < 0.05$). While pooling masks the heterogeneous effects of individual incentive types, these results confirm that performance-based incentives overall enhance group innovation.

showing that quantity-based incentives indirectly foster originality by encouraging a broader exploration of the idea space.

Figure 4: Bias against Originality



Note: The bar chart reports the average originality of all clue pairs before feedback and revision (\emptyset), after feedback and revision (\emptyset FB), and of the group’s final selected clue pair (Final). In addition, it depicts the mean originality of clue pairs discarded during revision (Remove), of modified clue pairs before (Change) and after revision (Change FB), and of each group’s most original clue pair before (Max) and after feedback and revision (Max FB). Significance stars are based on pairwise t-tests comparing \emptyset and \emptyset FB, \emptyset FB and Final, \emptyset and Remove, Change and Change FB, and Max and Max FB. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Error bars indicate 90% confidence intervals.

Despite the generative benefits of quantity, Figure 4 reveals a marked and systematic decline in originality throughout the group process of idea evaluation, refinement, and selection. Specifically, the average originality of all clue pairs before (\emptyset) is higher than after feedback and revision (\emptyset FB), and originality falls even further for the group’s final selected clue pair (Final). This pattern of a steady reduction in originality indicates that original ideas are not only difficult to generate but also difficult to retain and select during the creative process. Further insights come from examining how groups handle ideas during the revision process. Clue pairs that are deleted (Remove) or modified (Change) during revision are, on average, more original than those retained. In particular, the comparison between Change and Change FB reveals that revisions systematically reduce originality: ideas are more original before being modified than after. This also applies to the most original ideas. The Max and Max FB values, representing the originality of the single most original clue pair generated by each group before and after revision, demonstrate that groups are capable of producing highly original ideas. Nevertheless, these ideas are rarely selected as the final output. This provides strong evidence of a selection bias, where original ideas are present in the idea pool but are not recognized or chosen during group decision-making. This phenomenon is consistent with prior research documenting a bias against originality (see Section 2). Original ideas often face greater skepticism or resistance, especially in uncertain environments or group settings.

Novel contributions are often perceived as less feasible, riskier, or more difficult for groups to reach consensus on, which can result in their systematic devaluation.

Table 6: Originality Process Differences

	(1)	(2)	(3)	(4)	(5)
	Difference FB	Difference Final	Difference Change	Difference Removal	Difference Max
Incentive Originality	0.003 (0.250)	0.375 (0.617)	0.607 (0.812)	0.980 (1.820)	0.025 (0.416)
Incentive Quality	0.212 (0.263)	0.074 (0.680)	0.648 (0.909)	0.801 (1.521)	0.079 (0.429)
Incentive Quantity	0.016 (0.246)	0.225 (0.699)	0.030 (0.895)	2.151 (1.598)	-0.269 (0.458)
Constant	1.399 (1.527)	8.963** (4.298)	5.255 (6.318)	-3.568 (11.591)	4.020 (2.789)
Group Controls	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓
Obs.	960	960	693	228	960
R2	0.046	0.085	0.054	0.078	0.025

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category. Difference FB is calculated as the average originality before feedback and revision minus the average originality after feedback and revision. Difference Final is the average originality after feedback and revision minus the originality of the final clue pair. Difference Change is the difference between the average originality of modified pairs before revision and their average originality after revision. Difference Removal is the originality of the removed pairs minus the average originality before feedback and revision. Difference Max is the change in originality between the most original clue pair before and after feedback and revision.

Perhaps most striking is the finding that none of the incentive treatments is able to mitigate this bias. Table 6, as well as Figure A2 and Table A22 in the Appendix, show that even the Originality incentive, which directly rewards groups for producing and selecting the most original clue pair, fails to prevent the decline in originality during feedback and selection. Groups in this treatment condition behave no differently in terms of originality loss than those in the Control group. This finding contradicts the intuitive assumption that aligning incentives with originality should enhance not only idea generation but also recognition and selection of original ideas. This failure may stem from two distinct but related mechanisms. First, groups may simply be unable to identify the most original idea, lacking a shared understanding of what constitutes originality. Second, even if original ideas are recognized, unconscious bias against originality may prevent their selection (cf. Rietzschel et al., 2019). While quantity-based incentives can increase the supply of original ideas, they do not ensure that such ideas are retained or selected. The persistent attrition of originality throughout the evaluation, revision, and selection phases represents a key bottleneck in the creative process.

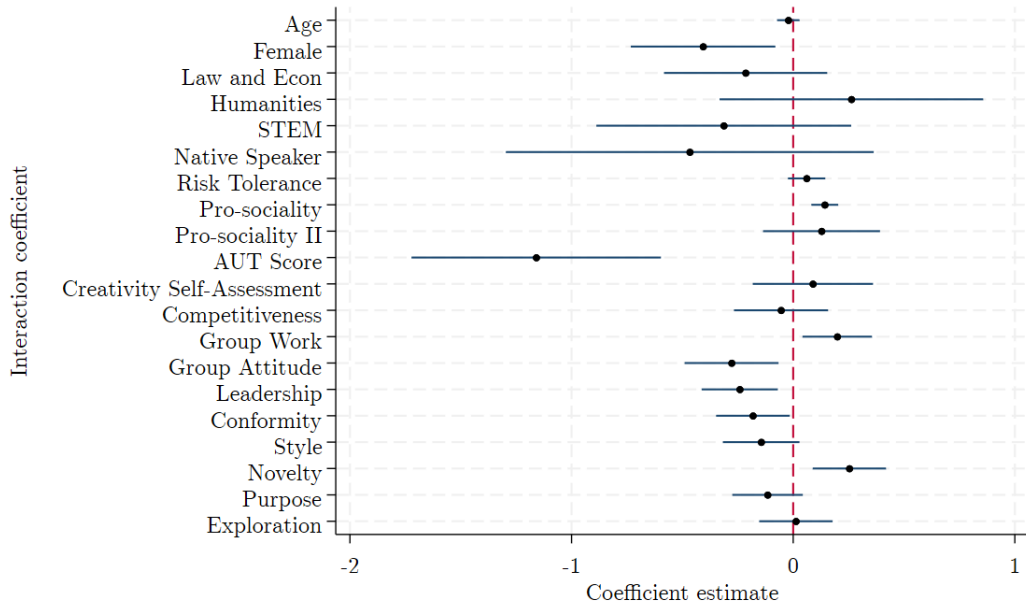
6 Mechanisms and Heterogeneity

6.1 Heterogeneous Effects

While the general positive effects of incentivizing quality, originality, and, in particular, quantity on the Innovation Index were established in the previous section, average treatment effects

can obscure substantial variation across different group compositions and characteristics. To better understand the conditions under which an incentive enhances innovation, I examine group-level heterogeneous effects across demographics, individual traits, group-related attitudes, creative capabilities, and cognitive styles. This approach allows me to identify boundary conditions for incentive efficacy and to unpack why certain types of incentives—Originality, Quantity, and Quality—may produce differential outcomes. Figures 5, 6, and 7 display the point estimates and 90% confidence intervals for interaction coefficients of each incentive treatment and heterogeneity dimension, using linear regressions with standard errors clustered at the group level. Figure 5 depicts the interactions for the Originality incentive, Figure 6 for the Quality incentive, and Figure 7 for the Quantity incentive⁹.

Figure 5: Heterogeneities Innovation Index Originality-PRP



Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Originality incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

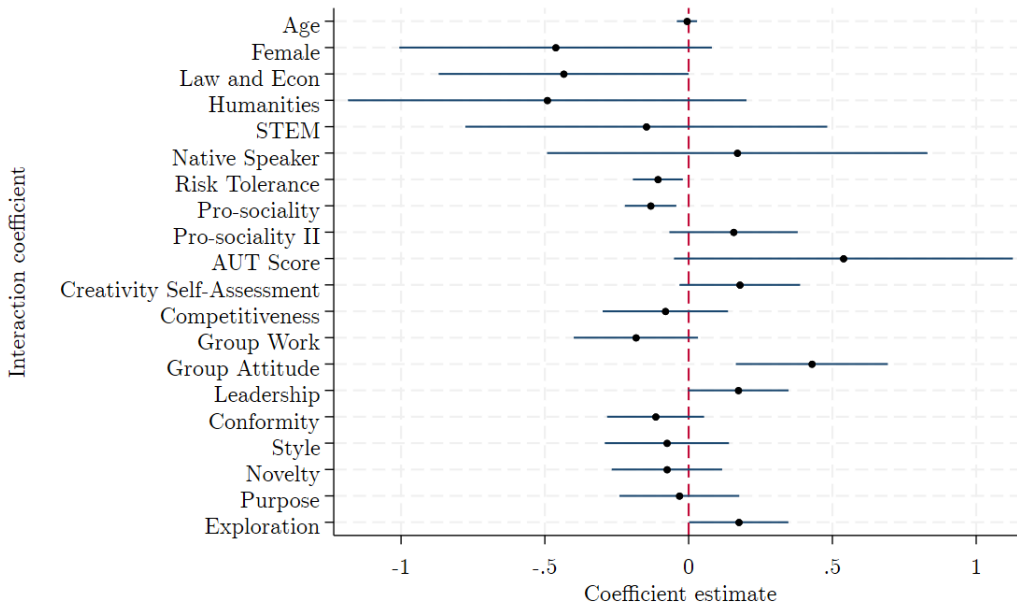
Starting with demographic moderators, gender composition shows selective moderation. Specifically, for Originality and Quantity incentives, groups with a higher proportion of female participants exhibit a significantly reduced effect on the Innovation Index, indicating lower responsiveness to these rewards. This is consistent with [Erat and Gneezy \(2016\)](#), who find

⁹ Further analyses of heterogeneous treatment effects across creativity dimensions are reported in the Appendix. Heterogeneous effects of the Quantity incentive on average group-level originality are shown before and after feedback and revision in Figures A3 and A4, with corresponding individual-level effects in Figures A5 and A6. Group- and individual-level heterogeneous effects on idea quantity are presented in Figures A7, A8, A9, and A10. Heterogeneous effects of the Quality and Originality incentive on idea quality are shown in Figures A11 and A12, respectively. Figures A13 and A14 illustrate heterogeneous effects on the binary indicator of innovative ideas for the Quantity and Quality incentives.

that competitive incentives reduce creativity for women. By contrast, the proportion of native speakers does not significantly affect incentive responses, likely due to the small number of non-native speakers and their high German proficiency. Notably, groups with more non-native speakers even display higher average originality independent of incentives (see Table A23 in the Appendix), in line with evidence that multilingualism enhances creativity (Fürst and Grin, 2023).

Turning to individual traits, risk tolerance moderates incentives differently. High-risk groups are slightly less responsive to quality-focused rewards, while responsiveness to Originality and Quantity incentives remains largely unaffected. This contrasts with findings such as those by Ederer and Manso (2013), where risk tolerance strongly shapes exploratory behavior. A plausible explanation is that their setting rewards exploration over time, whereas incentives here target specific output dimensions with immediate feedback, potentially leaving less scope for risk preferences to shape behavior. Pro-sociality exhibits incentive-specific heterogeneity. Under the Originality incentive, higher group-average pro-sociality (measured via the Dictator Game) significantly amplifies the treatment effect on innovation. No such moderation appears under Quantity or Quality incentives. Pro-sociality II (measured as willingness to make a sacrifice for the group on a Likert scale) does not moderate any incentive effect, suggesting that different pro-sociality measures capture distinct motivational mechanisms. Although Pro-sociality II is not related to incentives as in Charness and Grieco (2023), it is positively associated with higher creative output and originality (see Table A23 in the Appendix).

Figure 6: Heterogeneities Innovation Index Quality-PRP

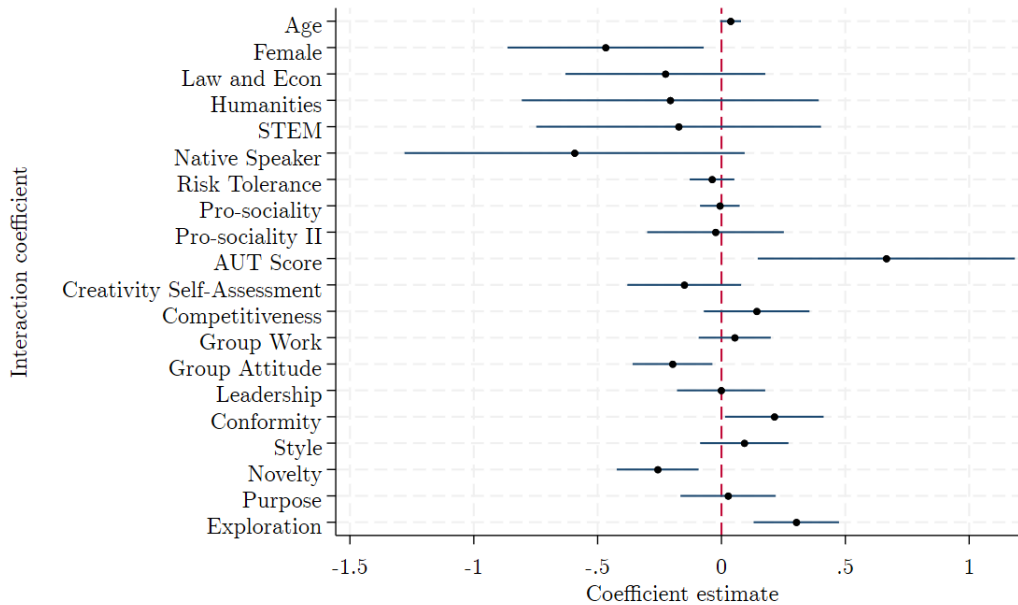


Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quality incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Following the analysis of individual traits, I turn to group-related attitudes. Groups with stronger preferences for collaboration exhibit significantly larger treatment effects under Originality incentives, while no significant moderation appears for Quantity or Quality rewards. Leadership orientation moderates incentives asymmetrically: it dampens the effect of Originality incentives but strengthens responsiveness to Quality incentives. This suggests that more hierarchical groups benefit from Quality rewards, where structured coordination supports convergence on high-quality solutions, whereas less hierarchical groups respond more to Originality incentives that foster exploration and divergent idea generation.

For Originality incentives, groups that view teamwork as more effective than individual work experience smaller gains, whereas for Quality incentives, these beliefs enhance responsiveness. Quantity incentives show only weak negative interaction. Conformity reduces the effectiveness of Originality incentives, indicating that groups composed of more conformity-prone members benefit less from originality-focused rewards. This suggests that individuals who prefer consensus over dissent may be less willing to explore unconventional ideas, limiting the effectiveness of incentives designed to promote divergent thinking. By contrast, conformity slightly enhances responsiveness to Quantity incentives, with a positive but only marginally significant interaction.

Figure 7: Heterogeneities Innovation Index Quantity-PRP



Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Finally, I consider creativity-related moderators. Originality incentives are most effective for groups with lower baseline AUT scores, suggesting a compensatory effect: groups with less inherent divergent capacity gain more from extrinsic originality rewards. Conversely,

Quantity incentives benefit groups with higher AUT scores more strongly, indicating that the incentive and ability are complementary. The external reward leverages an existing strength by channeling divergent thinking into measurable production. This aligns with evidence from real-effort tasks showing that baseline productivity and self-assessed relative ability are decisive for performance under exogenously imposed incentive schemes, while personality traits and economic preferences play at most a marginal role (Bašić et al., 2024). Independent of moderation, higher actual and self-assessed creative ability is strongly associated with superior creative performance in the task, validating that the task effectively captures meaningful aspects of creativity in groups (see Table A23 in the Appendix).

Creative traits related to work approach and motivation show selective moderation. Creative Style, reflecting a trial-and-error, unstructured approach, and Creative Purpose, capturing goal-directed intentions, do not significantly moderate any incentive. Novelty seeking (the propensity to pursue new and unfamiliar experiences) amplifies responsiveness to Originality incentives, suggesting that participants predisposed to exploring new ideas are particularly motivated when originality is rewarded. Conversely, novelty seeking slightly dampens responsiveness to Quantity incentives, implying that highly novelty-oriented individuals may prioritize idea quality or originality over sheer output. Creative Exploration, reflecting willingness to engage in open-ended, risk-prone creative processes, positively moderates both Quantity and Quality incentives. These results underscore a critical distinction: Novelty Seeking is outcome-oriented, aligning with incentives that reward generating unusual ideas (Originality), whereas Creative Exploration is process-oriented, aligning with incentives that reward effortful or iterative production (Quantity, Quality). This explains why the individuals may respond very differently across incentive types, despite both traits reflecting general creative tendencies.

Synthesizing the heterogeneous effects across incentive types reveals three clear patterns. **Originality incentives** are most effective in groups with stronger collaborative attitudes, stronger novelty-seeking tendencies, and lower baseline AUT scores, indicating a compensatory role for less inherently divergent groups. Their impact is weakened in groups with a higher share of female participants, greater conformity, and stronger leadership orientation, suggesting that hierarchy and consensus-seeking constrain responsiveness to divergence-focused rewards. **Quality incentives** are strongest in groups with pronounced leadership orientation, higher Creative Exploration, and stronger beliefs in group productivity, consistent with structured coordination and iterative experimentation supporting quality improvements. These effects are dampened by higher risk tolerance, while pro-sociality and novelty seeking play little role. **Quantity incentives** perform best in groups with higher baseline AUT scores and stronger Creative Exploration, leveraging existing divergent ability and exploratory work styles. Their effects are marginally strengthened by conformity but reduced in groups with a higher proportion of female participants and, to a lesser extent, by novelty seeking.

6.2 Mechanisms

In this section, I examine the mechanisms through which incentive treatments influence creative performance in groups. I focus in particular on the Quantity incentive, which shows the strongest and most consistent effects across quantity, originality, and innovation. Tables 7 and 8 report the results of a mediation analysis for idea quantity after feedback and revision and average originality, respectively¹⁰.

Mediation analysis provides a framework to decompose the total effect of a treatment into components that operate directly and indirectly through intermediate variables (mediators). More theoretical background can be found in *Appendix E: Mediation Analysis*. In the present analysis, I estimate the mediated and direct components using linear specifications for both the mediator and outcome equations. The same group-level control variables included in the main analysis are incorporated in both equations, and standard errors are clustered at the group level. For each potential mediator, I report the proportion of the total effect mediated, and corresponding standard errors and p-values. This implementation allows for a systematic assessment of the relative importance of different behavioral and psychological channels through which incentive treatments affect creative performance, while clearly distinguishing indirect pathways from direct treatment effects.

Table 7 reports the results of the mediation analysis for idea quantity after feedback and revision under the Quantity incentive, examining both behavioral and survey-based measures. Behavioral measures, recorded for each round of the task, capture effort via duration (time spent) and intensity (ideas generated, modifications made, or suggestions provided per minute). Additional behavioral indicators include the share of clue pairs changed or removed, the share of positive feedback, the number of suggestions provided and incorporated, and voting patterns. Communication is measured via chat sentiment and the number of chat messages. Survey-based measures, mostly collected at the end of the experiment, assess perceptions of received and given feedback, teamwork, leadership, peer pressure, group effort, self-efficacy, pro-social norms, motivation, enjoyment, and task difficulty.

For the outcome of quantity, the share of changed or removed pairs, the number or share of used suggestions, voting behavior, chat sentiment, and perceived group effort do not significantly mediate the effect. The Quantity incentive does not primarily operate through changes in group dynamics, feedback interactions, or communication. Instead, effort during ideation emerges as the key mechanism: both the duration and intensity of idea generation mediate a

¹⁰ Further mediation analyses across creativity dimensions and the composite measures of innovation are reported in the Appendix. Mechanisms of the Quantity incentive on average group-level originality before feedback and revision are shown in Table A24, with corresponding individual-level effects in Tables A25 and A26. Group- and individual-level mediating effects on idea quantity are presented in Tables A27, A28, and A29. Mediators of the effects of the Quality and Originality incentive on idea quality are shown in Tables A30 and A31, respectively. Tables A32 and A33 illustrate mechanisms on the binary indicator of innovative ideas for the Quantity and Quality incentives. Tables A34, A35, and A36 show the mediation analyses of the Innovation Index under Originality, Quality, and Quantity incentive treatments.

Table 7: Mediation Quantity after Feedback and Revision

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quantity FB	Incentive Quantity	Share of changed pairs	0.048	0.031	0.119
Quantity FB	Incentive Quantity	Share of removed pairs	-0.004	0.018	0.826
Quantity FB	Incentive Quantity	Share of given positive ratings	0.052	0.062	0.402
Quantity FB	Incentive Quantity	Random Vote	0.007	0.010	0.461
Quantity FB	Incentive Quantity	Unanimous Vote	-0.001	0.006	0.845
Quantity FB	Incentive Quantity	Share of used suggestions	-0.035	0.036	0.328
Quantity FB	Incentive Quantity	Number of suggestions	0.008	0.021	0.718
Quantity FB	Incentive Quantity	Teamwork	0.021	0.031	0.503
Quantity FB	Incentive Quantity	Leadership	0.000	0.017	0.991
Quantity FB	Incentive Quantity	Feedback Assessment	-0.021	0.048	0.655
Quantity FB	Incentive Quantity	Chat Sentiment	0.024	0.028	0.386
Quantity FB	Incentive Quantity	Number Chat Messages	-0.113	0.075	0.130
Quantity FB	Incentive Quantity	Group Effort	-0.008	0.034	0.809
Quantity FB	Incentive Quantity	Peer Pressure	0.011	0.095	0.907
Quantity FB	Incentive Quantity	Enjoyment	0.046	0.057	0.423
Quantity FB	Incentive Quantity	Communication	-0.018	0.042	0.659
Quantity FB	Incentive Quantity	Difficulty Quality	0.007	0.059	0.906
Quantity FB	Incentive Quantity	Difficulty Originality	-0.020	0.034	0.561
Quantity FB	Incentive Quantity	Difficulty Quantity	0.049	0.146	0.738
Quantity FB	Incentive Quantity	Duration Ideagen	0.373	0.108	0.001
Quantity FB	Incentive Quantity	Duration FB	0.062	0.090	0.494
Quantity FB	Incentive Quantity	Duration Revision	0.024	0.070	0.733
Quantity FB	Incentive Quantity	Intensity Ideagen	0.332	0.102	0.001
Quantity FB	Incentive Quantity	Intensity FB	-0.001	0.006	0.884
Quantity FB	Incentive Quantity	Intensity Revision	-0.013	0.037	0.725
No control group					
Quantity FB	Incentive Quantity	Self-efficacy	0.403	0.177	0.023
Quantity FB	Incentive Quantity	Pro-social Group	0.012	0.043	0.769
Quantity FB	Incentive Quantity	Motivation Competition	0.191	0.092	0.038
Quantity FB	Incentive Quantity	Motivation Payment	-0.042	0.070	0.549

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

substantial portion of the incentive effect (37% and 33%, respectively). In contrast, duration and intensity during feedback and revision do not mediate the effect. This indicates that the incentive primarily increases the time and intensity devoted to generating ideas.

Survey-based measures of dimension-specific self-efficacy and competitive motivation also mediate the effect of the Quantity incentive. The incentive shifts participants' attention to the incentivized dimension and strengthens their motivation to compete, which in turn leads to greater idea output. Because self-efficacy is measured specifically for each incentive dimension—confidence in producing many ideas under the Quantity incentive, original ideas under the Originality incentive, and high-quality ideas under the Quality incentive—it functions as a mediator rather than a moderator. The incentive directly activates domain-specific confidence, which then drives performance within that dimension. As these two measures are only collected in the three Treatment groups, their mediation effects should be interpreted as operating between Treatment groups, not as causal effects relative to the Control group. The strong mediation via self-efficacy under the Quantity incentive aligns with expectations: generating a high number of ideas is a controllable, effort-based task that allows participants to act directly on their confidence, whereas originality or quality is more uncertain and less amenable to self-efficacy-driven performance.

Overall, the Quantity incentive increases quantity primarily by boosting effort during idea generation. Participants spend more time, engage more intensely, and rely on their confidence and competitive drive to generate more clue pairs. The effect does not appear to operate through social dynamics or collaborative interactions.

The mediation pattern for average originality under the Quantity incentive is similar, but introduces an important additional channel (see Table 8). Effort again plays a central role: time spent in the ideation phase significantly mediates about 35% of the effect of the Quantity incentive, indicating that longer engagement facilitates broader exploration and increases the likelihood of original ideas. Based on the findings in section 5, I include quantity as a potential mediator and, consistent with the previous results (see Figure 3), quantity clearly mediates a substantial portion of the effect (nearly 40%), showing that Quantity incentives increase the number of ideas that, in turn, translates into higher average originality. Other behavioral measures, such as the share of changed or removed pairs, the share of given positive ratings, the number and implementation of suggestions, and voting behavior have no significant mediation contributions. Likewise, group-level processes, including teamwork, leadership, communication, feedback assessment, group effort, and peer pressure do not mediate significantly. Self-efficacy contributes marginally, suggesting that participants' confidence in their own ability to produce a larger number of ideas supports originality, although this effect is weaker than for quantity. Unlike for quantity, the intensity of effort plays no significant role. Originality appears to rely more on sustained exploration and producing a sufficiently large number of ideas to increase the chance of generating unique solutions.

Taken together, the mediation analyses consistently show that individual effort is the primary mechanism through which the Quantity incentive affects creative performance. For

Table 8: Mediation Originality after Feedback and Revision

Outcome	Treatment	Mediator	Proportion	SE	p-value
Originality \emptyset FB	Incentive Quantity	Share of changed pairs	0.070	0.074	0.343
Originality \emptyset FB	Incentive Quantity	Share of removed pairs	-0.001	0.006	0.901
Originality \emptyset FB	Incentive Quantity	Share of given positive ratings	0.044	0.055	0.422
Originality \emptyset FB	Incentive Quantity	Random Vote	0.009	0.017	0.592
Originality \emptyset FB	Incentive Quantity	Unanimous Vote	-0.004	0.020	0.843
Originality \emptyset FB	Incentive Quantity	Share of used suggestions	0.043	0.045	0.337
Originality \emptyset FB	Incentive Quantity	Number of suggestions	0.014	0.040	0.720
Originality \emptyset FB	Incentive Quantity	Teamwork	0.049	0.071	0.491
Originality \emptyset FB	Incentive Quantity	Leadership	-0.000	0.034	0.991
Originality \emptyset FB	Incentive Quantity	Feedback Assessment	-0.035	0.071	0.620
Originality \emptyset FB	Incentive Quantity	Chat Sentiment	-0.089	0.065	0.170
Originality \emptyset FB	Incentive Quantity	Number Chat Messages	-0.183	0.132	0.165
Originality \emptyset FB	Incentive Quantity	Group Effort	0.009	0.035	0.786
Originality \emptyset FB	Incentive Quantity	Peer Pressure	0.165	0.146	0.259
Originality \emptyset FB	Incentive Quantity	Enjoyment	0.047	0.085	0.580
Originality \emptyset FB	Incentive Quantity	Communication	0.038	0.058	0.510
Originality \emptyset FB	Incentive Quantity	Difficulty Quality	0.004	0.037	0.906
Originality \emptyset FB	Incentive Quantity	Difficulty Originality	-0.032	0.053	0.547
Originality \emptyset FB	Incentive Quantity	Difficulty Quantity	0.032	0.100	0.747
Originality \emptyset FB	Incentive Quantity	Quantity FB	0.393	0.150	0.009
Originality \emptyset FB	Incentive Quantity	Duration Ideagen	0.352	0.164	0.032
Originality \emptyset FB	Incentive Quantity	Duration FB	0.012	0.028	0.653
Originality \emptyset FB	Incentive Quantity	Duration Revision	-0.001	0.009	0.889
Originality \emptyset FB	Incentive Quantity	Intensity Ideagen	-0.014	0.059	0.816
Originality \emptyset FB	Incentive Quantity	Intensity FB	-0.007	0.043	0.873
Originality \emptyset FB	Incentive Quantity	Intensity Revision	-0.008	0.022	0.714
No control group					
Originality \emptyset FB	Incentive Quantity	Tradeoff	0.016	0.105	0.881
Originality \emptyset FB	Incentive Quantity	Self-efficacy	0.425	0.238	0.074
Originality \emptyset FB	Incentive Quantity	Pro-social Group	0.056	0.104	0.588
Originality \emptyset FB	Incentive Quantity	Motivation Competition	0.268	0.192	0.163
Originality \emptyset FB	Incentive Quantity	Motivation Payment	0.017	0.095	0.859

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

quantity, participants' prolonged and intensive engagement, reinforced by self-efficacy and competitive motivation, directly increases output. For originality, the mechanism is partly indirect: greater engagement and higher idea production expand the solution space, increasing the probability of novel ideas emerging. Group processes do not meaningfully mediate these effects. Beyond treatment mediation, effort appears to be a central determinant of creative performance overall: both duration and intensity in the idea generation phase are positively associated with innovation, while effort during the feedback phase may reduce originality but can enhance quality (see Table A39 in the Appendix). Although feedback does not mediate treatment effects, more critical feedback, both in ratings and in the chat, is positively related to average originality (see Tables A38 and A37 in the Appendix).

6.3 Secondary Outcomes

To shed further light on the crucial exploratory processes in idea generation, I analyze two additional measures of semantic exploration: idea divergence and flexibility. Unlike originality, which quantifies semantic distance relative to a target word, these metrics capture the dispersion and variation of ideas independently of the target word, reflecting the extent to which participants explore the semantic space more generally. Divergence measures the spread of all ideas generated by a participant in a single round around their centroid, calculated as the mean cosine distance between each word and the average vector of all words. Flexibility is the mean pairwise cosine distance among all ideas, capturing within-round variation. Tables 9 and 10 report OLS results of group-level idea divergence and flexibility across incentive treatments. Individual-level results can be found in Tables A40 and A41 in the Appendix.

These measures build on classical theories of divergent thinking, which enables both the generation of varied ideas and the ability to switch between different conceptual categories (Guilford, 1967). Higher divergence and flexibility indicate greater exploration, reflecting the ability to access and combine ideas from distant regions of semantic memory (Olson et al., 2021; Baader et al., 2025; Yaman et al., 2025). Overall, group-level flexibility is relatively high (pre-feedback mean: 71.88), while idea divergence is comparatively low (pre-feedback mean: 29.96). This indicates that participants generate ideas that are semantically distinct from one another within a round but remain clustered within a relatively narrow region of the semantic space. In other words, participants readily explore different directions in their thinking, yet the overall breadth of exploration is constrained. A plausible explanation for this pattern is that all ideas are generated with respect to the same target word, which anchors the conceptual space and limits the extent to which groups venture into entirely new semantic regions.

Table 9 shows that Quantity incentives substantially increase idea divergence across all specifications. In the pre-feedback phase, the effect is large and highly significant (column 1, 3.514, $p < 0.01$), demonstrating that groups produce a wider range of ideas when the number of ideas is rewarded. This effect remains robust after including group controls and round fixed

Table 9: Group Idea Divergence

	before feedback and revision				after feedback and revision			
	(1) Divergence	(2) Divergence	(3) Divergence	(4) Divergence	(5) Divergence FB	(6) Divergence FB	(7) Divergence FB	(8) Divergence FB
Incentive Originality	0.665 (0.944)	1.010 (0.885)	1.010 (0.887)	0.615 (0.459)	0.610 (0.939)	0.805 (0.865)	0.805 (0.868)	0.458 (0.478)
Incentive Quality	0.743 (0.973)	1.194 (0.912)	1.194 (0.915)	0.891** (0.434)	0.286 (0.953)	0.775 (0.899)	0.775 (0.902)	0.634 (0.483)
Incentive Quantity	3.514*** (1.050)	3.980*** (0.943)	3.980*** (0.946)	0.356 (0.474)	3.309*** (1.024)	3.735*** (0.936)	3.735*** (0.938)	0.299 (0.490)
Quality				0.179 (0.312)				0.152 (0.320)
Quantity				1.779*** (0.180)				
Quality × Quantity				-0.034 (0.052)				
Quantity FB								1.756*** (0.177)
Quality × Quantity FB								-0.039 (0.055)
Constant	28.725*** (0.724)	17.280*** (5.345)	12.840** (5.396)	16.367*** (2.556)	28.223*** (0.718)	14.394*** (5.125)	10.219** (5.163)	14.114*** (2.777)
Group Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.039	0.132	0.273	0.646	0.039	0.122	0.250	0.619

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

effects (columns 2 and 3, 3.980, $p < 0.01$). After feedback and revision, the effect persists (columns 6–7, 3.735, $p < 0.01$), suggesting that receiving feedback and making revisions does not diminish its positive impact on idea divergence. In contrast, Originality incentives do not significantly affect divergence, and Quality incentives show only modest effects if controlling for quantity and quality (column 4, 0.891, $p < 0.05$) in one specification. Importantly, when controlling for the number of ideas generated (columns 4 and 8), part of the idea divergence boost from the Quantity incentive appears mediated by increased idea production. These results indicate that quantity-based incentives encourage groups to expand the semantic space they explore, independently of the target word, reinforcing previous findings that generating more ideas allows participants to uncover less conventional associations.

Table 10 reveals a complementary pattern. In the pre-feedback phase, Quantity incentives modestly increase flexibility (columns 2–3, 0.959, $p < 0.1$), suggesting that rewarding a higher number of ideas slightly enhances variation among ideas. Quality incentives show a similar, marginally significant pattern (columns 2–3, 1.083, $p < 0.1$), whereas Originality incentives have no significant effect. In the post-feedback phase (columns 5–8), none of the incentive treatments significantly affect flexibility. Only the direct effect of quantity on flexibility is highly significant (column 8, 0.304, $p < 0.01$).

Table 10: Group Flexibility

	before feedback and revision				after feedback and revision			
	(1) Flexibility	(2) Flexibility	(3) Flexibility	(4) Flexibility	(5) Flexibility FB	(6) Flexibility FB	(7) Flexibility FB	(8) Flexibility FB
Incentive Originality	0.107 (0.622)	0.240 (0.612)	0.240 (0.614)	0.184 (0.604)	0.042 (0.589)	-0.017 (0.569)	-0.017 (0.570)	-0.015 (0.558)
Incentive Quality	0.822 (0.595)	1.083* (0.549)	1.083* (0.551)	1.055* (0.539)	0.254 (0.676)	0.603 (0.631)	0.603 (0.632)	0.671 (0.627)
Incentive Quantity	0.838 (0.613)	0.959* (0.546)	0.959* (0.547)	0.419 (0.547)	0.519 (0.544)	0.648 (0.547)	0.648 (0.548)	0.264 (0.535)
Quality				0.270 (0.442)				0.253 (0.386)
Quantity				0.322*** (0.101)				
Quality × Quantity				-0.054 (0.060)				
Quantity FB								0.304*** (0.087)
Quality × Quantity FB								-0.084 (0.055)
Constant	71.435*** (0.412)	69.740*** (3.599)	68.653*** (3.647)	69.182*** (3.663)	71.358*** (0.388)	65.637*** (3.781)	64.452*** (3.824)	64.960*** (3.846)
Group Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓		✓	✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.004	0.038	0.154	0.165	0.001	0.033	0.153	0.164

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Taken together, flexibility and idea divergence complement the main originality results (see Table 1 and Figure 3). Generating more clue pairs within a round raises originality, and Quantity incentives stimulate this higher output. Quantity incentives enable groups to systematically explore a larger and more varied idea space, which in turn increases the likelihood of original associations.

Results for further secondary outcomes are reported in the Appendix in Tables A42 and A43. Groups self-assess their final clue pair for quality and originality (Quality III and Originality III), while quality assessors rate all clue pairs on quality, originality, and innovation (Quality II, Originality II, Innovation II). Consistent with the primary outcomes, the Quality incentive increases both assessors' average and final clue pair quality ratings. In contrast, the Quantity incentive, which performs best in the primary outcomes, slightly increases groups' self-assessed originality but does not affect the assessors' ratings. The Originality incentive, which shows weak to no effects on the primary outcomes, significantly lowers assessors' ratings of originality and innovation. Overall, these subjective evaluations by a single human rater per clue pair are less reliable than the objective primary outcomes. Assessors are not independent judges: their ratings correlate with their own performance in the preceding guessing task (see Table A44). Nonetheless, this pattern reinforces one of the main findings: directly rewarding originality does not enhance creative performance.

7 Discussion and Conclusion

Despite the growing importance of creativity and teamwork in the labor market, groups often fail to fully realize their creative potential. Although teams benefit from diverse perspectives and complementary skills, group processes—particularly during idea evaluation and selection—frequently hinder the generation, recognition, and translation of creative contributions into implementable solutions. Little is known about how these group processes and incentive structures shape creativity across multiple dimensions. This study addresses this gap by opening the “black box” of group interaction, examining how idea generation and selection can be effectively incentivized, and identifying which incentive schemes optimize overall innovative performance. Using a novel task that captures the entire creative process—idea generation, evaluation, and selection—this paper provides the first experimental evidence on incentivizing and measuring different dimensions of group creative performance. This approach enables a deeper understanding of creativity in all its facets and informs the design of incentive schemes that effectively stimulate all creativity dimensions.

Quantity incentives lead groups to produce more ideas and systematically explore a larger and more varied idea space. As participants generate more clue pairs for the same target word, later ideas are systematically more original than earlier ones. This exploration translates not only into higher output but also into higher originality—an indirect effect that arises because generating more ideas increases the likelihood of discovering novel combinations (cf. [Ederer and Manso, 2013](#)). In contrast, Quality incentives encourage exploitation, focusing more on a limited set of ideas, improving quality but limiting exploratory breadth. Originality incentives, however, do not lead to higher originality, potentially due to the crowding out of intrinsic motivation that may be higher for originality than for other dimensions and uncertainty about what constitutes originality. Originality emerges more reliably as a by-product of broad exploration than as a direct response to incentives targeting it. In contrast, quality responds positively to incentives targeting any creativity dimension, suggesting that it is a fundamental and consistently valued aspect in all treatment groups.

In promoting the most innovative ideas, Quantity incentives consistently outperform those focused on quality or originality, as they not only increase idea production but also create spillovers on both quality and originality. Quality incentives appear to yield smaller but directionally similar gains, while directly rewarding originality fails to produce measurable improvements on the indicator for the most innovative ideas. These findings are consistent with [Laske et al. \(2024\)](#), who show at the individual level that piece-rate incentives for idea generation can increase the number of innovative ideas. Building on this, the present results show that group-level relative performance incentives directly rewarding originality are also less effective than those rewarding quantity in a task encompassing idea generation, evaluation, and selection, likely reflecting higher perceived self-efficacy and a clearer effort-performance link.

Despite the benefits of incentives during idea generation, my analyses reveal a systematic decline in originality as ideas progress through evaluation, revision, and final selection. Groups frequently discard or modify highly original ideas, resulting in a steady decline of originality across the creative process. This pattern reflects social-cognitive mechanisms documented in prior psychological research: evaluation phases are particularly susceptible to cognitive biases and conformity pressures that disadvantage unconventional ideas (Mueller et al., 2011). The persistence of this pattern across all incentive conditions suggests that motivational alignment alone cannot overcome structural biases that suppress originality. Even the Originality incentive, which explicitly rewards groups for selecting the most original clue pair, fails to prevent this decline. Two interrelated mechanisms may account for this failure. First, groups may lack a shared understanding of what constitutes originality, making it difficult to identify the most original ideas. Second, even when originality is recognized, unconscious bias against originality may inhibit the selection of unconventional ideas. The group setting likely amplifies these tendencies, as individuals may defer to the perceived safety of consensus or avoid advocating for ideas seen as too unconventional (Rietzschel et al., 2019).

Drawing on rich process and survey-based measures, I examine how incentive treatments influence creative performance in groups and how specific behaviors and interaction patterns facilitate or constrain creativity in multiple dimensions. Mediation analyses indicate that the Quantity incentive enhances performance primarily by increasing individual effort: participants invest more time generating ideas. In contrast to Englmaier et al. (2023, 2024, 2025), who examine performance in non-routine analytical tasks, group-level processes such as coordination and leadership do not significantly mediate treatment effects in this setting. Instead, the individual idea generation phase is the central locus of treatment effects, highlighting that incentives operate through motivational rather than organizational channels. Beyond incentive effects, early-stage effort and critical feedback are central determinants of creative group performance. Sustained engagement during idea generation is associated with higher quantity and originality, while intensive feedback is positively linked to quality, demonstrating the inherent trade-offs in creative collaboration. While I do not find evidence that incentives operate through the pro-social mechanism seen in Charness and Grieco (2023), the average pro-social attitude within groups is positively associated with group performance in my setting.

These findings should be interpreted in light of three main considerations. First, the study captures quality only for the final clue pair, limiting the ability to track how perceptions of feasibility and usefulness evolve throughout the creative process. Therefore, potential biases favoring more practical at the expense of less original solutions cannot be systematically identified. Future research should extend quality measurement to intermediate stages, enabling a finer analysis of how evaluative trade-offs between quality and originality develop over time. Second, the laboratory setting provides strong causal identification but differs from real-world teams, especially in terms of time horizons, repeated interactions, and hierarchical coordination. Future work can improve external validity by studying these dynamics in organizational contexts. Third, the rapid integration of generative AI into creative work, particularly in

group contexts, represents a novel research frontier. As AI tools increasingly contribute to ideation and evaluation, they may fundamentally alter the social dynamics of creativity in groups. AI could reduce production blocking and broaden exploration by providing immediate associative input, but it might also exacerbate conformity if used as an authority in evaluation and idea generation. Investigating how exposure to AI changes idea generation, evaluation, and selection represents an important direction for future studies seeking to understand creativity in hybrid human–AI teams.

From a managerial and policy perspective, the results reveal a fundamental challenge in managing creativity in groups: while quantity-based incentives can increase the generation of original ideas, they do not ensure their retention or selection. The persistent attrition of originality, even in controlled and incentivized environments, represents a critical bottleneck in the creative process and indicates that even aligned incentives are not sufficient to overcome biases against original ideas. This insight suggests that organizations aiming to foster creativity should focus less on rewarding originality directly and more on designing environments that encourage extensive idea exploration and protect unconventional ideas from premature rejection. Organizations can strategically combine individual thinking phases with collective interaction, allowing employees to develop ideas independently before exposing them to group evaluation. Clear, quantity-oriented incentive schemes can support this process by increasing output, which indirectly enhances both originality and quality. Managers should actively encourage deviation and dissent to ensure that original contributions are not filtered out too early in the creative process. Finally, evaluating originality independently of feasibility may allow original ideas to be judged on their own merits before practicality is considered. Designing incentive systems that encourage exploration, while mitigating group-level biases against originality, remains a key challenge for promoting effective creative collaboration.

References

- Amabile, T. M., Conti, R., Coon, H., Lazenby, J., and Herron, M. (1996). Assessing the Work Environment for Creativity. *The Academy of Management Journal*, 39(5):1154–1184.
- Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2009). Large Stakes and Big Mistakes. *The Review of Economic Studies*, 76(2):451–469.
- Arthur, W. B. (2007). The structure of invention. *Research Policy*, 36(2):274–287.
- Attanasi, G., Chessa, M., Gil-Gallen, S., and Llerena, P. (2021). A survey on experimental elicitation of creativity in economics. *Revue d'économie industrielle*, 174(2):273–324.
- Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.
- Autor, D. H. and Price, B. (2013). The Changing Task Composition of the US Labor Market: An Update of Autor, Levy, and Murnane (2003). Working paper.
- Azoulay, P., Graff Zivin, J. S., and Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3):527–554.
- Baader, M., Fischbacher, U., Starmer, C., and Tufano, F. (2025). Measuring Creativity: Associative Thinking in Semantic Networks. Working Paper 2025-01, CeDEx Discussion Paper Series.
- Baer, M., Leenders, R. T. A. J., Oldham, G. R., and Vadera, A. K. (2010). Win or Lose the Battle for Creativity: The Power and Perils of Intergroup Competition. *Academy of Management Journal*, 53(4):827–845.
- Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.
- Bašić, Z., Bortolotti, S., Salicath, D., Schmidt, S., Schneider, S. O., and Sutter, M. (2024). One Size Fits All? The Interplay of Incentives, Effort Provision, and Personality. Available at SSRN 4954146.
- Beaty, R., Patterson, J., Merseal, H., Johnson, D., Agnoli, S., Baas, M., Baker, B., Barbot, B., Benedek, M., Borhani, K., Chen, Q., Christensen, J., Corazza, G. E., Forthmann, B., Karwowski, M., Kazemian, N., Kreisberg-Nitzav, A., Kenett, Y., Link, A., Lubart, T., Mercier, M., Miroshnik, K., Ovando-Tellez, M., Primi, R., Puente-Díaz, R., Said-Metwaly, S., Stevenson, C., Vartanian, M., Volle, E., and van Hell, J. (2023). Multilingual semantic distance: Automatic verbal creativity assessment in many languages. *Psychology of Aesthetics, Creativity, and the Arts*, 17(4):495–507.
- Beaty, R. E. and Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2):757–780.
- Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., and Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & cognition*, 42(7):1186–1197.

- Blaufus, K., Piehl, K., and Schröder, M. (2025). Input Versus Output Incentives in Idea Generation - An Experimental Analysis. *Available at SSRN 5101127*.
- 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.
- Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.
- Charness, G. and Grieco, D. (2019). Creativity and Incentives. *Journal of the European Economic Association*, 17(2):454–496.
- Charness, G. and Grieco, D. (2023). Creativity and Corporate Culture*. *The Economic Journal*, 133(653):1846–1870.
- Chen, C. X., Williamson, M. G., and Zhou, F. H. (2012). Reward System Design and Group Creativity: An Experimental Investigation. *The Accounting Review*, 87(6):1885–1911.
- 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.
- Cross, R., Rebele, R., and Grant, A. (2016). Collaborative Overload. *Harvard Business Review*. Section: Collaboration and teams.
- De Dreu, C. K. W., Nijstad, B. A., Bechtoldt, M. N., and Baas, M. (2011). Group creativity and innovation: A motivated information processing perspective. *Psychology of Aesthetics, Creativity, and the Arts*, 5(1):81–89.
- Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627.
- Deming, D. J. and Silliman, M. I. (2024). Skills and human capital in the labor market. Working Paper 32908, National Bureau of Economic Research.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.
- Duncker, K. (1945). On problem-solving. *Psychological monographs*, 58(5).
- Eckartz, K., Kirchkamp, O., and Schunk, D. (2012). How Do Incentives Affect Creativity? Working Paper 4049, CESifo Working Paper.
- Ederer, F. and Manso, G. (2013). Is Pay for Performance Detrimental to Innovation? *Management Science*, 59(7):1496–1513.
- Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2023). The Efficacy of Tournaments for Non-Routine Team Tasks. *Journal of Labor Economics*, page 725553.
- Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2024). The Effect of Incentives in Non-Routine Analytical Team Tasks. *Journal of Political Economy*, page 729443.

- Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2025). The value of leadership: Evidence from a large-scale field experiment. *The Leadership Quarterly*, 36(3):101869.
- Erat, S. and Gneezy, U. (2016). Incentives for creativity. *Experimental Economics*, 19(2):269–280.
- Frederiksen, A., Hansen, D. B. S., and Manchester, C. F. (2024). Group-Based Incentives and Individual Performance: A Study of the Effort Response. *ILR Review*, 77(2):273–293.
- Friebel, G., Heinz, M., Krueger, M., and Zubanov, N. (2017). Team Incentives and Performance: Evidence from a Retail Chain. *American Economic Review*, 107(8):2168–2203.
- Fürst, G. and Grin, F. (2023). Multilingualism, multicultural experience, cognition, and creativity. *Frontiers in Psychology*, 14:1155158.
- 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.
- Gneezy, U., Laske, K., and Schröder, M. (2022). Teams do not outperform individuals in a simple creative task. *Applied Economics Letters*, 29(15):1393–1398.
- Goette, L. and Senn, J. (2024). Incentivizing interdependent tasks: Evidence from a real-effort experiment. *Journal of Economic Behavior & Organization*, 227:106718.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning Word Vectors for 157 Languages. arXiv:1802.06893[cs.CL].
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.
- Gross, D. P. (2020). Creativity Under Fire: The Effects of Competition on Creative Production. *The Review of Economics and Statistics*, 102(3):583–599.
- Grossman, G. M. and Helpman, E. (1994). Endogenous Innovation in the Theory of Growth. *Journal of Economic Perspectives*, 8(1):23–44.
- Guhr, O., Schumann, A.-K., Bahrman, F., and Böhme, H. J. (2020). Training a broad-coverage german sentiment classification model for dialog systems. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1620–1625, Marseille, France. European Language Resources Association.
- Guilford, J. P. (1967). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior*, 1(1):3–14.
- Kachelmeier, S. J., Reichert, B. E., and Williamson, M. G. (2008). Measuring and Motivating Quantity, Creativity, or Both. *Journal of Accounting Research*, 46(2):341–373.
- Kachelmeier, S. J., Webb, A., and Williamson, M. G. (2020). Do Performance-Contingent Incentives Help or Hinder Divergent Thinking? Available at SSRN 3745282.
- Kandel, E. and Lazear, E. P. (1992). Peer Pressure and Partnerships. *Journal of Political Economy*, 100(4):801–817.

- Karwowski, M., Lebuda, I., and Wiśniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *The International Journal of Creativity & Problem Solving*, 28(1):45–57.
- Kleine, M. (2021). No eureka! Incentives Hurt Creative Breakthrough Irrespective of the Incentives' Frame. *Available at SSRN 3848440*.
- Korde, R. and Paulus, P. B. (2017). Alternating individual and group idea generation: Finding the elusive synergy. *Journal of Experimental Social Psychology*, 70:177–190.
- Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Riemann, D., and Nissen, C. (2014). Entwicklung von 130 deutschsprachigen Compound Remote Associate (CRA)-Worträtseln zur Untersuchung kreativer Prozesse im deutschen Sprachraum. *Psychologische Rundschau*, 65(4):200–211.
- Laske, K., Römer, N., and Schröder, M. (2024). Piece-rate incentives and idea generation — An experimental analysis. Working Paper 11594, CESifo Working Paper.
- Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review*, 90(5):1346–1361.
- Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy*, 89(5):841–864.
- Leopold, T., Di Battista, A., Jativa, X., Sharma, S., Li, R., and Grayling, S. (2025). Future of jobs report 2025. World Economic Forum.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3):220–232.
- Mueller, J., Melwani, S., and Goncalo, J. (2011). The Bias Against Creativity: Why People Desire But Reject Creative Ideas. *Psychological science*, 23:13–7.
- Nguyen, T. Q., Schmid, I., Ogburn, E. L., and Stuart, E. A. (2022). Clarifying causal mediation analysis: Effect identification via three assumptions and five potential outcomes. *Journal of Causal Inference*, 10(1):246–279.
- Nijstad, B. A., Stroebe, W., and Lodewijckx, H. F. (2003). Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, 39(6):531–548.
- Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., and Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25):e2022340118.
- Organisciak, P., Acar, S., Dumas, D., and Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, 49:101356.
- Paulus, P. B., Putman, V. L., Dugosh, K. L., Dzindolet, M. T., and Coskun, H. (2002). Social and Cognitive Influences in Group Brainstorming: Predicting Production Gains and Losses. *European Review of Social Psychology*, 12(1):299–325.

- Ramm, J., Tjøtta, S., and Torsvik, G. (2013). Incentives and Creativity in Groups. *Available at SSRN 2319864*.
- Rietzschel, E. F., Nijstad, B. A., and Stroebe, W. (2019). Why Great Ideas Are Often Overlooked: A Review and Theoretical Analysis of Research on Idea Evaluation and Selection. In Paulus, P. B. and Nijstad, B. A., editors, *The Oxford Handbook of Group Creativity and Innovation*, pages 178–197. Oxford University Press.
- Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5):71–102.
- Sampson, T. (2023). Technology Gaps, Trade, and Income. *American Economic Review*, 113(2):472–513.
- Yaman, A., Tian, S., and Lindström, B. (2025). Semantic knowledge guides innovation and drives cultural evolution. arXiv:2510.12837 [cs.MA].

Appendix A: Supplementary Tables

Table A1: Questionnaire I Ideators

No.	Question	Response Options
1	I found it difficult to create useful clue pairs.	5-point Likert scale
2	I found it difficult to create original clue pairs.	5-point Likert scale
3	I found it difficult to create many clue pairs.	5-point Likert scale
4	What strategy did you use for clue pair creation?	Open-ended question (optional)
5	I consider the feedback I gave to be helpful.	5-point Likert scale
6	The feedback I received was helpful.	5-point Likert scale
7	My group members were very willing to accept feedback.	5-point Likert scale
8	I felt that my ideas and I were valued.	5-point Likert scale
9	Our collaboration in the group was productive.	5-point Likert scale
10	One person was dominant in leading the group.	5-point Likert scale
11	My group developed a real team spirit.	5-point Likert scale
12	Our individual skills and perspectives complemented each other well.	5-point Likert scale
13	All group members made an effort.	5-point Likert scale
14	All group members contributed equally.	5-point Likert scale
15	In general, I prefer to agree with the majority rather than cause conflict.	5-point Likert scale
16	I enjoyed the task.	5-point Likert scale
17	I didn't want to contribute less than the others in my group.	5-point Likert scale
18	The performance of the guessers met my expectations.	5-point Likert scale
19	We imagined what the people guessing might be like and tailored our clue pairs accordingly.	5-point Likert scale
20	I think that my performance on the task was better than that of an AI.	5-point Likert scale
Only for treatment groups:		
21	The competition with the other groups motivated me.	5-point Likert scale
22	The bonus payment motivated me.	5-point Likert scale
23	Since our pay depended on the group's success, we supported each other strongly.	5-point Likert scale
24	Since our pay depended on the group's success, we all felt obligated to do our best.	5-point Likert scale
25	My group members and I focused heavily on the quantity/quality/originality of clue pairs.	5-point Likert scale
26	By focusing on quantity/quality/originality, we have neglected quality/quantity/originality and quality/quantity/originality.	5-point Likert scale
27	By putting in more effort, I was able to improve my ideas in terms of quantity/quality/originality.	5-point Likert scale

Note: This questionnaire is translated from German and was displayed to the ideators. The 5-point Likert scale provides five options: strongly agree, agree, neutral, disagree, strongly disagree.

Table A2: Questionnaire II Ideators

No.	Question	Response Options
1	Age	Open-ended question
2	Gender	Male/Female/Diverse
3	Field of Study	Humanities/Sports/Law, Economics, and Social Sciences/Mathematics, Natural Sciences/Human Medicine, Health Sciences/Agricultural, Forestry, and Nutritional Sciences, Veterinary Medicine/Arts, Art History/Engineering/Other Native language/Very good/Good/Okay/Rather poor/Not at all
4	How well can you read and write German?	0-10 scale
5	How would you personally assess yourself: Are you generally a risk-taker, or do you try to avoid risks?	5-point Likert scale
6	I think that I am an above-average creative person.	5-point Likert scale
7	It is important to me to be a creative person.	5-point Likert scale
8	I am good at finding creative solutions to problems.	5-point Likert scale
9	Creative work is mainly trial and error without a precise goal and detailed plan.	5-point Likert scale
10	I find it boring to see the same old faces.	5-point Likert scale
11	I work creatively to produce something that serves a purpose.	5-point Likert scale
12	I like exploring foreign cities on my own, even at the risk of getting lost.	5-point Likert scale
13	I enjoy competing with others.	5-point Likert scale
14	I perform better when I compete against others.	5-point Likert scale
15	I generally prefer working in a group rather than alone.	5-point Likert scale
16	People in a group are generally more productive than people who work alone.	5-point Likert scale
17	People in a group should be willing to put themselves first for the good of the group.	5-point Likert scale
18	For a group to make a good decision, someone always has to take the lead.	5-point Likert scale

Note: This questionnaire is translated from German and was displayed to the ideators. The 5-point Likert scale provides five options: strongly agree, agree, neutral, disagree, strongly disagree.

Table A3: Questionnaire I Quality Assessors

No.	Question	Response Options
1	What strategy did you use when guessing?	Open-ended question (optional)
2	I first considered the two clue words individually before looking for a connection between them.	5-point Likert scale
3	I intuitively looked for associations with the clue pairs.	5-point Likert scale
4	I tried to find an overarching concept or theme that united both clue words.	5-point Likert scale
5	I consciously tried to think outside the box to discover more unusual connections.	5-point Likert scale
6	I took into account that particularly obvious words (taboo words) might be excluded.	5-point Likert scale
7	I often chose the first word that came to mind.	5-point Likert scale
8	I found it difficult to find answers based on the clue pairs.	5-point Likert scale
9	I found it difficult to decide on a final answer.	5-point Likert scale
10	I enjoyed the task.	5-point Likert scale
11	The competition with the other guessers motivated me.	5-point Likert scale
12	The bonus payment motivated me.	5-point Likert scale
13	I think that my performance on the task was better than that of an AI.	5-point Likert scale

Note: This questionnaire is translated from German and was displayed to the quality assessors. The 5-point Likert scale provides five options: strongly agree, agree, neutral, disagree, strongly disagree.

Table A4: Questionnaire II Quality Assessors

No.	Question	Response Options
1	Age	Open-ended question
2	Gender	Male/Female/Diverse
3	Field of Study	Humanities/Sports/Law, Economics, and Social Sciences/Mathematics, Natural Sciences/Human Medicine, Health Sciences/Agricultural, Forestry, and Nutritional Sciences, Veterinary Medicine/Arts, Art History/Engineering/Other Native language/Very good/Good/Okay/Rather poor/Not at all
4	How well can you read and write German?	0-10 scale
5	How would you personally assess yourself: Are you generally a risk-taker, or do you try to avoid risks?	5-point Likert scale
6	I think I am an above-average creative person.	5-point Likert scale
7	It is important to me to be a creative person.	5-point Likert scale
8	I am good at finding creative solutions to problems.	5-point Likert scale
9	Creative work is mainly trial and error without a precise goal or detailed plan.	5-point Likert scale
10	I find it boring to always see the same old faces.	5-point Likert scale
11	I work creatively to produce something that serves a purpose.	5-point Likert scale
12	I enjoy exploring foreign cities on my own, even at the risk of getting lost.	5-point Likert scale
13	I enjoy competing with others.	5-point Likert scale
14	I perform better when I compete against others.	5-point Likert scale

Note: This questionnaire is translated from German and was displayed to the quality assessors. The 5-point Likert scale provides five options: strongly agree, agree, neutral, disagree, strongly disagree.

Table A5: Descriptives Ideators

	Mean (SD)				Min	Max	p-value
	Control	Quantity	Quality	Originality			
Age	22.983 (4.744)	23.533 (4.126)	23.375 (4.770)	23.492 (3.567)	18	62	0.319
Female	0.662 (0.473)	0.617 (0.484)	0.604 (0.489)	0.704 (0.451)	0	1	0.359
Law and Econ	0.400 (0.492)	0.442 (0.499)	0.392 (0.490)	0.392 (0.490)	0	1	0.837
Humanities	0.217 (0.414)	0.183 (0.389)	0.175 (0.382)	0.233 (0.425)	0	1	0.639
STEM	0.133 (0.341)	0.158 (0.367)	0.192 (0.395)	0.183 (0.389)	0	1	0.616
Native Speaker	0.925 (0.264)	0.833 (0.374)	0.925 (0.264)	0.917 (0.278)	0	1	0.047
Risk Tolerance	5.567 (2.263)	5.533 (2.041)	5.750 (2.231)	5.467 (2.050)	0	10	0.657
Pro-sociality	3.904 (2.056)	3.618 (2.024)	3.325 (2.301)	3.629 (2.415)	0	10	0.361
Pro-sociality II	0.992 (0.804)	0.875 (0.795)	0.775 (0.939)	1.000 (0.767)	-2	2	0.151
AUT Score	2.143 (0.323)	2.158 (0.341)	2.142 (0.317)	2.100 (0.358)	1.000	2.800	0.637
Crea. Self-Assess.	-0.020 (1.059)	0.013 (1.012)	-0.072 (1.006)	0.231 (0.969)	-2.752	1.933	0.076
Competitiveness	0.038 (1.083)	0.126 (1.014)	0.115 (0.977)	-0.002 (0.999)	-1.810	1.690	0.700
Group Work	-0.125 (1.058)	-0.225 (1.111)	-0.075 (1.101)	-0.158 (1.108)	-2	2	0.716
Group Attitude	-0.075 (1.109)	-0.083 (1.089)	-0.075 (1.078)	-0.217 (1.055)	-2	2	0.642
Leadership Attitude	0.058 (1.117)	0.033 (1.076)	-0.008 (1.104)	0.100 (1.133)	-2	2	0.906
Conformity	0.042 (1.205)	0.275 (1.130)	-0.025 (1.148)	0.100 (1.226)	-2	2	0.225
Creative Style	-0.117 (1.101)	-0.125 (1.097)	-0.133 (1.092)	-0.250 (1.125)	-2	2	0.766
Novelty Seeking	-0.158 (1.145)	-0.208 (1.208)	-0.267 (1.186)	-0.350 (1.150)	-2	2	0.631
Creative Purpose	0.508 (1.077)	0.275 (1.045)	0.200 (1.089)	0.517 (1.004)	-2	2	0.030
Creative Exploration	1.167 (1.103)	0.767 (1.248)	1.117 (1.109)	1.125 (1.042)	-2	2	0.025
Obs.	120	120	120	120			

Note: The last column reports the p-values corresponding to the χ^2 statistics of a Kruskal–Wallis rank test, adjusted for ties. Detailed variable descriptions can be found in the *Appendix D: Detailed Variable Descriptions*.

Table A6: Descriptives Quality Assessors

	Mean (SD)				Min	Max	p-value
	Control	Quantity	Quality	Originality			
Age	22.575 (3.836)	23.950 (4.437)	23.000 (4.602)	23.350 (3.302)	18	47	0.232
Female	0.738 (0.438)	0.675 (0.461)	0.625 (0.490)	0.600 (0.496)	0	1	0.606
Law and Econ	0.425 (0.501)	0.375 (0.490)	0.400 (0.496)	0.250 (0.439)	0	1	0.374
Humanities	0.250 (0.439)	0.300 (0.464)	0.225 (0.423)	0.300 (0.464)	0	1	0.837
STEM	0.175 (0.385)	0.225 (0.423)	0.150 (0.362)	0.175 (0.385)	0	1	0.851
Native Speaker	0.875 (0.335)	0.875 (0.335)	0.750 (0.439)	0.825 (0.385)	0	1	0.397
Risk Tolerance	5.400 (2.240)	5.075 (2.347)	4.650 (2.237)	5.000 (2.407)	0	10	0.582
RAT	1.625 (0.838)	1.875 (0.853)	1.375 (0.628)	1.800 (0.911)	0	5	0.041
Creativity Self-Assessment	-0.273 (1.078)	-0.078 (0.837)	0.010 (0.854)	-0.117 (1.007)	-2.362	1.933	0.587
Competitiveness	-0.180 (0.799)	-0.355 (0.888)	0.071 (1.035)	-0.366 (0.904)	-1.810	1.690	0.129
Creative Style	-0.275 (0.905)	-0.250 (0.899)	-0.225 (1.121)	-0.250 (1.006)	-2	2	0.999
Novelty Seeking	-0.400 (1.033)	-0.275 (1.198)	-0.275 (1.358)	-0.525 (0.987)	-2	2	0.817
Creative Purpose	0.350 (0.921)	0.150 (0.949)	0.300 (0.992)	0.225 (1.025)	-2	2	0.801
Creative Exploration	0.675 (1.421)	0.975 (1.143)	0.550 (1.395)	1.150 (1.122)	-2	2	0.201
Strategy	-0.075 (1.248)	-0.525 (1.012)	-0.175 (1.217)	-0.300 (1.159)	-2	2	0.395
Strategy II	1.375 (0.628)	1.475 (0.679)	1.400 (0.632)	1.550 (0.597)	0	2	0.528
Strategy III	1.325 (0.764)	0.950 (0.959)	1.100 (0.709)	1.050 (1.011)	-2	2	0.275
Strategy IV	0.250 (0.899)	-0.225 (0.974)	-0.250 (0.954)	-0.400 (1.150)	-2	2	0.024
Strategy V	1.150 (0.770)	1.175 (0.813)	1.025 (0.947)	0.725 (1.132)	-2	2	0.285
Taboo Words	0.175 (1.196)	-0.050 (1.011)	0.300 (1.114)	-0.050 (1.395)	-2	2	0.527
Difficulty Generation	0.925 (0.797)	0.575 (0.958)	0.700 (0.883)	0.775 (1.050)	-2	2	0.433
Difficulty Selection	0.800 (0.939)	0.325 (1.095)	0.250 (1.214)	0.200 (1.159)	-2	2	0.073
Enjoyment	0.200 (1.043)	0.475 (0.877)	0.600 (0.982)	0.300 (1.244)	-2	2	0.249
Motivation Competition	0.125 (1.343)	-0.075 (1.269)	0.050 (1.449)	0.075 (1.328)	-2	2	0.906
Motivation Payment	0.850 (1.189)	0.525 (1.154)	0.825 (1.107)	0.700 (1.181)	-2	2	0.430
Obs.	40	40	40	40			

Note: The last column reports the p-values corresponding to the χ^2 statistics of a Kruskal–Wallis rank test, adjusted for ties. Detailed variable descriptions can be found in *Appendix D: Detailed Variable Descriptions*.

Table A7: Descriptives by Treatment

	Mean (SD)					
	Control	Quantity	Quality	Originality	Min	Max
Quality	0.900 (1.240)	1.050 (1.293)	1.158 (1.326)	1.154 (1.337)	0	3
Quantity	5.371 (2.261)	7.171 (3.838)	5.400 (2.164)	5.567 (2.319)	1	29
Quantity FB	5.138 (2.252)	6.912 (3.776)	5.071 (2.043)	5.321 (2.267)	1	28
Originality \emptyset	61.452 (5.501)	62.745 (5.240)	61.309 (5.812)	61.724 (5.959)	39.290	81.348
Originality \emptyset FB	60.475 (5.933)	61.798 (5.443)	60.186 (6.289)	60.871 (6.417)	39.290	81.348
Originality	58.956 (9.411)	60.363 (10.824)	58.936 (9.377)	59.318 (9.860)	30.503	88.571
Innovation 75	0.046 (0.210)	0.113 (0.317)	0.087 (0.283)	0.071 (0.257)	0	1
Inno Index	-0.131 (0.925)	0.066 (1.037)	0.019 (1.031)	0.046 (0.998)	-2.852	3.172
Obs.	240	240	240	240		

Table A8: Descriptives by Round

	Mean (SD)						
	1 Wind	2 Meal	3 Grain	4 Science	5 Emergency	6 Farm	p-value
Quality	1.663 (1.359)	0.719 (1.161)	0.812 (1.260)	0.944 (1.250)	0.819 (1.192)	1.438 (1.302)	<0.001
Quantity	4.194 (1.486)	4.825 (2.051)	6.044 (2.566)	7.006 (3.121)	6.231 (2.715)	6.963 (3.413)	<0.001
Quantity FB	4.050 (1.470)	4.650 (2.013)	5.737 (2.507)	6.706 (3.115)	5.912 (2.672)	6.606 (3.389)	<0.001
Originality \emptyset	59.336 (4.042)	61.138 (4.354)	64.605 (4.600)	57.252 (6.323)	66.952 (3.490)	61.563 (4.709)	<0.001
Originality \emptyset FB	58.496 (4.140)	60.548 (4.416)	63.378 (5.246)	55.539 (6.762)	66.246 (3.543)	60.788 (5.364)	<0.001
Originality	58.073 (6.566)	61.678 (8.550)	61.308 (9.010)	50.512 (11.622)	65.740 (6.759)	59.048 (8.983)	<0.001
Innovation 75	0.044 (0.205)	0.081 (0.274)	0.087 (0.283)	0.037 (0.191)	0.131 (0.339)	0.094 (0.292)	0.022
Inno Index	0.248 (0.869)	-0.027 (0.941)	-0.001 (0.947)	-0.756 (1.096)	0.345 (0.814)	0.191 (0.916)	<0.001
Obs.	160	160	160	160	160	160	

Note: The last column reports the p-values corresponding to the χ^2 statistics of a Kruskal–Wallis rank test, adjusted for ties.

Table A9: Correlation Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Quality	1.000							
(2) Quantity	-0.057*	1.000						
(3) Quantity FB	-0.058*	0.976***	1.000					
(4) Originality \emptyset	-0.161***	0.192***	0.176***	1.000				
(5) Originality \emptyset FB	-0.147***	0.185***	0.191***	0.891***	1.000			
(6) Originality	-0.139***	0.040	0.044	0.569***	0.628***	1.000		
(7) Innovation 75	0.276***	0.120***	0.115***	0.170***	0.186***	0.345***	1.000	
(8) Inno Index	0.656***	-0.013	-0.011	0.311***	0.366***	0.656***	0.473***	1.000

*** p<0.01, ** p<0.05, * p<0.1, n=960, Pearson correlations

Table A10: Treatment Effects on Effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Duration Ideagen	Duration FB	Duration Revision	Intensity Ideagen	Intensity FB	Intensity Revision
Incentive Originality	6.186 (5.585)	3.471 (2.528)	-0.227 (2.400)	-0.029 (0.058)	-0.001 (0.065)	-0.042 (0.084)
Incentive Quality	10.208* (5.903)	0.312 (2.874)	-0.364 (2.500)	-0.096* (0.052)	-0.024 (0.059)	-0.041 (0.078)
Incentive Quantity	17.713*** (5.538)	2.562 (2.617)	0.433 (2.462)	0.093 (0.062)	-0.017 (0.062)	-0.057 (0.074)
Constant	3.065 (36.164)	1.418 (18.389)	10.618 (15.801)	1.205*** (0.354)	0.330 (0.369)	0.268 (0.555)
Group Controls	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓
Obs.	960	960	960	960	960	960
R2	0.297	0.179	0.134	0.143	0.047	0.085

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A11: Quantile Regressions Originality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Originality O 50	Originality O 50	Originality O 50	Originality O 50	Originality O 75	Originality O 75	Originality O 75	Originality O 75	Originality O 90	Originality O 90	Originality O 90	Originality O 90
Incentive Originality	-0.035 (0.635)	0.562 (0.495)	0.239 (0.642)	0.861* (0.500)	0.244 (0.642)	-0.434 (0.627)	0.337 (0.642)	0.479 (0.591)	0.310 (0.707)	0.336 (0.671)	0.242 (0.581)	0.177 (0.582)
Incentive Quality	-0.035 (0.639)	0.386 (0.636)	0.402 (0.734)	0.457 (0.598)	-0.032 (0.664)	-0.281 (0.583)	0.573 (0.396)	0.785 (0.535)	0.071 (0.672)	0.608 (0.587)	0.551 (0.670)	0.362 (0.624)
Incentive Quantity	1.388** (0.592)	1.640** (0.664)	1.244* (0.721)	0.976* (0.560)	1.371** (0.693)	0.894 (0.694)	1.509** (0.660)	1.506*** (0.567)	1.484** (0.677)	1.723*** (0.608)	1.340* (0.710)	0.729 (0.717)
Quality				-0.176 (0.399)				-0.120 (0.449)				-0.268 (0.430)
Quantity				0.423*** (0.117)				0.246*** (0.078)				0.117 (0.113)
Quality × Quantity				-0.034 (0.056)				-0.052 (0.062)				-0.035 (0.063)
Constant	61.847*** (0.418)	66.150*** (4.059)	57.892*** (4.068)	57.090*** (4.061)	65.263*** (0.479)	66.089*** (3.683)	63.864*** (5.012)	65.241*** (3.817)	68.258*** (0.502)	63.194*** (3.569)	63.432*** (4.187)	59.866*** (4.990)
Group Controls		✓	✓	✓		✓	✓	✓		✓	✓	✓
Round FE												
Obs.	960	960	960	960	960	960	960	960	960	960	960	960
R2	0.009	0.034	0.357	0.386	0.010	0.027	0.337	0.360	0.010	0.021	0.318	0.341

Note: The table reports coefficients of quantile regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A12: Group Originality Final

	(1)	(2)	(3)	(4)	(5)
	Originality	Originality	Originality	Originality	Originality
Incentive Originality	0.362 (0.792)	-0.157 (0.774)	-0.157 (0.776)	0.022 (0.753)	-0.204 (0.625)
Incentive Quality	-0.020 (0.883)	-0.131 (0.868)	-0.131 (0.870)	0.169 (0.848)	0.103 (0.679)
Incentive Quantity	1.407 (0.902)	1.330 (0.892)	1.330 (0.894)	1.034 (0.827)	0.286 (0.697)
Quality				-0.699 (0.601)	-0.510 (0.508)
Quantity FB				0.271 (0.184)	-0.154 (0.146)
Quality \times Quantity FB				-0.034 (0.111)	0.008 (0.100)
Originality \emptyset FB					0.918*** (0.048)
Constant	58.956*** (0.635)	51.199*** (5.330)	49.879*** (5.359)	50.593*** (5.288)	-4.380 (5.234)
Group Controls		✓	✓	✓	✓
Round FE			✓	✓	✓
Obs.	960	960	960	960	960
R2	0.003	0.022	0.243	0.260	0.452

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A13: Group Originality Taboo Words

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Originality \emptyset	Originality \emptyset	Originality \emptyset	Originality \emptyset	Originality	Originality	Originality	Originality
Incentive Originality	0.209 (0.379)	0.202 (0.343)	0.202 (0.344)	0.183 (0.348)	0.123 (0.599)	0.149 (0.611)	0.149 (0.613)	0.147 (0.613)
Incentive Quality	0.053 (0.394)	0.242 (0.387)	0.242 (0.388)	0.248 (0.370)	-0.033 (0.701)	-0.116 (0.725)	-0.116 (0.727)	-0.061 (0.739)
Incentive Quantity	0.811* (0.415)	1.050*** (0.382)	1.050*** (0.383)	0.630* (0.372)	1.063 (0.704)	1.139 (0.720)	1.139 (0.722)	0.513 (0.691)
Quality				-0.013 (0.211)				0.017 (0.443)
Quantity				0.233*** (0.062)				0.390*** (0.132)
Quality \times Quantity				-0.018 (0.032)				-0.056 (0.074)
Constant	69.190*** (0.275)	69.195*** (2.274)	72.425*** (2.281)	72.878*** (2.201)	68.055*** (0.510)	63.278*** (4.412)	66.333*** (4.461)	67.053*** (4.545)
Group Controls		✓	✓	✓		✓	✓	✓
Round FE								
Obs.	960	960	960	960	960	960	960	960
R2	0.005	0.042	0.426	0.442	0.003	0.022	0.205	0.220

Note: The table reports coefficients of OLS (standard errors in parentheses, clustered at the group level). All variables are at the group level. Originality refers to the semantic distance between the final clue pair and the provided taboo words. Originality \emptyset refers to the average semantic distance between all generated clue pairs and the taboo words. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A14: Quality Determinants

	(1)	(2)	(3)	(4)
	Quality	Quality	Quality	Quality
	OLS	OLS	OLogit	OLogit
Age	-0.018*	-0.018*	-0.033**	-0.035**
	(0.010)	(0.010)	(0.015)	(0.016)
Female	-0.036	-0.036	-0.040	-0.048
	(0.081)	(0.081)	(0.129)	(0.139)
Law and Econ	0.092	0.092	0.190	0.203
	(0.096)	(0.096)	(0.159)	(0.171)
Humanities	0.165	0.165	0.300*	0.316*
	(0.101)	(0.102)	(0.161)	(0.172)
STEM	0.032	0.032	0.133	0.117
	(0.130)	(0.130)	(0.212)	(0.223)
Native Speaker	0.161*	0.161*	0.378***	0.408***
	(0.086)	(0.086)	(0.145)	(0.156)
Risk Tolerance	-0.003	-0.003	-0.005	-0.009
	(0.013)	(0.013)	(0.022)	(0.023)
RAT	0.011	0.011	0.019	0.009
	(0.045)	(0.046)	(0.071)	(0.075)
Creativity Self-Assessment	0.008	0.008	0.012	0.014
	(0.040)	(0.040)	(0.064)	(0.068)
Competitiveness	-0.033	-0.033	-0.048	-0.060
	(0.050)	(0.050)	(0.076)	(0.081)
Creative Style	-0.089***	-0.089***	-0.140***	-0.150***
	(0.030)	(0.030)	(0.049)	(0.053)
Novelty Seeking	-0.016	-0.016	-0.011	-0.015
	(0.031)	(0.031)	(0.048)	(0.052)
Creative Purpose	0.028	0.028	0.061	0.072
	(0.034)	(0.034)	(0.052)	(0.055)
Creative Exploration	-0.032	-0.032	-0.056	-0.059
	(0.023)	(0.023)	(0.038)	(0.041)
Strategy	0.045	0.045	0.082*	0.090*
	(0.031)	(0.031)	(0.047)	(0.050)
Strategy II	0.077	0.077	0.115	0.128
	(0.060)	(0.060)	(0.094)	(0.100)
Strategy III	-0.073	-0.073	-0.114*	-0.110
	(0.046)	(0.046)	(0.069)	(0.075)
Strategy IV	-0.012	-0.012	-0.030	-0.027
	(0.037)	(0.037)	(0.057)	(0.061)
Strategy V	-0.027	-0.027	-0.045	-0.049
	(0.038)	(0.038)	(0.058)	(0.063)
Taboo Words	-0.057**	-0.057**	-0.077*	-0.079*
	(0.028)	(0.029)	(0.043)	(0.046)
Difficulty Gen	-0.291***	-0.291***	-0.459***	-0.490***
	(0.041)	(0.041)	(0.064)	(0.069)
Difficulty Select	-0.022	-0.022	-0.036	-0.038
	(0.032)	(0.032)	(0.050)	(0.053)
Enjoyment	0.026	0.026	0.022	0.036
	(0.037)	(0.037)	(0.058)	(0.062)
Motivation Competition	0.052	0.052	0.087	0.087
	(0.034)	(0.034)	(0.054)	(0.057)
Motivation Payment	0.079**	0.079**	0.115**	0.128**
	(0.034)	(0.034)	(0.053)	(0.057)
Constant	1.436***	2.033***		
	(0.290)	(0.312)		
Round FE		✓		✓
Obs.	960	960	960	960
R2	0.091	0.166		
Pseudo R2			0.044	0.082

Note: The table reports coefficients of OLS and ordered logit regressions (standard errors in parentheses, clustered at the individual level). Ordered logit coefficients are reported as logits, not as odds ratios or marginal effects. Cutpoints (thresholds) exist in the model but are omitted from this table for clarity. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A15: Difficulty Quality Assessors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Difficulty Gen	Difficulty Gen	Difficulty Gen	Difficulty Gen	Difficulty Gen	Difficulty Select	Difficulty Select	Difficulty Select	Difficulty Select	Difficulty Select
	OLS	OLS	OLS	OLogit	OLogit	OLS	OLS	OLS	OLogit	OLogit
Incentive Originality	-0.150 (0.207)	-0.122 (0.215)	0.041 (0.211)	-0.201 (0.412)	0.239 (0.459)	-0.600** (0.247)	-0.630** (0.244)	-0.589** (0.235)	-0.951** (0.406)	-1.066** (0.446)
Incentive Quality	-0.225 (0.207)	-0.305 (0.213)	-0.102 (0.211)	-0.399 (0.400)	-0.283 (0.440)	-0.550** (0.247)	-0.786*** (0.242)	-0.684*** (0.234)	-0.830** (0.410)	-1.369*** (0.459)
Incentive Quantity	-0.350* (0.207)	-0.337 (0.210)	-0.223 (0.203)	-0.642 (0.407)	-0.502 (0.446)	-0.475* (0.247)	-0.439* (0.239)	-0.326 (0.231)	-0.717* (0.401)	-0.592 (0.425)
Difficulty Select			0.258*** (0.071)		0.597*** (0.159)					
Difficulty Gen								0.335*** (0.092)		0.670*** (0.182)
Constant	0.925*** (0.147)	1.465** (0.613)	1.133* (0.595)			0.800*** (0.175)	1.283* (0.698)	0.792 (0.683)		
Guesser Controls		✓	✓		✓		✓	✓		✓
Obs.	160	160	160	160	160	160	160	160	160	160
R2	0.019	0.145	0.219			0.046	0.244	0.309		
Pseudo R2				0.007	0.104				0.015	0.130

Note: The table reports coefficients of OLS and ordered logit regressions (standard errors in parentheses). Ordered logit coefficients are reported as logits, not as odds ratios or marginal effects. Cutpoints (thresholds) exist in the model but are omitted from this table for clarity. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A16: Quantile Regressions Quantity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Quantity 50	Quantity 50	Quantity 50	Quantity 50	Quantity 75	Quantity 75	Quantity 75	Quantity 75	Quantity 90	Quantity 90	Quantity 90	Quantity 90
Incentive Originality	1.000*** (0.293)	0.411* (0.233)	0.328 (0.200)	0.355* (0.187)	1.000*** (0.207)	0.593** (0.297)	0.439* (0.235)	0.272 (0.221)	1.000** (0.391)	0.742* (0.423)	0.722* (0.396)	0.766** (0.329)
Incentive Quality	1.000*** (0.249)	0.183 (0.239)	0.139 (0.238)	0.198 (0.218)	1.000*** (0.234)	0.518 (0.368)	0.288 (0.226)	0.206 (0.295)	1.000*** (0.256)	0.467 (0.311)	0.398 (0.340)	0.286 (0.268)
Incentive Quantity	1.000*** (0.225)	0.822*** (0.273)	0.522** (0.214)	0.559*** (0.203)	1.000*** (0.263)	0.893** (0.440)	0.816*** (0.304)	0.756** (0.353)	2.000*** (0.405)	1.477*** (0.386)	1.414** (0.570)	1.365*** (0.447)
Quality				-0.015 (0.415)				-0.178 (0.659)				0.702 (1.717)
Originality \emptyset				0.030** (0.014)				0.023 (0.020)				-0.011 (0.034)
Quality \times Originality \emptyset				-0.000 (0.007)				0.003 (0.011)				-0.011 (0.028)
Constant	1.000*** (0.123)	-1.108 (1.848)	-1.016 (1.692)	-2.597 (1.741)	2.000*** (0.140)	-0.330 (1.432)	-1.625 (1.313)	-2.802 (1.844)	3.000*** (0.198)	-1.031 (1.757)	-1.564 (2.787)	-0.320 (3.072)
Group Controls		✓	✓	✓		✓	✓	✓		✓	✓	✓
Round FE												
Obs.	960	960	960	960	960	960	960	960	960	960	960	960
R2	0.015	0.075	0.135	0.132	0.015	0.069	0.129	0.134	0.034	0.082	0.131	0.132

Note: The table reports coefficients of quantile regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A17: Individual Originality

	before feedback and revision				after feedback and revision			
	(1) Originality	(2) Originality	(3) Originality	(4) Originality	(5) Originality FB	(6) Originality FB	(7) Originality FB	(8) Originality FB
Incentive Originality	0.178 (0.485)	0.241 (0.468)	0.257 (0.464)	0.139 (0.451)	0.268 (0.472)	0.278 (0.463)	0.289 (0.457)	0.177 (0.437)
Incentive Quality	-0.083 (0.496)	0.092 (0.474)	0.104 (0.472)	0.073 (0.454)	-0.291 (0.495)	-0.259 (0.489)	-0.229 (0.487)	-0.197 (0.460)
Incentive Quantity	1.302*** (0.488)	1.377*** (0.480)	1.391*** (0.481)	0.665 (0.476)	1.254*** (0.477)	1.247*** (0.470)	1.270*** (0.469)	0.418 (0.450)
Quantity				1.122*** (0.111)				
Quantity FB								1.370*** (0.117)
Constant	61.404*** (0.335)	62.689*** (1.871)	60.358*** (1.867)	59.469*** (1.869)	60.481*** (0.331)	61.768*** (1.855)	59.709*** (1.853)	58.762*** (1.814)
Individual Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	2,739	2,716	2,716	2,716	2,713	2,690	2,690	2,690
R2	0.005	0.029	0.185	0.214	0.005	0.020	0.190	0.231

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A18: Individual Quantity

	before feedback and revision				after feedback and revision			
	(1) Quantity	(2) Quantity	(3) Quantity	(4) Quantity	(5) Quantity FB	(6) Quantity FB	(7) Quantity FB	(8) Quantity FB
Incentive Originality	0.065 (0.106)	0.084 (0.104)	0.084 (0.104)	0.097 (0.101)	0.061 (0.103)	0.074 (0.100)	0.074 (0.100)	0.071 (0.097)
Incentive Quality	0.010 (0.106)	0.034 (0.104)	0.034 (0.104)	0.024 (0.100)	-0.022 (0.101)	-0.003 (0.099)	-0.003 (0.099)	-0.014 (0.095)
Incentive Quantity	0.600*** (0.139)	0.642*** (0.142)	0.642*** (0.142)	0.602*** (0.140)	0.592*** (0.136)	0.624*** (0.139)	0.624*** (0.139)	0.575*** (0.135)
Originality				0.032*** (0.004)				
Originality FB								0.037*** (0.004)
Constant	1.790*** (0.077)	1.181*** (0.401)	0.623 (0.398)	-1.140** (0.453)	1.712*** (0.075)	1.035*** (0.396)	0.517 (0.392)	-1.524*** (0.440)
Individual Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	2,880	2,856	2,856	2,716	2,880	2,856	2,856	2,690
R2	0.033	0.071	0.135	0.172	0.035	0.070	0.127	0.179

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A19: Quantile Regressions Innovation Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Inno Index 50	Inno Index 50	Inno Index 50	Inno Index 75	Inno Index 75	Inno Index 75	Inno Index 90	Inno Index 90	Inno Index 90
Incentive Originality	0.163 (0.102)	0.237** (0.103)	0.151 (0.105)	0.254* (0.137)	0.146 (0.139)	0.249** (0.111)	0.276** (0.125)	0.146 (0.123)	0.098 (0.123)
Incentive Quality	0.140 (0.109)	0.169 (0.108)	0.185* (0.112)	0.274** (0.137)	0.247 (0.156)	0.307** (0.135)	0.340** (0.135)	0.383** (0.157)	0.293* (0.155)
Incentive Quantity	0.177 (0.130)	0.134 (0.110)	0.128 (0.119)	0.347*** (0.127)	0.289* (0.156)	0.403*** (0.117)	0.340** (0.165)	0.389** (0.155)	0.374*** (0.134)
Constant	-0.116 (0.076)	-1.419** (0.695)	-1.561** (0.758)	0.440*** (0.068)	-2.271** (0.916)	-1.513** (0.647)	1.102*** (0.098)	-0.769 (0.847)	-0.472 (0.840)
Group Controls		✓	✓		✓	✓		✓	✓
Round FE									
Obs.	960	960	960	960	960	960	960	960	960
R2	0.006	0.022	0.154	0.006	0.025	0.137	0.005	0.024	0.122

Note: The table reports coefficients of quantile regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A20: Pooled Performance Pay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quantity	Quantity FB	Originality \emptyset	Originality \emptyset FB	Originality	Innovation 75	Inno Index
Performance Pay	0.274*** (0.082)	0.770** (0.312)	0.719** (0.303)	0.607 (0.378)	0.532 (0.415)	0.300 (0.696)	0.046** (0.020)	0.183*** (0.063)
Constant	0.249 (0.657)	-1.736 (2.696)	-2.018 (2.592)	60.438*** (3.490)	58.944*** (3.601)	50.141*** (5.600)	-0.028 (0.194)	-1.192** (0.485)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.100	0.230	0.213	0.357	0.346	0.240	0.034	0.163

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Performance pay is a treatment dummy.

Table A21: Quantity and Originality

	(1)	(2)	(3)
	Originality \emptyset	Originality \emptyset	Originality \emptyset
Number of Pairs	1.543*** (0.115)	1.508*** (0.120)	1.498*** (0.120)
Constant	59.762*** (0.259)	62.819*** (1.846)	60.134*** (1.821)
Individual Controls		✓	✓
Round FE			✓
Obs.	5,642	5,577	5,577
R2	0.042	0.052	0.136

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A22: Group Originality Process

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Originality O	Originality O FB	Originality	Originality Removed	Originality Change	Originality Change FB	Max Originality	Max Originality FB
Incentive Originality	0.221 (0.436)	0.218 (0.443)	-0.157 (0.776)	2.312 (1.921)	-0.306 (0.827)	-0.604 (0.893)	1.326* (0.716)	1.301* (0.743)
Incentive Quality	0.155 (0.502)	-0.058 (0.558)	-0.131 (0.870)	1.574 (1.675)	0.066 (0.831)	0.045 (1.001)	0.759 (0.763)	0.680 (0.755)
Incentive Quantity	1.572*** (0.505)	1.556*** (0.541)	1.330 (0.894)	5.073*** (1.701)	1.709* (0.952)	1.493 (1.106)	3.626*** (0.819)	3.894*** (0.827)
Constant	60.240*** (3.243)	58.842*** (3.362)	49.879*** (5.359)	54.007*** (12.131)	60.072*** (5.728)	53.984*** (6.129)	65.935*** (5.487)	61.915*** (5.306)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	960	960	960	228	741	693	960	960
R2	0.366	0.355	0.243	0.158	0.163	0.168	0.213	0.205

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A23: Group Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quantity	Quantity FB	Originality \emptyset	Originality \emptyset FB	Originality	Innovation 75	Inno Index
Age	0.018 (0.016)	0.060 (0.048)	0.059 (0.046)	-0.008 (0.062)	0.025 (0.064)	0.108 (0.102)	-0.001 (0.003)	0.019 (0.012)
Female	-0.201 (0.169)	-0.638 (0.564)	-0.610 (0.551)	-1.104 (0.712)	-0.995 (0.803)	0.311 (1.268)	-0.017 (0.039)	-0.094 (0.110)
Law and Econ	-0.103 (0.158)	-0.119 (0.626)	0.186 (0.606)	0.837 (0.859)	1.201 (0.872)	1.325 (1.423)	0.046 (0.039)	0.042 (0.120)
Humanities	0.093 (0.253)	-0.717 (0.737)	-0.168 (0.713)	-0.623 (0.971)	-0.385 (0.975)	0.228 (1.876)	-0.048 (0.045)	0.072 (0.177)
STEM	-0.142 (0.210)	-0.872 (0.826)	-0.684 (0.783)	-0.162 (1.113)	0.191 (1.127)	3.155 (1.948)	0.066 (0.058)	0.160 (0.168)
Native Speaker	0.210 (0.225)	-0.232 (0.986)	-0.312 (0.944)	-3.478*** (1.063)	-2.800*** (1.006)	-1.424 (1.534)	-0.036 (0.077)	0.013 (0.171)
Risk Tolerance	-0.058 (0.036)	-0.027 (0.140)	-0.007 (0.137)	0.024 (0.157)	-0.232 (0.172)	-0.210 (0.245)	-0.011 (0.009)	-0.050** (0.023)
Pro-sociality	0.026 (0.035)	0.040 (0.124)	0.061 (0.120)	0.212 (0.147)	0.190 (0.160)	0.018 (0.236)	-0.004 (0.008)	0.017 (0.024)
Pro-sociality II	0.007 (0.099)	1.087*** (0.364)	1.094*** (0.354)	1.017** (0.416)	0.845* (0.445)	1.260** (0.592)	0.010 (0.021)	0.101 (0.063)
AUT	0.431** (0.207)	1.844** (0.917)	1.759** (0.887)	-0.122 (1.089)	0.209 (1.138)	2.185 (1.808)	0.053 (0.058)	0.421*** (0.148)
Creativity Self-Assessment	0.161* (0.097)	0.324 (0.329)	0.265 (0.320)	0.271 (0.366)	0.636 (0.434)	1.092* (0.640)	0.013 (0.020)	0.179*** (0.068)
Competitiveness	0.045 (0.085)	-0.058 (0.274)	0.022 (0.265)	-0.086 (0.338)	0.213 (0.313)	0.710 (0.554)	0.005 (0.019)	0.081 (0.062)
Group Work	0.060 (0.059)	0.238 (0.251)	0.159 (0.245)	0.324 (0.253)	0.520* (0.281)	-0.229 (0.463)	0.000 (0.017)	0.017 (0.041)
Group Attitude	-0.016 (0.068)	0.184 (0.301)	0.196 (0.296)	-0.169 (0.334)	-0.225 (0.357)	0.155 (0.591)	0.018 (0.014)	0.003 (0.052)
Leadership Attitude	-0.061 (0.061)	-0.104 (0.236)	-0.186 (0.227)	-0.033 (0.306)	-0.354 (0.297)	-0.158 (0.491)	-0.001 (0.015)	-0.048 (0.039)
Conformity	0.007 (0.062)	-0.105 (0.198)	-0.090 (0.190)	-0.238 (0.257)	-0.276 (0.255)	-0.274 (0.449)	0.006 (0.017)	-0.017 (0.049)
Creative Style	0.003 (0.066)	-0.307 (0.257)	-0.267 (0.253)	-0.024 (0.293)	0.018 (0.302)	0.102 (0.555)	0.001 (0.016)	0.010 (0.052)
Novelty Seeking	-0.007 (0.071)	-0.071 (0.226)	-0.128 (0.218)	-0.485 (0.337)	-0.562 (0.360)	-0.761 (0.499)	-0.005 (0.017)	-0.063 (0.052)
Creative Purpose	0.146* (0.076)	-0.489* (0.265)	-0.416 (0.256)	0.005 (0.326)	0.093 (0.331)	-0.685 (0.503)	0.000 (0.017)	0.032 (0.052)
Creative Exploration	-0.039 (0.059)	-0.024 (0.227)	-0.094 (0.218)	0.479* (0.284)	0.409 (0.296)	0.364 (0.440)	0.018 (0.014)	0.005 (0.047)
Constant	-0.072 (0.690)	0.723 (2.694)	0.268 (2.603)	63.521*** (3.396)	61.816*** (3.547)	51.763*** (5.468)	0.054 (0.185)	-1.255*** (0.473)
Obs.	960	960	960	960	960	960	960	960
R2	0.019	0.081	0.079	0.034	0.029	0.018	0.016	0.026

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Mediation Originality before FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Originality \emptyset	Incentive Quantity	Share of changed pairs	-0.039	0.060	0.517
Originality \emptyset	Incentive Quantity	Share of removed pairs	0.012	0.053	0.816
Originality \emptyset	Incentive Quantity	Share of given positive ratings	0.051	0.061	0.402
Originality \emptyset	Incentive Quantity	Random Vote	0.006	0.015	0.694
Originality \emptyset	Incentive Quantity	Unanimous Vote	-0.003	0.016	0.844
Originality \emptyset	Incentive Quantity	Share of used suggestions	0.044	0.048	0.368
Originality \emptyset	Incentive Quantity	Number of suggestions	0.016	0.044	0.723
Originality \emptyset	Incentive Quantity	Teamwork	0.050	0.072	0.493
Originality \emptyset	Incentive Quantity	Leadership	-0.001	0.053	0.991
Originality \emptyset	Incentive Quantity	Feedback Assessment	-0.023	0.053	0.664
Originality \emptyset	Incentive Quantity	Chat Sentiment	-0.096	0.069	0.166
Originality \emptyset	Incentive Quantity	Number Chat Messages	-0.176	0.130	0.175
Originality \emptyset	Incentive Quantity	Group Effort	0.002	0.012	0.869
Originality \emptyset	Incentive Quantity	Peer Pressure	0.123	0.139	0.375
Originality \emptyset	Incentive Quantity	Enjoyment	0.129	0.110	0.242
Originality \emptyset	Incentive Quantity	Communication	0.039	0.060	0.522
Originality \emptyset	Incentive Quantity	Difficulty Quality	0.007	0.062	0.907
Originality \emptyset	Incentive Quantity	Difficulty Originality	-0.051	0.081	0.530
Originality \emptyset	Incentive Quantity	Difficulty Quantity	0.033	0.102	0.745
Originality \emptyset	Incentive Quantity	Quantity	0.465	0.164	0.005
Originality \emptyset	Incentive Quantity	Duration Ideagen	0.372	0.173	0.032
Originality \emptyset	Incentive Quantity	Duration FB	0.027	0.045	0.545
Originality \emptyset	Incentive Quantity	Duration Revision	0.015	0.044	0.740
Originality \emptyset	Incentive Quantity	Intensity Ideagen	0.039	0.063	0.529
Originality \emptyset	Incentive Quantity	Intensity FB	-0.007	0.043	0.872
Originality \emptyset	Incentive Quantity	Intensity Revision	-0.020	0.053	0.710
No control group					
Originality \emptyset	Incentive Quantity	Tradeoff	-0.051	0.100	0.606
Originality \emptyset	Incentive Quantity	Self-efficacy	0.330	0.264	0.210
Originality \emptyset	Incentive Quantity	Pro-social Group	0.024	0.069	0.731
Originality \emptyset	Incentive Quantity	Motivation Competition	0.227	0.192	0.239
Originality \emptyset	Incentive Quantity	Motivation Payment	-0.021	0.103	0.835

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A25: Mediation Individual Originality before FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Originality	Incentive Quantity	Share of changed pairs	-0.121	0.069	0.081
Originality	Incentive Quantity	Share of removed pairs	-0.027	0.044	0.545
Originality	Incentive Quantity	Share of given positive ratings	0.000	0.006	0.951
Originality	Incentive Quantity	Share of used suggestions	0.004	0.018	0.818
Originality	Incentive Quantity	Number of suggestions	-0.016	0.040	0.687
Originality	Incentive Quantity	Group Effort	-0.002	0.015	0.878
Originality	Incentive Quantity	Peer Pressure	-0.126	0.104	0.225
Originality	Incentive Quantity	Teamwork	0.013	0.026	0.627
Originality	Incentive Quantity	Leadership	-0.001	0.009	0.910
Originality	Incentive Quantity	Feedback Assessment	0.007	0.021	0.734
Originality	Incentive Quantity	Difficulty Quality	0.008	0.025	0.762
Originality	Incentive Quantity	Difficulty Originality	-0.043	0.089	0.625
Originality	Incentive Quantity	Difficulty Quantity	0.026	0.055	0.635
Originality	Incentive Quantity	Enjoyment	0.032	0.050	0.526
Originality	Incentive Quantity	Quantity	0.464	0.180	0.010
Originality	Incentive Quantity	Duration Ideagen	0.241	0.130	0.065
Originality	Incentive Quantity	Duration FB	0.000	0.003	0.969
Originality	Incentive Quantity	Duration Revision	-0.023	0.074	0.761
Originality	Incentive Quantity	Intensity Ideagen	0.060	0.052	0.251
Originality	Incentive Quantity	Intensity FB	-0.005	0.018	0.765
Originality	Incentive Quantity	Intensity Revision	-0.045	0.042	0.283
No control group					
Originality	Incentive Quantity	Motivation Competition	-0.002	0.095	0.984
Originality	Incentive Quantity	Motivation Payment	-0.076	0.075	0.312
Originality	Incentive Quantity	Tradeoff	-0.045	0.066	0.491
Originality	Incentive Quantity	Self-efficacy	0.264	0.173	0.127

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the individual level).

Table A26: Mediation Individual Originality after FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Originality FB	Incentive Quantity	Share of changed pairs	0.077	0.056	0.164
Originality FB	Incentive Quantity	Share of removed pairs	0.004	0.013	0.753
Originality FB	Incentive Quantity	Share of given positive ratings	-0.006	0.016	0.700
Originality FB	Incentive Quantity	Share of used suggestions	0.008	0.022	0.711
Originality FB	Incentive Quantity	Number of suggestions	-0.012	0.031	0.698
Originality FB	Incentive Quantity	Group Effort	0.002	0.012	0.834
Originality FB	Incentive Quantity	Peer Pressure	-0.059	0.066	0.372
Originality FB	Incentive Quantity	Teamwork	0.017	0.030	0.576
Originality FB	Incentive Quantity	Leadership	-0.001	0.006	0.919
Originality FB	Incentive Quantity	Feedback Assessment	-0.003	0.012	0.797
Originality FB	Incentive Quantity	Difficulty Quality	0.007	0.026	0.786
Originality FB	Incentive Quantity	Difficulty Originality	-0.038	0.069	0.582
Originality FB	Incentive Quantity	Difficulty Quantity	0.035	0.074	0.640
Originality FB	Incentive Quantity	Enjoyment	0.026	0.046	0.563
Originality FB	Incentive Quantity	Quantity FB	0.573	0.197	0.004
Originality FB	Incentive Quantity	Duration Ideagen	0.286	0.138	0.037
Originality FB	Incentive Quantity	Duration FB	-0.002	0.019	0.904
Originality FB	Incentive Quantity	Duration Revision	-0.008	0.027	0.754
Originality FB	Incentive Quantity	Intensity Ideagen	0.073	0.055	0.187
Originality FB	Incentive Quantity	Intensity FB	-0.007	0.017	0.683
Originality FB	Incentive Quantity	Intensity Revision	0.001	0.018	0.964
No control group					
Originality FB	Incentive Quantity	Motivation Competition	0.045	0.087	0.604
Originality FB	Incentive Quantity	Motivation Payment	-0.015	0.051	0.763
Originality FB	Incentive Quantity	Tradeoff	-0.041	0.060	0.495
Originality FB	Incentive Quantity	Self-efficacy	0.294	0.158	0.063

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the individual level).

Table A27: Mediation Quantity before FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quantity	Incentive Quantity	Share of changed pairs	0.029	0.029	0.319
Quantity	Incentive Quantity	Share of removed pairs	0.001	0.005	0.798
Quantity	Incentive Quantity	Share of given positive ratings	0.055	0.065	0.399
Quantity	Incentive Quantity	Random Vote	0.007	0.010	0.483
Quantity	Incentive Quantity	Unanimous Vote	-0.001	0.007	0.843
Quantity	Incentive Quantity	Share of used suggestions	-0.033	0.034	0.332
Quantity	Incentive Quantity	Number of suggestions	0.009	0.024	0.717
Quantity	Incentive Quantity	Teamwork	0.024	0.035	0.484
Quantity	Incentive Quantity	Leadership	0.000	0.022	0.991
Quantity	Incentive Quantity	Feedback Assessment	-0.022	0.049	0.652
Quantity	Incentive Quantity	Chat Sentiment	0.018	0.026	0.491
Quantity	Incentive Quantity	Number Chat Messages	-0.122	0.079	0.123
Quantity	Incentive Quantity	Group Effort	-0.007	0.031	0.812
Quantity	Incentive Quantity	Peer Pressure	0.012	0.097	0.901
Quantity	Incentive Quantity	Enjoyment	0.049	0.059	0.411
Quantity	Incentive Quantity	Communication	-0.016	0.039	0.686
Quantity	Incentive Quantity	Difficulty Quality	0.007	0.059	0.906
Quantity	Incentive Quantity	Difficulty Originality	-0.020	0.035	0.562
Quantity	Incentive Quantity	Difficulty Quantity	0.049	0.147	0.738
Quantity	Incentive Quantity	Duration Ideagen	0.387	0.107	0.000
Quantity	Incentive Quantity	Duration FB	0.065	0.095	0.493
Quantity	Incentive Quantity	Duration Revision	0.027	0.078	0.732
Quantity	Incentive Quantity	Intensity Ideagen	0.334	0.101	0.001
Quantity	Incentive Quantity	Intensity FB	-0.001	0.008	0.879
Quantity	Incentive Quantity	Intensity Revision	-0.017	0.049	0.724
No control group					
Quantity	Incentive Quantity	Self-efficacy	0.417	0.179	0.020
Quantity	Incentive Quantity	Pro-social Group	0.020	0.048	0.681
Quantity	Incentive Quantity	Motivation Competition	0.185	0.091	0.044
Quantity	Incentive Quantity	Motivation Payment	-0.045	0.073	0.534

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A28: Mediation Individual Quantity before FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quantity	Incentive Quantity	Share of changed pairs	0.051	0.020	0.009
Quantity	Incentive Quantity	Share of removed pairs	-0.003	0.007	0.631
Quantity	Incentive Quantity	Share of given positive ratings	0.000	0.002	0.893
Quantity	Incentive Quantity	Share of used suggestions	-0.023	0.050	0.646
Quantity	Incentive Quantity	Number of suggestions	-0.010	0.032	0.746
Quantity	Incentive Quantity	Group Effort	0.002	0.031	0.949
Quantity	Incentive Quantity	Peer Pressure	0.057	0.051	0.265
Quantity	Incentive Quantity	Teamwork	0.012	0.021	0.580
Quantity	Incentive Quantity	Leadership	-0.005	0.020	0.783
Quantity	Incentive Quantity	Feedback Assessment	-0.012	0.027	0.667
Quantity	Incentive Quantity	Difficulty Quality	0.002	0.009	0.807
Quantity	Incentive Quantity	Difficulty Originality	0.002	0.011	0.861
Quantity	Incentive Quantity	Difficulty Quantity	0.050	0.094	0.597
Quantity	Incentive Quantity	Enjoyment	0.017	0.037	0.640
Quantity	Incentive Quantity	Duration Ideagen	0.294	0.099	0.003
Quantity	Incentive Quantity	Duration FB	0.002	0.006	0.772
Quantity	Incentive Quantity	Duration Revision	-0.021	0.065	0.754
Quantity	Incentive Quantity	Intensity Ideagen	0.314	0.087	0.000
Quantity	Incentive Quantity	Intensity FB	-0.005	0.011	0.692
Quantity	Incentive Quantity	Intensity Revision	-0.028	0.026	0.281
No control group					
Quantity	Incentive Quantity	Motivation Competition	0.027	0.066	0.689
Quantity	Incentive Quantity	Motivation Payment	-0.024	0.040	0.543
Quantity	Incentive Quantity	Tradeoff	-0.095	0.070	0.174
Quantity	Incentive Quantity	Self-efficacy	0.255	0.116	0.028

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the individual level).

Table A29: Mediation Individual Quantity after FB

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quantity FB	Incentive Quantity	Share of changed pairs	0.075	0.025	0.002
Quantity FB	Incentive Quantity	Share of removed pairs	0.011	0.017	0.504
Quantity FB	Incentive Quantity	Share of given positive ratings	0.000	0.002	0.883
Quantity FB	Incentive Quantity	Share of used suggestions	-0.027	0.059	0.645
Quantity FB	Incentive Quantity	Number of suggestions	-0.009	0.028	0.745
Quantity FB	Incentive Quantity	Group Effort	0.002	0.035	0.949
Quantity FB	Incentive Quantity	Peer Pressure	0.052	0.049	0.291
Quantity FB	Incentive Quantity	Teamwork	0.008	0.018	0.646
Quantity FB	Incentive Quantity	Leadership	-0.004	0.015	0.785
Quantity FB	Incentive Quantity	Feedback Assessment	-0.010	0.025	0.671
Quantity FB	Incentive Quantity	Difficulty Quality	0.002	0.009	0.804
Quantity FB	Incentive Quantity	Difficulty Originality	0.002	0.011	0.869
Quantity FB	Incentive Quantity	Difficulty Quantity	0.049	0.093	0.597
Quantity FB	Incentive Quantity	Enjoyment	0.014	0.036	0.704
Quantity FB	Incentive Quantity	Duration Ideagen	0.283	0.096	0.003
Quantity FB	Incentive Quantity	Duration FB	0.002	0.006	0.767
Quantity FB	Incentive Quantity	Duration Revision	-0.017	0.055	0.753
Quantity FB	Incentive Quantity	Intensity Ideagen	0.308	0.086	0.000
Quantity FB	Incentive Quantity	Intensity FB	-0.004	0.011	0.693
Quantity FB	Incentive Quantity	Intensity Revision	-0.016	0.016	0.314
No control group					
Quantity FB	Incentive Quantity	Motivation Competition	0.027	0.065	0.674
Quantity FB	Incentive Quantity	Motivation Payment	-0.017	0.037	0.641
Quantity FB	Incentive Quantity	Tradeoff	-0.096	0.069	0.163
Quantity FB	Incentive Quantity	Self-efficacy	0.220	0.110	0.045

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the individual level).

Table A30: Mediation Quality Quality-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quality	Incentive Quality	Share of changed pairs	0.004	0.032	0.898
Quality	Incentive Quality	Share of removed pairs	-0.010	0.043	0.825
Quality	Incentive Quality	Share of given positive ratings	-0.002	0.015	0.876
Quality	Incentive Quality	Random Vote	-0.030	0.043	0.483
Quality	Incentive Quality	Unanimous Vote	-0.020	0.042	0.640
Quality	Incentive Quality	Share of used suggestions	-0.009	0.078	0.911
Quality	Incentive Quality	Number of suggestions	0.001	0.018	0.942
Quality	Incentive Quality	Teamwork	0.039	0.102	0.702
Quality	Incentive Quality	Leadership	0.039	0.115	0.737
Quality	Incentive Quality	Feedback Assessment	0.000	0.009	0.972
Quality	Incentive Quality	Chat Sentiment	-0.066	0.094	0.480
Quality	Incentive Quality	Number Chat Messages	-0.018	0.059	0.765
Quality	Incentive Quality	Group Effort	-0.002	0.026	0.939
Quality	Incentive Quality	Peer Pressure	0.158	0.214	0.459
Quality	Incentive Quality	Enjoyment	0.132	0.155	0.395
Quality	Incentive Quality	Communication	-0.014	0.049	0.769
Quality	Incentive Quality	Difficulty Quality	-0.119	0.171	0.484
Quality	Incentive Quality	Difficulty Originality	-0.069	0.102	0.499
Quality	Incentive Quality	Difficulty Quantity	0.030	0.164	0.854
Quality	Incentive Quality	Duration Ideagen	-0.035	0.095	0.710
Quality	Incentive Quality	Duration FB	0.015	0.045	0.746
Quality	Incentive Quality	Duration Revision	0.010	0.044	0.824
Quality	Incentive Quality	Intensity Ideagen	-0.163	0.166	0.325
Quality	Incentive Quality	Intensity FB	-0.011	0.034	0.735
Quality	Incentive Quality	Intensity Revision	0.005	0.045	0.907
No control group					
Quality	Incentive Quality	Self-efficacy	-0.038	0.358	0.915
Quality	Incentive Quality	Pro-social Group	0.004	0.233	0.985
Quality	Incentive Quality	Motivation Competition	0.178	0.316	0.574
Quality	Incentive Quality	Motivation Payment	0.249	0.438	0.570

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A31: Mediation Quality Originality-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Quality	Incentive Originality	Share of changed pairs	0.083	0.186	0.653
Quality	Incentive Originality	Share of removed pairs	-0.030	0.101	0.766
Quality	Incentive Originality	Share of given positive ratings	0.027	0.111	0.806
Quality	Incentive Originality	Random Vote	-0.054	0.152	0.721
Quality	Incentive Originality	Unanimous Vote	-0.289	0.506	0.567
Quality	Incentive Originality	Share of used suggestions	-0.057	0.100	0.570
Quality	Incentive Originality	Number of suggestions	0.058	0.157	0.710
Quality	Incentive Originality	Teamwork	-0.104	0.295	0.723
Quality	Incentive Originality	Leadership	0.707	0.876	0.420
Quality	Incentive Originality	Feedback Assessment	0.178	0.314	0.570
Quality	Incentive Originality	Chat Sentiment	0.010	0.179	0.957
Quality	Incentive Originality	Number Chat Messages	-0.141	0.268	0.599
Quality	Incentive Originality	Group Effort	0.125	0.450	0.781
Quality	Incentive Originality	Peer Pressure	-0.005	0.054	0.919
Quality	Incentive Originality	Enjoyment	-0.126	0.518	0.809
Quality	Incentive Originality	Communication	-0.491	0.801	0.540
Quality	Incentive Originality	Difficulty Quality	0.032	0.202	0.873
Quality	Incentive Originality	Difficulty Originality	-0.678	1.272	0.594
Quality	Incentive Originality	Difficulty Quantity	-0.000	0.165	0.999
Quality	Incentive Originality	Originality	0.129	0.186	0.488
Quality	Incentive Originality	Originality \emptyset FB	0.091	0.197	0.643
Quality	Incentive Originality	Duration Ideagen	0.079	0.204	0.698
Quality	Incentive Originality	Duration FB	-0.180	0.271	0.505
Quality	Incentive Originality	Duration Revision	-0.001	0.014	0.932
Quality	Incentive Originality	Intensity Ideagen	0.015	0.056	0.795
Quality	Incentive Originality	Intensity FB	0.032	0.161	0.841
Quality	Incentive Originality	Intensity Revision	0.000	0.018	0.983
No control group					
Quality	Incentive Originality	Tradeoff	0.179	2.491	0.943
Quality	Incentive Originality	Self-efficacy	2.220	9.819	0.821
Quality	Incentive Originality	Pro-social Group	-1.560	13.389	0.907
Quality	Incentive Originality	Motivation Competition	5.991	493.710	0.990
Quality	Incentive Originality	Motivation Payment	0.454	7.981	0.955

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A32: Mediation Innovation Quantity-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Innovation 75	Incentive Quantity	Share of changed pairs	-0.168	0.150	0.262
Innovation 75	Incentive Quantity	Share of removed pairs	0.001	0.007	0.931
Innovation 75	Incentive Quantity	Share of given positive ratings	0.024	0.043	0.585
Innovation 75	Incentive Quantity	Random Vote	0.004	0.025	0.863
Innovation 75	Incentive Quantity	Unanimous Vote	-0.006	0.031	0.844
Innovation 75	Incentive Quantity	Share of used suggestions	-0.062	0.082	0.450
Innovation 75	Incentive Quantity	Number of suggestions	0.028	0.072	0.703
Innovation 75	Incentive Quantity	Teamwork	0.107	0.149	0.471
Innovation 75	Incentive Quantity	Leadership	0.000	0.037	0.991
Innovation 75	Incentive Quantity	Feedback Assessment	-0.071	0.152	0.637
Innovation 75	Incentive Quantity	Chat Sentiment	0.002	0.060	0.969
Innovation 75	Incentive Quantity	Number Chat Messages	-0.257	0.195	0.189
Innovation 75	Incentive Quantity	Group Effort	0.026	0.099	0.794
Innovation 75	Incentive Quantity	Peer Pressure	-0.061	0.212	0.775
Innovation 75	Incentive Quantity	Enjoyment	0.170	0.196	0.386
Innovation 75	Incentive Quantity	Communication	0.054	0.101	0.596
Innovation 75	Incentive Quantity	Difficulty Quality	0.003	0.030	0.920
Innovation 75	Incentive Quantity	Difficulty Originality	-0.020	0.066	0.759
Innovation 75	Incentive Quantity	Difficulty Quantity	0.031	0.093	0.742
Innovation 75	Incentive Quantity	Duration Ideagen	0.230	0.158	0.145
Innovation 75	Incentive Quantity	Duration FB	0.048	0.082	0.557
Innovation 75	Incentive Quantity	Duration Revision	0.034	0.098	0.728
Innovation 75	Incentive Quantity	Intensity Ideagen	0.128	0.144	0.375
Innovation 75	Incentive Quantity	Intensity FB	-0.018	0.113	0.875
Innovation 75	Incentive Quantity	Intensity Revision	-0.033	0.099	0.735
No control group					
Innovation 75	Incentive Quantity	Tradeoff	-0.257	0.301	0.393
Innovation 75	Incentive Quantity	Self-efficacy	0.007	0.486	0.989
Innovation 75	Incentive Quantity	Pro-social Group	0.093	0.156	0.551
Innovation 75	Incentive Quantity	Motivation Competition	-0.163	0.283	0.565
Innovation 75	Incentive Quantity	Motivation Payment	-0.442	0.491	0.368

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A33: Mediation Innovation Quality-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Innovation 75	Incentive Quality	Share of changed pairs	-0.001	0.020	0.952
Innovation 75	Incentive Quality	Share of removed pairs	-0.016	0.078	0.835
Innovation 75	Incentive Quality	Share of given positive ratings	0.023	0.145	0.874
Innovation 75	Incentive Quality	Random Vote	0.040	0.137	0.770
Innovation 75	Incentive Quality	Unanimous Vote	-0.009	0.118	0.940
Innovation 75	Incentive Quality	Share of used suggestions	0.027	0.089	0.758
Innovation 75	Incentive Quality	Number of suggestions	0.018	0.081	0.822
Innovation 75	Incentive Quality	Teamwork	-0.062	0.251	0.806
Innovation 75	Incentive Quality	Leadership	0.033	0.154	0.830
Innovation 75	Incentive Quality	Feedback Assessment	0.035	0.296	0.904
Innovation 75	Incentive Quality	Chat Sentiment	0.039	0.161	0.809
Innovation 75	Incentive Quality	Number Chat Messages	0.062	0.229	0.787
Innovation 75	Incentive Quality	Group Effort	0.057	0.641	0.929
Innovation 75	Incentive Quality	Peer Pressure	0.171	0.835	0.838
Innovation 75	Incentive Quality	Enjoyment	0.178	0.484	0.713
Innovation 75	Incentive Quality	Communication	0.079	0.284	0.781
Innovation 75	Incentive Quality	Difficulty Quality	0.198	0.603	0.743
Innovation 75	Incentive Quality	Difficulty Originality	-0.538	1.085	0.620
Innovation 75	Incentive Quality	Difficulty Quantity	0.056	0.310	0.857
Innovation 75	Incentive Quality	Duration Ideagen	0.111	0.346	0.748
Innovation 75	Incentive Quality	Duration FB	0.146	0.324	0.652
Innovation 75	Incentive Quality	Duration Revision	0.065	0.298	0.828
Innovation 75	Incentive Quality	Intensity Ideagen	0.014	0.574	0.980
Innovation 75	Incentive Quality	Intensity FB	-0.012	0.072	0.865
Innovation 75	Incentive Quality	Intensity Revision	-0.001	0.013	0.955
No control group					
Innovation 75	Incentive Quality	Tradeoff	4.908	72.111	0.946
Innovation 75	Incentive Quality	Self-efficacy	3.305	111.537	0.976
Innovation 75	Incentive Quality	Pro-social Group	2.112	52.605	0.968
Innovation 75	Incentive Quality	Motivation Competition	3.367	27.646	0.903
Innovation 75	Incentive Quality	Motivation Payment	10.651	518.304	0.984

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A34: Mediation Innovation Index Originality-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Inno Index	Incentive Originality	Share of changed pairs	0.284	1.848	0.878
Inno Index	Incentive Originality	Share of removed pairs	-1.226	7.251	0.866
Inno Index	Incentive Originality	Share of given positive ratings	-0.005	0.112	0.965
Inno Index	Incentive Originality	Random Vote	-0.097	0.631	0.878
Inno Index	Incentive Originality	Unanimous Vote	-0.074	0.500	0.882
Inno Index	Incentive Originality	Share of used suggestions	-0.295	0.441	0.503
Inno Index	Incentive Originality	Number of suggestions	0.319	1.933	0.869
Inno Index	Incentive Originality	Teamwork	-3.649	52.676	0.945
Inno Index	Incentive Originality	Leadership	3.001	19.208	0.876
Inno Index	Incentive Originality	Feedback Assessment	0.564	3.494	0.872
Inno Index	Incentive Originality	Chat Sentiment	0.019	0.361	0.957
Inno Index	Incentive Originality	Number Chat Messages	-0.826	4.652	0.859
Inno Index	Incentive Originality	Group Effort	0.766	2.199	0.728
Inno Index	Incentive Originality	Peer Pressure	-0.269	1.661	0.871
Inno Index	Incentive Originality	Enjoyment	-0.366	3.194	0.909
Inno Index	Incentive Originality	Communication	0.333	1.708	0.845
Inno Index	Incentive Originality	Difficulty Quality	0.006	1.032	0.995
Inno Index	Incentive Originality	Difficulty Originality	-12.945	536.086	0.981
Inno Index	Incentive Originality	Difficulty Quantity	-0.001	0.431	0.999
Inno Index	Incentive Originality	Duration Ideagen	0.106	0.746	0.887
Inno Index	Incentive Originality	Duration FB	-3.556	30.533	0.907
Inno Index	Incentive Originality	Duration Revision	-0.030	0.270	0.910
Inno Index	Incentive Originality	Intensity Ideagen	0.333	1.686	0.843
Inno Index	Incentive Originality	Intensity FB	0.218	1.668	0.896
Inno Index	Incentive Originality	Intensity Revision	0.027	0.222	0.904
No control group					
Inno Index	Incentive Originality	Tradeoff	-0.018	0.092	0.842
Inno Index	Incentive Originality	Self-efficacy	-0.066	0.759	0.930
Inno Index	Incentive Originality	Pro-social Group	-0.190	0.340	0.577
Inno Index	Incentive Originality	Motivation Competition	-0.037	0.153	0.806
Inno Index	Incentive Originality	Motivation Payment	0.366	0.636	0.565

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A35: Mediation Innovation Index Quality-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Inno Index	Incentive Quality	Share of changed pairs	0.016	0.119	0.893
Inno Index	Incentive Quality	Share of removed pairs	-0.038	0.169	0.824
Inno Index	Incentive Quality	Share of given positive ratings	-0.008	0.049	0.868
Inno Index	Incentive Quality	Random Vote	0.008	0.057	0.892
Inno Index	Incentive Quality	Unanimous Vote	-0.029	0.091	0.748
Inno Index	Incentive Quality	Share of used suggestions	0.173	1.062	0.871
Inno Index	Incentive Quality	Number of suggestions	0.044	0.111	0.695
Inno Index	Incentive Quality	Teamwork	0.027	0.147	0.854
Inno Index	Incentive Quality	Leadership	0.069	0.210	0.744
Inno Index	Incentive Quality	Feedback Assessment	0.007	0.060	0.906
Inno Index	Incentive Quality	Chat Sentiment	-0.042	0.190	0.826
Inno Index	Incentive Quality	Number Chat Messages	-0.027	0.097	0.781
Inno Index	Incentive Quality	Group Effort	0.004	0.041	0.929
Inno Index	Incentive Quality	Peer Pressure	0.217	0.419	0.604
Inno Index	Incentive Quality	Enjoyment	0.155	0.278	0.578
Inno Index	Incentive Quality	Communication	0.011	0.056	0.838
Inno Index	Incentive Quality	Difficulty Quality	-0.193	0.371	0.603
Inno Index	Incentive Quality	Difficulty Originality	-0.110	0.243	0.652
Inno Index	Incentive Quality	Difficulty Quantity	0.051	0.276	0.854
Inno Index	Incentive Quality	Duration Ideagen	0.008	0.050	0.872
Inno Index	Incentive Quality	Duration FB	0.083	0.150	0.577
Inno Index	Incentive Quality	Duration Revision	0.005	0.032	0.880
Inno Index	Incentive Quality	Intensity Ideagen	0.227	0.338	0.502
Inno Index	Incentive Quality	Intensity FB	0.008	0.032	0.799
Inno Index	Incentive Quality	Intensity Revision	0.019	0.158	0.904
No control group					
Inno Index	Incentive Quality	Tradeoff	-2.419	18.991	0.899
Inno Index	Incentive Quality	Self-efficacy	0.621	3.147	0.844
Inno Index	Incentive Quality	Pro-social Group	0.390	2.006	0.846
Inno Index	Incentive Quality	Motivation Competition	-1.354	8.168	0.868
Inno Index	Incentive Quality	Motivation Payment	0.201	1.960	0.919

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A36: Mediation Innovation Index Quantity-PRP

Outcome	Treatment	Mediator	Proportion	SE	p-value
Inno Index	Incentive Quantity	Share of changed pairs	-0.128	0.208	0.540
Inno Index	Incentive Quantity	Share of removed pairs	-0.007	0.036	0.845
Inno Index	Incentive Quantity	Share of given positive ratings	0.026	0.055	0.639
Inno Index	Incentive Quantity	Random Vote	0.009	0.039	0.822
Inno Index	Incentive Quantity	Unanimous Vote	-0.013	0.065	0.843
Inno Index	Incentive Quantity	Share of used suggestions	-0.034	0.088	0.697
Inno Index	Incentive Quantity	Number of suggestions	0.040	0.102	0.693
Inno Index	Incentive Quantity	Teamwork	0.080	0.129	0.536
Inno Index	Incentive Quantity	Leadership	0.002	0.140	0.991
Inno Index	Incentive Quantity	Feedback Assessment	-0.050	0.111	0.657
Inno Index	Incentive Quantity	Chat Sentiment	-0.043	0.113	0.705
Inno Index	Incentive Quantity	Number Chat Messages	-0.308	0.290	0.287
Inno Index	Incentive Quantity	Group Effort	0.029	0.110	0.791
Inno Index	Incentive Quantity	Peer Pressure	0.387	0.401	0.335
Inno Index	Incentive Quantity	Enjoyment	0.357	0.306	0.243
Inno Index	Incentive Quantity	Communication	0.024	0.081	0.773
Inno Index	Incentive Quantity	Difficulty Quality	-0.013	0.113	0.908
Inno Index	Incentive Quantity	Difficulty Originality	0.068	0.151	0.653
Inno Index	Incentive Quantity	Difficulty Quantity	0.014	0.051	0.777
Inno Index	Incentive Quantity	Duration Ideagen	0.225	0.226	0.321
Inno Index	Incentive Quantity	Duration FB	0.062	0.099	0.526
Inno Index	Incentive Quantity	Duration Revision	0.026	0.080	0.743
Inno Index	Incentive Quantity	Intensity Ideagen	-0.404	0.300	0.178
Inno Index	Incentive Quantity	Intensity FB	-0.026	0.171	0.878
Inno Index	Incentive Quantity	Intensity Revision	0.002	0.029	0.942
No control group					
Inno Index	Incentive Quantity	Tradeoff	0.135	0.582	0.817
Inno Index	Incentive Quantity	Self-efficacy	0.695	1.322	0.599
Inno Index	Incentive Quantity	Pro-social Group	0.219	0.510	0.668
Inno Index	Incentive Quantity	Motivation Competition	-0.552	1.246	0.658
Inno Index	Incentive Quantity	Motivation Payment	-0.448	0.998	0.654

Note: The table reports proportions mediated and corresponding standard errors and p-values of a mediation analysis based on linear outcome and mediator models (standard errors clustered at the group level).

Table A37: Process

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quantity	Quantity FB	Originality \emptyset	Originality \emptyset FB	Originality	Innovation 75	Inno Index
Share of changed pairs	0.231 (0.207)	-2.024*** (0.340)	-1.914*** (0.337)	-0.324 (0.920)	-2.508** (1.029)	-0.762 (1.703)	0.037 (0.044)	0.077 (0.183)
Share of removed pairs	0.011 (0.540)	4.538*** (0.834)	-1.312* (0.723)	0.611 (2.353)	-4.342 (2.799)	-4.124 (4.842)	0.062 (0.112)	-0.312 (0.454)
Share Invalid FB	0.147 (0.269)	-2.858*** (0.470)	-2.885*** (0.480)	3.239*** (1.153)	2.961** (1.281)	5.482*** (2.022)	0.067 (0.054)	0.509*** (0.194)
Share of given positive ratings	0.426 (0.303)	-3.491*** (0.562)	-3.406*** (0.541)	-6.059*** (1.129)	-6.253*** (1.127)	-2.480 (2.222)	-0.002 (0.062)	0.058 (0.225)
Random Vote	-0.178 (0.181)	0.297 (0.248)	0.227 (0.250)	1.446** (0.692)	1.125 (0.762)	1.776 (1.573)	0.005 (0.037)	0.032 (0.138)
Unanimous Vote	0.272 (0.220)	0.808*** (0.283)	0.698** (0.283)	-0.663 (0.828)	-0.999 (0.942)	-0.459 (1.913)	0.053 (0.051)	0.124 (0.168)
Vote deviation	0.141 (0.544)	1.586** (0.612)	1.357** (0.596)	-2.174 (2.272)	-3.562 (2.558)	0.114 (5.229)	0.156 (0.137)	0.091 (0.441)
Vote own contribution	0.129 (0.276)	-0.086 (0.355)	-0.056 (0.343)	-0.623 (1.014)	-0.219 (1.060)	-0.137 (2.167)	0.013 (0.045)	0.065 (0.215)
Contribution dominance	-0.130 (0.446)	0.136 (1.012)	0.104 (1.010)	1.751 (1.808)	2.973 (1.962)	-0.152 (3.131)	-0.099 (0.105)	-0.088 (0.316)
Number of suggestions	0.089 (0.122)	0.597** (0.268)	0.548** (0.271)	-0.394 (0.492)	-0.515 (0.533)	-0.853 (0.913)	-0.012 (0.024)	-0.013 (0.095)
Teamwork	-0.139 (0.106)	0.439* (0.226)	0.405* (0.221)	-0.552 (0.411)	-0.315 (0.412)	0.252 (0.681)	-0.033 (0.022)	-0.062 (0.079)
Leadership	0.113 (0.073)	-0.047 (0.188)	-0.041 (0.188)	0.023 (0.322)	0.191 (0.309)	0.002 (0.558)	0.031* (0.017)	0.066 (0.053)
Feedback Assessment	0.135 (0.095)	0.204 (0.250)	0.165 (0.248)	-0.180 (0.419)	0.061 (0.432)	-0.947 (0.655)	0.017 (0.021)	0.006 (0.069)
Chat Sentiment	0.055 (0.131)	0.289 (0.208)	0.259 (0.207)	-0.855* (0.479)	-1.130** (0.536)	0.512 (0.960)	0.023 (0.022)	0.072 (0.097)
Number Chat Messages	-0.007 (0.013)	0.031 (0.022)	0.031 (0.021)	0.165*** (0.051)	0.154*** (0.054)	0.065 (0.092)	0.001 (0.002)	0.001 (0.009)
Group Effort	0.034 (0.079)	-0.243 (0.195)	-0.217 (0.190)	-0.161 (0.306)	-0.171 (0.313)	0.243 (0.522)	0.051*** (0.017)	0.039 (0.053)
Peer Pressure	-0.022 (0.085)	0.198 (0.205)	0.208 (0.204)	-0.006 (0.329)	0.055 (0.372)	-0.104 (0.611)	0.008 (0.017)	-0.021 (0.063)
Enjoyment	-0.063 (0.099)	-0.016 (0.246)	-0.021 (0.242)	0.484 (0.412)	0.160 (0.491)	1.186 (0.718)	-0.021 (0.022)	0.055 (0.069)
Heterogeneity Originality	0.004 (0.011)	-0.091*** (0.018)	-0.086*** (0.018)	-0.265*** (0.056)	-0.364*** (0.059)	-0.460*** (0.124)	0.001 (0.002)	-0.033*** (0.009)
Heterogeneity Quantity	0.065 (0.065)	1.988*** (0.145)	1.935*** (0.146)	0.512** (0.233)	0.410 (0.248)	0.193 (0.563)	0.039*** (0.014)	0.053 (0.054)
Communication	0.057 (0.147)	-0.635** (0.281)	-0.569** (0.274)	0.884 (0.595)	0.537 (0.630)	-0.085 (1.195)	0.034 (0.036)	0.027 (0.119)
Difficulty Quality	-0.004 (0.087)	-0.084 (0.155)	-0.077 (0.155)	0.471 (0.357)	0.282 (0.368)	0.497 (0.628)	-0.005 (0.019)	0.036 (0.063)
Difficulty Originality	-0.006 (0.090)	0.256 (0.194)	0.237 (0.190)	-0.056 (0.361)	0.248 (0.387)	0.266 (0.579)	0.027 (0.021)	0.017 (0.066)
Difficulty Quantity	0.082 (0.076)	-0.348** (0.161)	-0.362** (0.158)	-0.195 (0.305)	-0.239 (0.306)	-0.274 (0.534)	-0.001 (0.016)	0.027 (0.058)
Constant	-0.420 (0.916)	2.359 (1.898)	2.688 (1.891)	66.325*** (4.216)	65.861*** (4.505)	54.668*** (7.058)	-0.033 (0.207)	-1.235* (0.649)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	906	906	906	906	906	906	906	906
R2	0.042	0.539	0.534	0.146	0.148	0.072	0.045	0.058

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A38: Process (controlled for effort)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quantity	Quantity FB	Originality \emptyset	Originality \emptyset FB	Originality	Innovation 75	Inno Index
Share of changed pairs	0.232 (0.366)	-1.829*** (0.324)	-1.131*** (0.317)	-0.325 (1.366)	-2.739* (1.550)	1.597 (2.672)	0.018 (0.085)	0.259 (0.303)
Share of removed pairs	0.294 (0.582)	-0.685 (0.548)	-6.193*** (0.596)	-0.450 (2.532)	-5.508* (2.941)	-1.682 (5.168)	0.015 (0.120)	0.043 (0.488)
Share Invalid FB	0.139 (0.272)	-0.460** (0.232)	-0.630** (0.245)	2.251* (1.289)	2.030 (1.388)	2.251 (2.264)	0.078 (0.052)	0.255 (0.212)
Share of given positive ratings	0.278 (0.323)	-0.735*** (0.261)	-0.740*** (0.260)	-5.350*** (1.216)	-5.487*** (1.182)	-2.730 (2.201)	0.031 (0.061)	-0.048 (0.234)
Random Vote	-0.153 (0.182)	-0.064 (0.137)	-0.115 (0.138)	1.238* (0.679)	0.941 (0.744)	1.834 (1.624)	-0.001 (0.037)	0.052 (0.141)
Unanimous Vote	0.277 (0.222)	0.116 (0.167)	0.031 (0.168)	-0.823 (0.819)	-1.177 (0.929)	-0.258 (1.912)	0.039 (0.050)	0.142 (0.170)
Vote deviation	0.108 (0.558)	0.517 (0.443)	0.353 (0.440)	-2.175 (2.270)	-3.622 (2.543)	0.878 (5.284)	0.130 (0.134)	0.131 (0.448)
Vote own contribution	0.129 (0.274)	0.140 (0.209)	0.150 (0.205)	-0.478 (1.035)	-0.065 (1.076)	-0.183 (2.202)	0.020 (0.046)	0.062 (0.214)
Contribution dominance	-0.035 (0.441)	-0.600 (0.557)	-0.589 (0.575)	1.227 (1.767)	2.535 (1.861)	-0.016 (3.255)	-0.102 (0.103)	-0.021 (0.333)
Number of suggestions	-0.160 (0.340)	0.009 (0.245)	0.089 (0.250)	-0.024 (1.244)	0.018 (1.456)	0.557 (2.271)	-0.105 (0.068)	-0.051 (0.235)
Teamwork	-0.124 (0.103)	0.067 (0.095)	0.045 (0.098)	-0.658 (0.402)	-0.434 (0.393)	0.273 (0.683)	-0.039* (0.021)	-0.052 (0.077)
Leadership	0.114 (0.071)	-0.038 (0.073)	-0.034 (0.077)	-0.034 (0.310)	0.138 (0.298)	-0.057 (0.561)	0.030* (0.017)	0.063 (0.052)
Feedback Assessment	0.145 (0.096)	0.058 (0.107)	0.029 (0.111)	-0.323 (0.420)	-0.092 (0.421)	-1.042 (0.675)	0.013 (0.021)	0.004 (0.071)
Chat Sentiment	0.052 (0.131)	0.034 (0.110)	-0.004 (0.117)	-0.920* (0.480)	-1.207** (0.530)	0.472 (0.947)	0.017 (0.022)	0.067 (0.097)
Number Chat Messages	-0.008 (0.013)	0.009 (0.010)	0.008 (0.011)	0.139*** (0.051)	0.132** (0.054)	0.063 (0.097)	0.000 (0.002)	0.000 (0.009)
Group Effort	0.024 (0.081)	-0.053 (0.093)	-0.035 (0.094)	-0.150 (0.303)	-0.167 (0.314)	0.133 (0.533)	0.053*** (0.017)	0.025 (0.054)
Peer Pressure	-0.000 (0.084)	-0.016 (0.088)	0.005 (0.089)	0.028 (0.314)	0.103 (0.354)	0.164 (0.642)	0.011 (0.017)	0.013 (0.065)
Enjoyment	-0.063 (0.101)	0.096 (0.091)	0.092 (0.095)	0.508 (0.396)	0.209 (0.467)	1.220* (0.728)	-0.020 (0.021)	0.057 (0.072)
Heterogeneity Originality	-0.000 (0.011)	-0.038*** (0.009)	-0.037*** (0.010)	-0.249*** (0.057)	-0.345*** (0.060)	-0.462*** (0.125)	0.002 (0.002)	-0.036*** (0.010)
Heterogeneity Quantity	0.102 (0.070)	0.901*** (0.074)	0.873*** (0.077)	0.199 (0.251)	0.098 (0.283)	0.400 (0.634)	0.022 (0.016)	0.091 (0.058)
Communication	0.032 (0.150)	0.031 (0.144)	0.069 (0.148)	1.161** (0.580)	0.826 (0.610)	-0.158 (1.202)	0.046 (0.036)	0.006 (0.119)
Difficulty Quality	0.010 (0.090)	-0.072 (0.093)	-0.051 (0.096)	0.512 (0.355)	0.317 (0.375)	0.509 (0.642)	-0.002 (0.020)	0.045 (0.062)
Difficulty Originality	-0.003 (0.088)	0.135 (0.099)	0.130 (0.098)	0.022 (0.362)	0.332 (0.380)	0.515 (0.596)	0.027 (0.020)	0.038 (0.066)
Difficulty Quantity	0.075 (0.078)	-0.139* (0.077)	-0.158** (0.079)	-0.281 (0.309)	-0.309 (0.313)	-0.398 (0.559)	-0.000 (0.017)	0.013 (0.061)
Constant	-0.417 (0.970)	-5.428*** (0.949)	-5.104*** (1.008)	64.151*** (4.156)	63.533*** (4.333)	54.705*** (7.111)	-0.213 (0.207)	-1.230* (0.676)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Effort Controls	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	906	906	906	906	906	906	906	906
R2	0.050	0.870	0.859	0.172	0.170	0.088	0.059	0.072

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A39: Effort

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality	Quantity	Quantity FB	Originality \emptyset	Originality \emptyset FB	Originality	Innovation 75	Inno Index
Duration Ideagen	-0.003* (0.002)	0.062*** (0.003)	0.059*** (0.003)	0.042*** (0.006)	0.045*** (0.006)	0.031*** (0.011)	0.001*** (0.000)	0.001 (0.001)
Duration FB	-0.003 (0.004)	0.004 (0.003)	0.004 (0.004)	-0.037*** (0.014)	-0.036** (0.016)	-0.059** (0.028)	-0.002** (0.001)	-0.006** (0.003)
Duration Revision	0.001 (0.003)	0.007* (0.004)	0.000 (0.004)	0.011 (0.014)	0.003 (0.015)	0.040 (0.026)	0.000 (0.001)	0.004 (0.003)
Intensity Ideagen	-0.024 (0.092)	4.435*** (0.261)	4.414*** (0.271)	0.294 (0.442)	0.606 (0.456)	-1.311 (0.828)	0.048** (0.023)	-0.115* (0.067)
Intensity FB	0.251** (0.112)	-0.107 (0.115)	0.060 (0.132)	0.416 (0.534)	0.205 (0.608)	0.687 (0.883)	0.033 (0.025)	0.200** (0.080)
Intensity Revision	0.000 (0.062)	0.058 (0.061)	-0.249*** (0.075)	0.330 (0.309)	-0.447 (0.347)	-1.134* (0.601)	0.013 (0.014)	-0.087 (0.054)
Constant	0.008 (0.710)	-7.219*** (0.978)	-7.397*** (1.096)	60.767*** (3.390)	59.010*** (3.527)	51.672*** (5.650)	-0.070 (0.184)	-1.215** (0.480)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	960	960	960	960	960	960	960	960
R2	0.030	0.813	0.767	0.092	0.080	0.045	0.035	0.041

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). All variables are at the group level. Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A40: Individual Idea Divergence

	before feedback and revision				after feedback and revision			
	(1) Divergence	(2) Divergence	(3) Divergence	(4) Divergence	(5) Divergence FB	(6) Divergence FB	(7) Divergence FB	(8) Divergence FB
Incentive Originality	0.640 (0.801)	0.791 (0.781)	0.811 (0.777)	0.213 (0.375)	0.550 (0.788)	0.591 (0.768)	0.621 (0.765)	0.157 (0.369)
Incentive Quality	0.764 (0.818)	0.847 (0.796)	0.871 (0.792)	0.715* (0.378)	0.253 (0.811)	0.285 (0.793)	0.325 (0.790)	0.456 (0.382)
Incentive Quantity	3.518*** (0.833)	3.474*** (0.819)	3.479*** (0.818)	-0.205 (0.416)	3.314*** (0.825)	3.164*** (0.811)	3.176*** (0.809)	-0.367 (0.385)
Quantity				5.694*** (0.314)				
Quantity FB								5.698*** (0.331)
Constant	28.753*** (0.577)	24.906*** (3.021)	20.537*** (2.990)	16.029*** (1.564)	28.278*** (0.571)	23.131*** (3.036)	19.056*** (3.009)	15.119*** (1.555)
Individual Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	2,739	2,716	2,716	2,716	2,713	2,690	2,690	2,690
R2	0.019	0.054	0.119	0.632	0.019	0.056	0.118	0.636

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A41: Individual Flexibility

	before feedback and revision				after feedback and revision			
	(1) Flexibility	(2) Flexibility	(3) Flexibility	(4) Flexibility	(5) Flexibility FB	(6) Flexibility FB	(7) Flexibility FB	(8) Flexibility FB
Incentive Originality	0.104 (0.544)	0.226 (0.541)	0.233 (0.542)	0.144 (0.536)	-0.047 (0.522)	-0.048 (0.518)	-0.050 (0.517)	-0.114 (0.509)
Incentive Quality	0.837 (0.550)	0.844 (0.539)	0.843 (0.541)	0.820 (0.531)	0.258 (0.561)	0.236 (0.560)	0.237 (0.561)	0.255 (0.547)
Incentive Quantity	0.846 (0.550)	0.787 (0.549)	0.803 (0.551)	0.255 (0.566)	0.457 (0.484)	0.276 (0.493)	0.291 (0.494)	-0.201 (0.493)
Quantity				0.846*** (0.135)				
Quantity FB								0.792*** (0.125)
Constant	71.398*** (0.377)	70.590*** (2.055)	69.730*** (2.079)	69.060*** (2.061)	71.372*** (0.341)	70.558*** (2.218)	69.707*** (2.222)	69.160*** (2.209)
Individual Controls		✓	✓	✓		✓	✓	✓
Round FE			✓	✓			✓	✓
Obs.	2,739	2,716	2,716	2,716	2,713	2,690	2,690	2,690
R2	0.002	0.011	0.060	0.072	0.000	0.013	0.066	0.077

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A42: Ratings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality \emptyset II	Quality II	Quality III	Originality \emptyset II	Originality II	Originality III	Innovation \emptyset II	Innovation II
Incentive Originality	-0.356** (0.176)	-0.405 (0.288)	-0.055 (0.143)	-0.471*** (0.166)	-0.292 (0.242)	0.202 (0.167)	-0.637*** (0.153)	-0.654** (0.252)
Incentive Quality	0.390** (0.161)	0.590** (0.278)	0.021 (0.142)	0.031 (0.163)	0.067 (0.249)	0.206 (0.188)	0.006 (0.141)	0.060 (0.279)
Incentive Quantity	0.039 (0.187)	-0.163 (0.293)	0.134 (0.140)	-0.206 (0.180)	-0.431* (0.226)	0.346* (0.189)	-0.103 (0.153)	-0.317 (0.234)
Constant	6.770*** (1.192)	8.771*** (1.741)	8.531*** (0.941)	6.299*** (1.211)	5.649*** (1.564)	7.943*** (1.153)	6.033*** (0.915)	4.872*** (1.588)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	960	959	960	960	960	960	960	959
R2	0.128	0.084	0.108	0.038	0.033	0.133	0.055	0.045

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A43: Ratings with Interaction Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quality \emptyset II	Quality II	Quality III	Originality \emptyset II	Originality II	Originality III	Innovation \emptyset II	Innovation II
Incentive Originality	-0.143 (0.152)	-0.359 (0.253)	-0.094 (0.127)	-0.303** (0.145)	-0.164 (0.215)	0.239 (0.148)	-0.666*** (0.155)	-0.693*** (0.254)
Incentive Quality	0.334** (0.137)	0.513** (0.239)	-0.021 (0.126)	-0.131 (0.141)	-0.107 (0.216)	0.207 (0.167)	-0.018 (0.142)	0.029 (0.281)
Incentive Quantity	0.252 (0.155)	-0.039 (0.273)	0.035 (0.126)	-0.221 (0.161)	-0.317 (0.212)	0.309* (0.177)	-0.120 (0.151)	-0.340 (0.232)
Rater Performance	0.157*** (0.046)	0.195*** (0.069)		-0.027 (0.038)	0.004 (0.069)		0.096** (0.047)	0.127* (0.070)
Originality \emptyset II	0.574*** (0.091)							
Quantity FB	-0.030 (0.074)	0.024 (0.101)	0.227** (0.093)	-0.011 (0.062)	-0.019 (0.063)	0.278** (0.120)		
Originality \emptyset II \times Quantity FB	-0.007 (0.013)							
Originality II		0.403*** (0.105)						
Originality II \times Quantity FB		-0.005 (0.016)						
Originality III			0.388*** (0.077)					
Originality III \times Quantity FB			-0.027** (0.012)					
Quality \emptyset II				0.430*** (0.073)				
Quality \emptyset II \times Quantity FB				0.003 (0.011)				
Quality II					0.304*** (0.071)			
Quality II \times Quantity FB					-0.001 (0.010)			
Quality III						0.512*** (0.088)		
Quality III \times Quantity FB						-0.032** (0.015)		
Constant	2.801** (1.122)	6.202*** (1.638)	5.468*** (1.063)	3.424*** (1.123)	2.853* (1.512)	3.535*** (1.190)	5.873*** (0.930)	4.661*** (1.603)
Group Controls	✓	✓	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	960	959	960	960	959	960	960	959
R2	0.352	0.195	0.184	0.270	0.144	0.202	0.059	0.049

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the group level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Table A44: Ratings Determinants Quality Assessors

	(1)	(2)	(3)	(4)	(5)	(6)
	Guesser Quality	Guesser Quality	Guesser Originality	Guesser Originality	Guesser Creativity	Guesser Creativity
Quality	0.233*** (0.046)	0.159*** (0.046)	0.064* (0.038)	0.032 (0.038)	0.091** (0.041)	0.068* (0.041)
Age	-0.058** (0.028)	-0.060** (0.028)	-0.069** (0.030)	-0.070** (0.030)	-0.077** (0.031)	-0.078** (0.031)
Female	0.409* (0.243)	0.404* (0.243)	0.476** (0.239)	0.475** (0.240)	0.553** (0.245)	0.552** (0.246)
Law and Econ	0.083 (0.248)	0.088 (0.246)	0.329 (0.295)	0.332 (0.296)	0.268 (0.278)	0.270 (0.278)
Humanities	-0.302 (0.292)	-0.291 (0.292)	-0.038 (0.305)	-0.033 (0.305)	-0.161 (0.295)	-0.157 (0.296)
STEM	0.137 (0.315)	0.139 (0.315)	0.228 (0.328)	0.229 (0.329)	0.049 (0.318)	0.050 (0.319)
Native Speaker	-0.495* (0.292)	-0.483 (0.292)	-0.640** (0.289)	-0.635** (0.290)	-0.379 (0.294)	-0.375 (0.295)
Risk Tolerance	-0.045 (0.047)	-0.046 (0.047)	0.045 (0.048)	0.044 (0.048)	0.029 (0.048)	0.029 (0.048)
RAT	-0.029 (0.104)	-0.029 (0.104)	0.080 (0.106)	0.081 (0.106)	0.036 (0.112)	0.036 (0.112)
Creativity Self-Assessment	0.152 (0.125)	0.155 (0.125)	0.122 (0.136)	0.122 (0.136)	0.159 (0.136)	0.159 (0.136)
Competitiveness	0.152 (0.139)	0.149 (0.139)	0.088 (0.143)	0.087 (0.143)	0.151 (0.141)	0.151 (0.141)
Creative Style	-0.055 (0.099)	-0.061 (0.099)	-0.177* (0.094)	-0.180* (0.095)	-0.134 (0.093)	-0.136 (0.093)
Novelty Seeking	-0.069 (0.106)	-0.069 (0.106)	0.000 (0.095)	-0.000 (0.096)	-0.078 (0.099)	-0.078 (0.099)
Creative Purpose	-0.074 (0.103)	-0.073 (0.103)	0.043 (0.097)	0.044 (0.097)	0.102 (0.096)	0.103 (0.096)
Creative Exploration	0.226*** (0.075)	0.223*** (0.075)	0.050 (0.081)	0.049 (0.081)	0.065 (0.081)	0.064 (0.081)
Strategy	-0.090 (0.082)	-0.087 (0.082)	-0.078 (0.087)	-0.077 (0.087)	-0.023 (0.083)	-0.021 (0.083)
Strategy II	-0.155 (0.151)	-0.148 (0.151)	-0.043 (0.163)	-0.041 (0.164)	-0.076 (0.155)	-0.074 (0.155)
Strategy III	0.006 (0.103)	0.000 (0.104)	0.025 (0.111)	0.022 (0.112)	-0.050 (0.095)	-0.052 (0.096)
Strategy IV	-0.010 (0.103)	-0.010 (0.103)	-0.119 (0.104)	-0.120 (0.104)	-0.047 (0.095)	-0.048 (0.095)
Strategy V	0.167 (0.102)	0.166 (0.101)	0.052 (0.107)	0.051 (0.107)	0.130 (0.108)	0.130 (0.108)
Taboo Words	0.091 (0.087)	0.087 (0.087)	0.098 (0.094)	0.096 (0.095)	0.138 (0.090)	0.136 (0.090)
Difficulty Gen	-0.206* (0.119)	-0.226* (0.118)	-0.102 (0.127)	-0.112 (0.128)	-0.132 (0.127)	-0.139 (0.128)
Difficulty Select	0.237** (0.102)	0.235** (0.102)	0.107 (0.101)	0.106 (0.102)	0.121 (0.099)	0.120 (0.099)
Enjoyment	0.033 (0.088)	0.035 (0.088)	0.148 (0.103)	0.149 (0.103)	0.075 (0.093)	0.076 (0.093)
Motivation Competition	-0.070 (0.084)	-0.065 (0.084)	-0.022 (0.090)	-0.021 (0.090)	-0.019 (0.082)	-0.018 (0.083)
Motivation Payment	0.063 (0.093)	0.068 (0.093)	0.032 (0.103)	0.034 (0.103)	0.007 (0.092)	0.009 (0.092)
Constant	6.953*** (0.768)	7.765*** (0.761)	6.405*** (0.827)	6.814*** (0.832)	6.715*** (0.822)	6.894*** (0.813)
Round FE		✓		✓		✓
Obs.	959	959	960	960	960	960
R2	0.094	0.163	0.090	0.106	0.099	0.114

Note: The table reports coefficients of OLS regressions (standard errors in parentheses, clustered at the individual level). Significance levels are indicated as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Incentive Originality, Incentive Quality, and Incentive Quantity are treatment dummies. Flat Pay is the omitted category.

Appendix B: Supplementary Figures

Figure A1: Correlation Outcomes

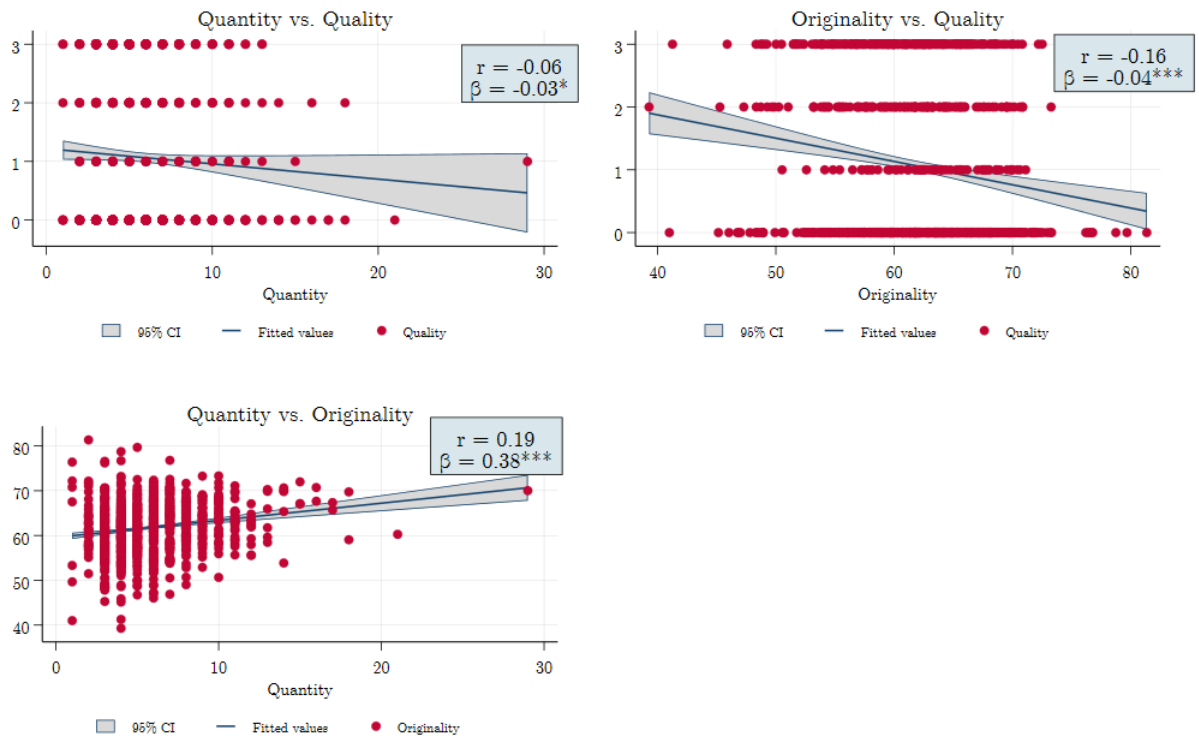
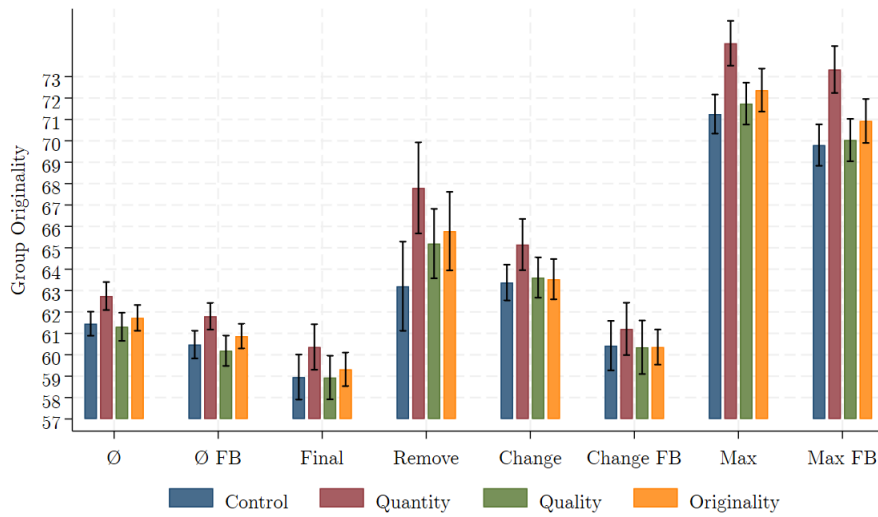
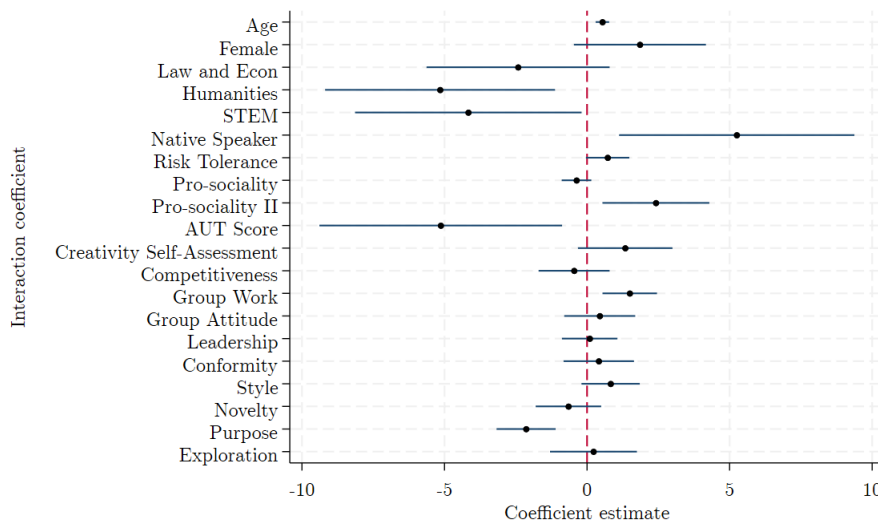


Figure A2: Bias against Originality by Treatment



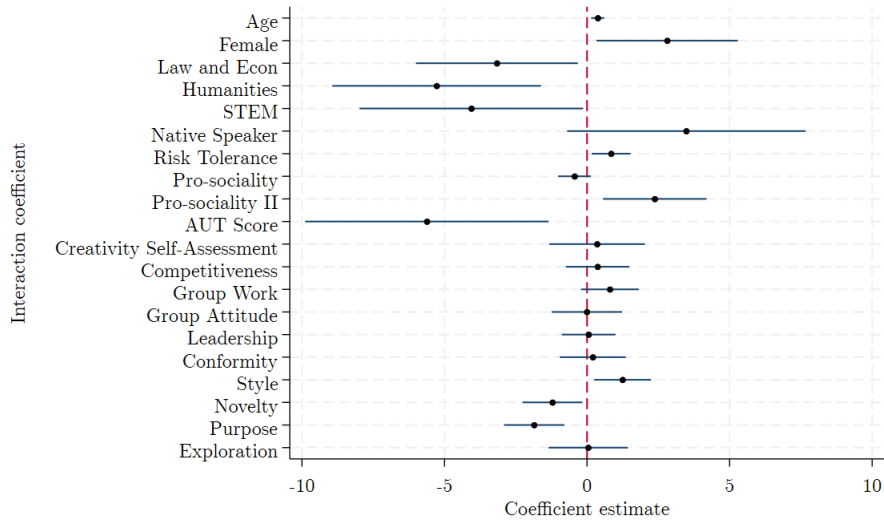
Note: The bar chart reports the average originality of all clue pairs before feedback and revision (\emptyset), after feedback and revision (\emptyset FB), and of the group's final selected clue pair (Final). In addition, it depicts the mean originality of clue pairs discarded during revision (Remove), of modified clue pairs before (Change) and after revision (Change FB), and of each group's most original clue pair before (Max) and after feedback and revision (Max FB). Error bars indicate 90% confidence intervals.

Figure A3: Heterogeneities Originality Quantity-PRP before FB



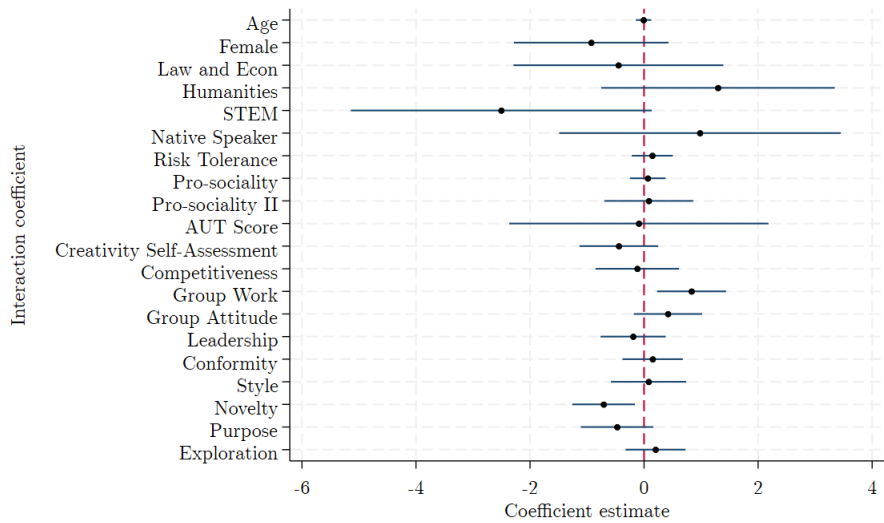
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A4: Heterogeneities Originality Quantity-PRP after FB



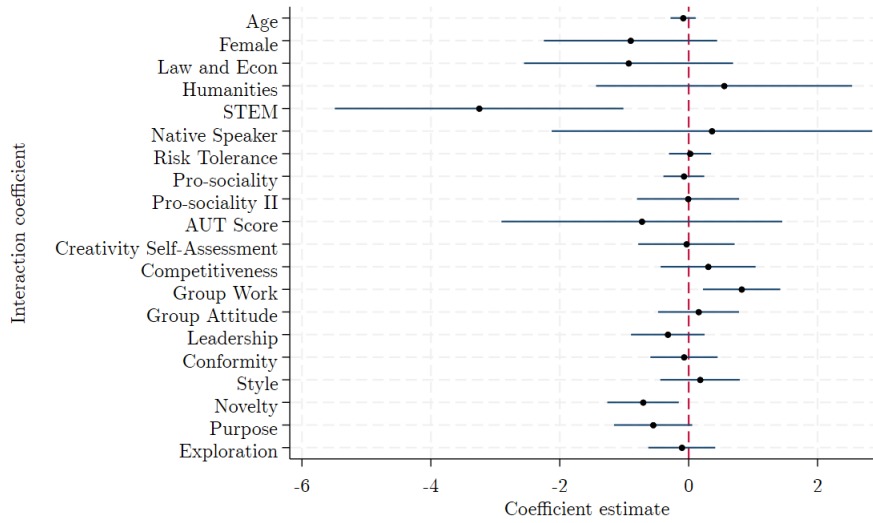
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A5: Heterogeneities Individual Originality Quantity-PRP before FB



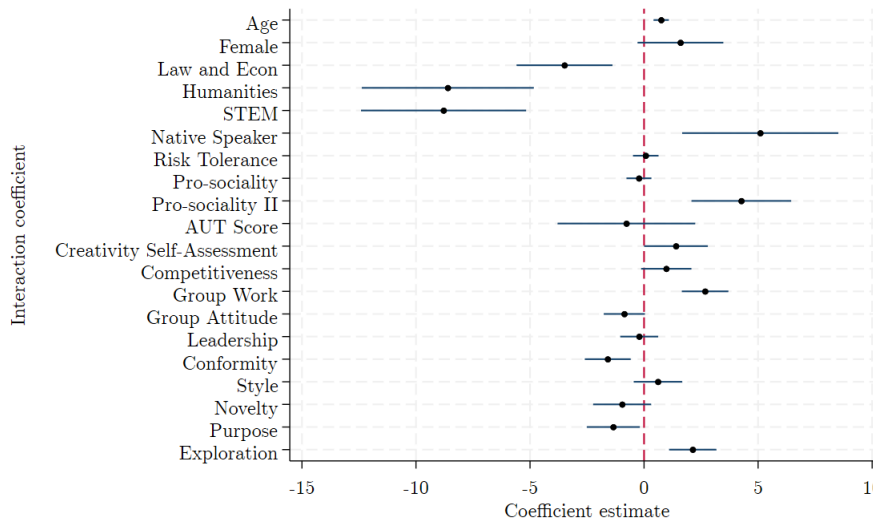
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the individual level.

Figure A6: Heterogeneities Individual Originality Quantity-PRP after FB



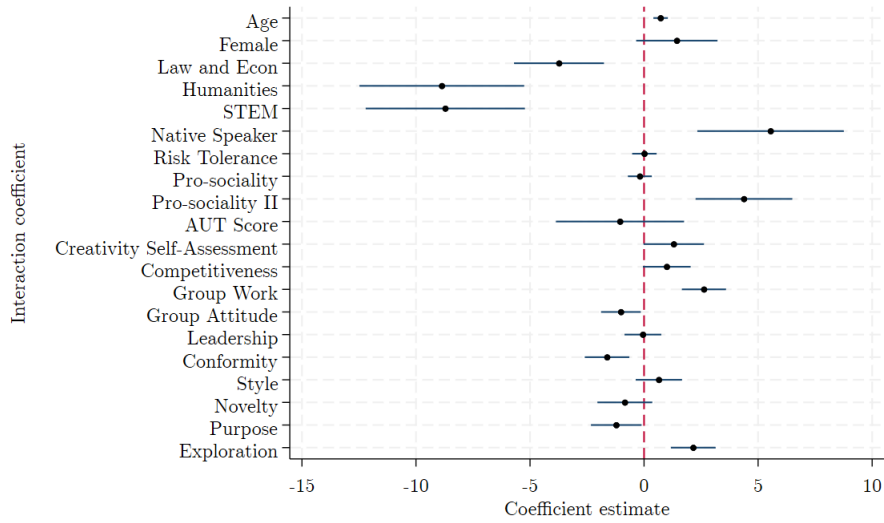
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the individual level.

Figure A7: Heterogeneities Quantity Quantity-PRP before FB



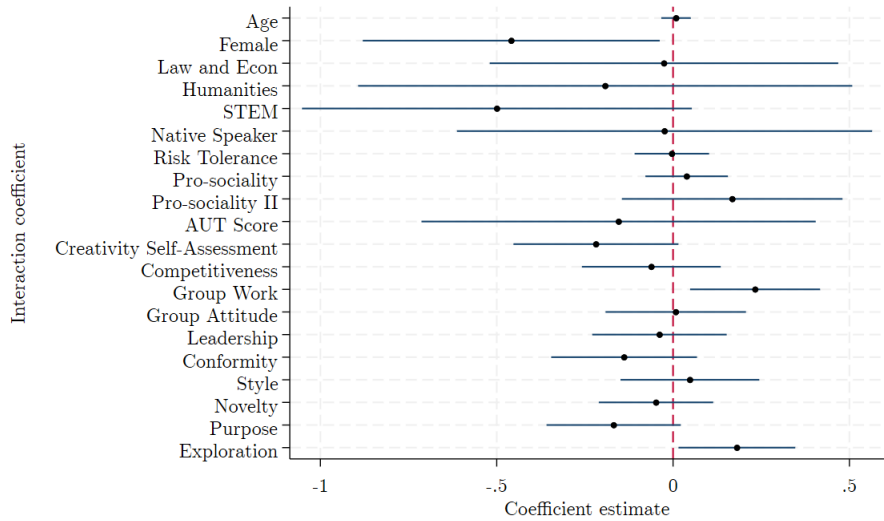
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A8: Heterogeneities Quantity Quantity-PRP after FB



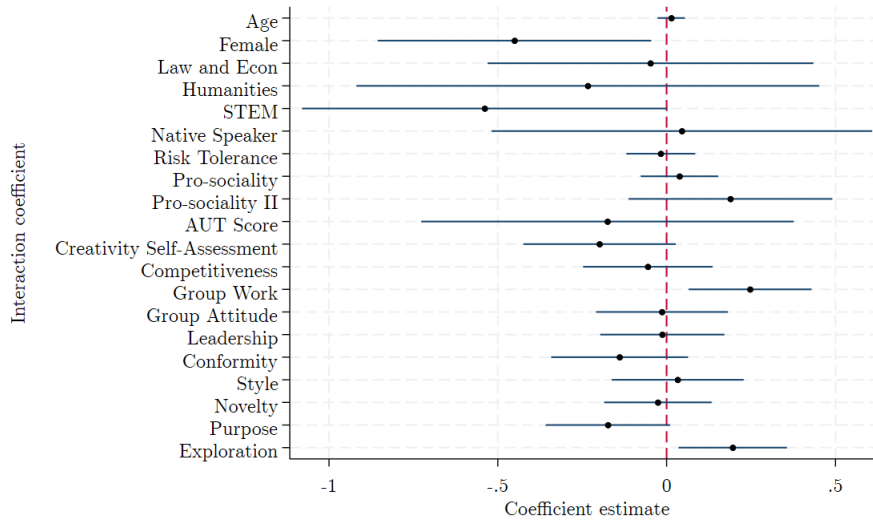
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A9: Heterogeneities Individual Quantity Quantity-PRP before FB



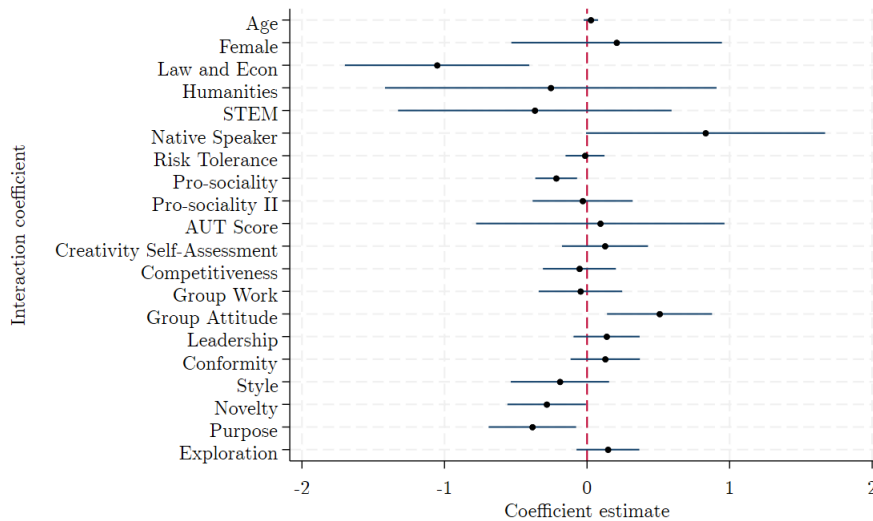
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the individual level.

Figure A10: Heterogeneities Individual Quantity Quantity-PRP after FB



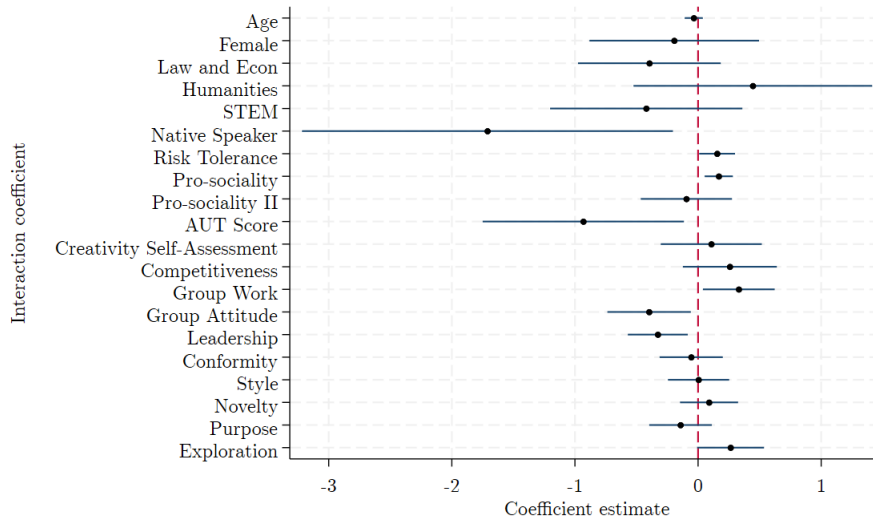
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the individual level.

Figure A11: Heterogeneities Quality Quality-PRP



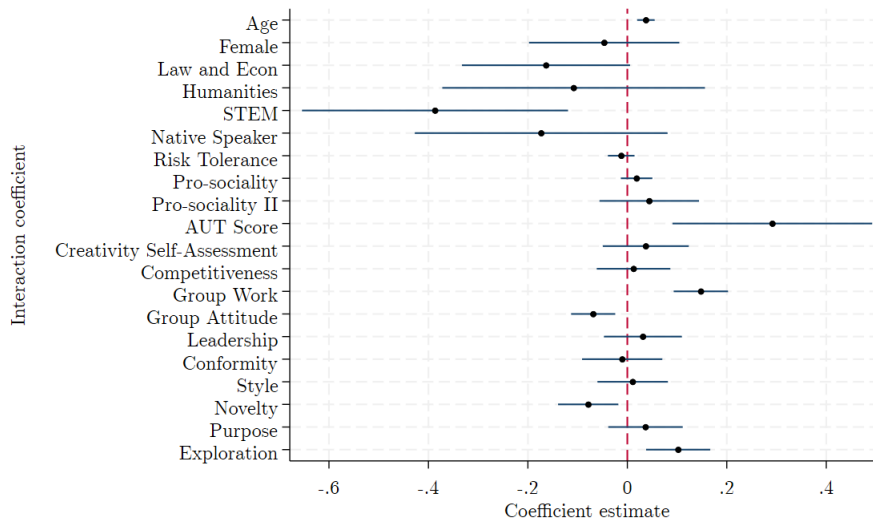
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quality incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A12: Heterogeneities Quality Originality-PRP



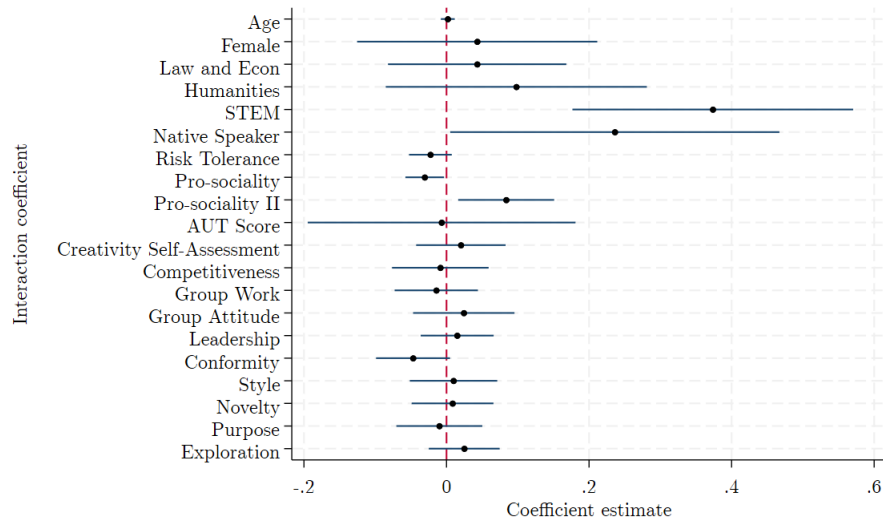
Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Originality incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A13: Heterogeneities Innovation Quantity-PRP



Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quantity incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Figure A14: Heterogeneities Innovation Quality-PRP



Note: The figure displays the point estimates and 90% confidence intervals for interaction coefficients of the Quality incentive and heterogeneity dimensions, using OLS with standard errors clustered at the group level.

Appendix C: Instructions

Figure A15: Instructions (translated from German)

Welcome!

Thank you for participating in this **study**.

You can earn money during this study – how much depends on your decisions.

Please read the following **instructions** carefully.

- You are **not** allowed to talk to other participants during the study.
- Please turn off your **smartphone** and remove it from the table.
- In this study, you - and everyone else - will act as anonymous participants.
- We therefore ask you not to share any information that could lead to your identification.
- Please only use the **functions on the computer** that are intended for the study.

Failure to comply with these rules will result **in exclusion from the study and all payments**.

Next

Study

You have been randomly assigned to a **group with two other participants**, which will remain unchanged throughout the study.

The study consists of three tasks: **Word Task 1**, **Decision**, **Word Task 2**

Word Task 1 is the main component of the study and consists of 6 rounds.

The other two tasks take a maximum of 5 minutes each.

In addition, you will complete two short questionnaires.

The amount of your payout depends on your performance and your decisions.

You will only receive the payout if you complete the entire study.

The payment will only be made after all participants have completed the study.

Next

Figure A16: Instructions (translated from German)

Word Task 1

The **goal** is to work together in your group

- create as many clue pairs as possible,
- that are original,
- and can be used by a person to guess a **target word**.

A clue pair consists of **two individual words**. You play a total of **6 rounds**, i.e. 6 different target words.

Three particularly obvious **taboo words** are shown at the beginning of each round. These words must **not** be part of your clue pairs.

Next

Word Task 1

Each round is divided into **five phases**:

1. Everyone in your group will be told which target word they can create clue pairs for.
2. Without communicating with each other, each group member **first creates** clue pairs on their own.
3. Then the group members **rate** (👍 🗣️) **and comment** on each other's previously created clue pairs.
4. After receiving the feedback, each group member can **revise** their own clue pairs.
5. The group decides together which **final clue pair** will be submitted for evaluation (see next page).

After each round, you will be informed of the result of the previous round.

Next

Word Task 1

Validity check

- In each round, it is manually checked whether the clue pairs created by the group are **valid**, i.e. whether they are related to the target word in some way.
- Invalid clue pairs will be **removed** and will not be included in the evaluation.

Evaluation according to the following criteria

- **Quantity:** The total number of valid clue pairs created by the group.
- **Quality:** Usefulness of the final clue pair, measured by whether and how quickly a person from this room (who is not part of a group) can guess the target word with it. The guessing person changes in each round.
- **Originality:** Novelty of the final clue pair, evaluated by an algorithm that calculates how far the clue pair is from the target word in all meanings and uses (semantic distance).

Next

Figure A17: Instructions (translated from German)

Example Round **Word Task 1**

Target word: *Time*

Taboo words that must **not** be part of the clue pairs:

Present, past, hour

Clue pairs of an example group: *clock + minute, moment + second, infinity + hands*

Final clue pair of the example group: *clock + minute*

Evaluation of the round:

Quality: The target word was guessed in **the 1st attempt** (out of 3 possible).

Originality: **9** out of 10 points on the originality scale (the higher the score, the more original the clue pair)

Quantity: **3** valid clue pairs have been created.

Next

Word problem 1

In addition to the manual validation **at the end of each round**, the following clue pairs are automatically considered **invalid** and will be removed immediately after entry.

Invalid clue pair:

- The clue pair contains more than a **single word** per clue.
 - Example: The clue pair "*very dangerous*" + "*green*" is not allowed because it contains more than two words in total.
- The clue pair contains the target word in a **different spelling**.
 - Example: *Fussball* is not allowed if the target word is *Fußball*.
- The clue pair contains the target word in a **different language**.
 - Example: *Black* is not allowed if the target word is *schwarz*.
- The clue pair contains a word with the same **root word as the target word or one of the three taboo words**.
 - Example: *Princess* is not allowed if the target word is *prince*.
- The clue pair **contains a made-up word**.
 - Example: *Sweetthing* is not allowed if the target word is *cake*.
- The clue pair contains a clue with **special characters or numbers**.
 - Example: *#sweet** is not allowed if the target word is *chocolate*.

Next

Figure A18: Payment Schemes by Treatment (translated from German)

Payment Scheme **Word Task 1**

Each group member receives the following for **Word Task 1**:

- **Participation fee:** 5€
- **Bonus:** up to 15€, based on your group's performance in the **Quality** dimension compared to the **4 other groups** in this session.

In each round, your group receives points if a person from this room (who is not part of a group) guesses the target word based on your final clue pair.

- Guessing on the 1st attempt: **3 points**
- Guessing on the 2nd attempt: **2 points**
- Guessing on the 3rd attempt: **1 point**

If the target word is not guessed within three attempts, your group **will not receive any points**. After 6 rounds, the points from all rounds are added together.

The bonus is awarded based on the total score of your group compared to the other groups:

Place	Your bonus	Your total payout
1.	15€	20€
2.	10€	15€
3.	7€	12€
4.	5€	10€
5.	3€	8€

Next

All payouts will only be made if you complete the entire study. If there is a tie after 6 rounds, the total prize will be split.

Payment Scheme **Word Task 1**

Each group member receives the following for **Word Task 1**:

- **Participation fee:** 5€
- **Bonus:** up to 15€, based on your **group's** performance in the quantity dimension compared to the **4 other groups** in this session.

In each round, your group receives **one point for each valid clue pair** created by you and your fellow players. After 6 rounds, the points from all rounds are added together.

The bonus is awarded based on the total score of your group compared to the other groups:

Place	Your bonus	Your total payout
1.	15€	20€
2.	10€	15€
3.	7€	12€
4.	5€	10€
5.	3€	8€

Next

All payouts will only be made if you complete the entire study. If there is a tie after 6 rounds, the total prize will be split.

Payment Scheme **Word Task 1**

Each group member receives the following for **Word Task 1**:

- **Participation fee:** 5€
- **Bonus:** up to €15 based on your group's performance in the **Originality** dimension compared to the **4 other groups** in this session.

In each round, the **final clue pair** selected by your group receives an **originality score** from an algorithm that calculates how far the clue pair is from the target word in all meanings and uses (semantic distance).

The originality rating is based on a **scale of 1 to 10 points**. The higher the score, the more original the pair. After 6 rounds, the points from all rounds are added together.

The bonus is awarded based on the total score of your group compared to the other groups:

Place	Your bonus	Your total payout
1.	15€	20€
2.	10€	15€
3.	7€	12€
4.	5€	10€
5.	3€	8€

Next

All payouts will only be made if you complete the entire study. If there is a tie after 6 rounds, the total prize will be split.

Figure A19: Comprehension Checks by Treatment (translated from German)

Everything clear?

Please answer the following questions.

Question 1: What three criteria are used to evaluate the clue pairs?

- Length, relevance, originality
- Speed, number, variety
- Quantity, quality, originality

Question 2: How many players form a group together?

- Two
- Three
- Five

Question 3: Which clue would be valid in a clue pair for the target word "Zeit"?

- Time
- Point in time
- Zeeit
- none of the above

Question 4: What happens if a group member's clue pair contains a taboo word?

- It will be taken into account for the evaluation if no one complains
- It will be removed and not scored
- It will be saved for the next round

Question 5: What is the maximum amount of money you can earn on **Word Task 1**?

- 13€ fixed
- €5 fixed plus up to €15 bonus for most clue pairs created
- €5 fixed plus up to €15 bonus for the highest quality clue pairs
- €5 fixed plus up to €15 bonus for the most original clue pairs

Question 6: How is the quality of a final clue pair measured?

- Based on the number of pairs created
- Using an algorithm that evaluates semantic distance
- Based on the number of attempts it takes a person to guess the target word

Question 7: What happens if the target word is guessed on the 1st attempt based on your final clue pair?

- No points
- 3 points
- 1 point

Next

► Summary of the rules

Everything clear?

Please answer the following questions.

Question 1: What three criteria are used to evaluate the clue pairs?

- Length, relevance, originality
- Speed, number, variety
- Quantity, quality, originality

Question 2: How many players form a group together?

- Two
- Three
- Five

Question 3: Which clue would be valid in a clue pair for the target word "Zeit"?

- Time
- Point in time
- Zeeit
- none of the above

Question 4: What happens if a group member's clue pair contains a taboo word?

- It will be taken into account for the evaluation if no one complains
- It will be removed and not scored
- It will be saved for the next round

Question 5: What is the maximum amount of money you can earn on **Word Task 1**?

- 13€ fixed
- €5 fixed plus up to €15 bonus for most clue pairs created
- €5 fixed plus up to €15 bonus for the highest quality clue pairs
- €5 fixed plus up to €15 bonus for the most original clue pairs

Question 6: How is the originality of a final clue pair measured?

- Based on the number of pairs created
- Using an algorithm that evaluates semantic distance
- Based on the number of attempts it takes a person to guess the target word

Next

► Summary of the rules

Everything clear?

Please answer the following questions.

Question 1: What three criteria are used to evaluate the clue pairs?

- Length, relevance, originality
- Speed, number, variety
- Quantity, quality, originality

Question 2: How many players form a group together?

- Two
- Three
- Five

Question 3: Which clue would be valid in a clue pair for the target word "Zeit"?

- Time
- Point of time
- Zeeit
- none of the above

Question 4: What happens if a group member's clue pair contains a taboo word?

- It will be taken into account for the evaluation if no one complains
- It will be removed and not scored
- It will be saved for the next round

Question 5: What is the maximum amount of money you can earn on **Word Task 1**?

- 13€ fixed
- €5 fixed plus up to €15 bonus for most clue pairs created
- €5 fixed plus up to €15 bonus for the highest quality clue pairs
- €5 fixed plus up to €15 bonus for the most original clue pairs

Question 6: How is the quantity of clue pairs measured?

- Based on the number of pairs created
- Using an algorithm that evaluates semantic distance
- Based on the number of attempts it takes a person to guess the target word

Next

► Summary of the rules

Figure A20: Idea Generation (translated from German)

Clue Pairs

The **target word** in this round is:

Wind

The three taboo words are:

- **Blow**
- **Breeze**
- **Squall**

Please enter the clue pairs that come to mind for this word:

(one clue per text field, each word may only appear in one clue pair)

 +

▶ Add a second clue pair

Next

▶ Summary of the rules

Figure A21: Evaluation (translated from German)

Discussion

The following **Clue Pairs** were created by your fellow players.

Rate how good you think each clue pair is and make **Suggestions** for which clues you would replace.

1. Hello + World

▼ Replace clue

Replace the

first word by: second word by:

2. Test + Yes

▶ Replace clue

Next

▶ Summary of the rules

Figure A22: Revision (translated from German)

Clue Pair

The **target word** in this round is:

Wind

The three taboo words are:

- **Blow**
- **Breeze**
- **Squall**

Here are your previously submitted clue pairs with **feedback** from your group members.

You can now **revise** them or replace them completely and create **new Clue Pairs**.

1. Clue pair:

hello + world

Rating from your group members: 👍👍

Replacement suggestion for the **first** word: **goodbye**

2. Clue pair:

test + yes

Rating from a group member: 👍

Next

► Summary of the rules

Figure A23: Selection (translated from German)

Voting

Time remaining for this page: **1:29**

The following **Clue Pairs** were created by you and your fellow players.

You now have **90 seconds** to decide on a final clue pair. If a **majority (at least 2 players)** votes for a clue pair, it will be used for the originality and quality assessment. Otherwise, a clue pair will be randomly selected.

You can coordinate within your group via the **chat**.

Selection	My Vote
hello + world	<input type="radio"/>
test + yes	<input type="radio"/>

Chat

► Summary of the rules

Figure A24: Dictator Game (translated from German)

Decision

Thank you for participating in **Word Task 1!** The second task is now to divide money once between yourself and another randomly selected person in this room.

You are in the role of the "allocator". Another person has the role of the "recipient".

You have 10€ at your disposal, the recipient has 0€.

You can decide **how much of your €10 you want to transfer to the recipient.** You can choose any amount between €0 and €10. The recipient receives the amount you decide to transfer; You will receive the amount you decide not to transfer and thus keep.

Payout: This decision will be made by a total of 15 participants in this session, **one of the 15 decisions will be drawn randomly** and the €10 will be paid out to the allocator and recipient at the end of the study according to the chosen distribution.

The randomly selected recipient is someone from your group.

How much of your €10 do you want to transfer to the recipient in your group?

 €

Next

Figure A25: AUT (translated from German)

Word Task 2

Time remaining for this page: **42:15**

You now have a total of five minutes to think of as many **alternative uses** as possible for two items.

Example: For the object *brick*, alternative uses could be e.g. *doorstop* or *paperweight*.

The first item is a **padlock**.



Provide one alternate usage **per row**.

The second item is a **hammer**.



Provide one alternate usage **per row**.

Figure A26: Quality Assessors Instructions (translated from German)

Word Task 1 and Evaluation

Word Task 1: The goal is to guess a target word in a maximum of **three attempts** based on a given clue pair.

A clue pair consists of **two individual words**. In total, you will try to guess **6 different target words**.

Evaluation The goal is to check the validity of clue pairs for each target word and to evaluate their creativity.

The clue pairs are created **by groups** in this session. They are confronted with three particularly obvious taboo words that they are not allowed to use.

In each new round, you will receive the clue pairs of a different group that you have to **guess, evaluate and check**.

Next

Example Round Word Task 1 and Evaluation

Target word: *Time*

Word Task 1:

Clue pair of an example group: *clock + minute*

1. Guess: Pointer

False!

2nd guess: Time

Correct!

2 points earned!

Validity Check

Clue pairs of another example group: *clock + minute, moment + ephemeral, infinity + hands, the + the*

the + the is **invalid** because there is no connection to the target word *time*.

Creativity Evaluation:

Clue pairs: *clock + minute, moment + ephemeral, infinity + hands*

Each valid clue pair is rated individually on a scale of 1-10 according to usefulness, originality and overall creativity.

Next

Figure A27: Quality Assessors Payment Scheme / Comprehension Checks (translated from German)

Payment Scheme **Word Task 1**

Your payout is made up of two components:

- **Participation fee:** 5€
- **Bonus:** up to €15, based on your performance compared to the 4 other guessers in this session.

In each round, you get points when you guess the target word.

- Guessing on the 1st attempt: **3 points**
- Guessing on the 2nd attempt: **2 points**
- Guessing on the 3rd attempt: **1 point**

If you don't guess the target word within 3 attempts, you **won't get any points**. It must be exactly the *one* target word for which the groups have created the clue pairs.

After 6 rounds, the points from all rounds are added together.

The bonus is awarded based on the total number of points compared to the other participants:

Place	Your bonus	Your total payout
1.	15€	20€
2.	10€	15€
3.	7€	12€
4.	5€	10€
5.	3€	8€

Next

All payouts will only be made if you complete the entire study. If there is a tie after 6 rounds, the total prize will be split.

Everything clear?

Please answer the following **questions**.

Question 1: How many attempts do you have to guess a target word?

- Unlimited attempts
- Three attempts
- Only one attempt

Question 2: How many points do you get if you guess the target word directly on the 1st attempt?

- 1 point
- 2 points
- 3 points

Question 3: What is the maximum amount of money you can earn on **Word Task 1**?

- 20€
- 15€
- 10€

Question 4: How many target words can you guess in total?

- 5 words
- 6 words
- 10 words

Question 5: What does the amount of the bonus depend on?

- Your achieved points in comparison to those of the other guessers
- The number of clue pairs you have created
- The quality of your evaluation of the clue pairs

Next

► Summary of the rules

Figure A28: Guessing / Validity Check (translated from German)^a

1. Guess

This is the **clue pair** you received from one of the groups:

hello + world

Please enter your first **guess** for the target word.

Your guess:

Next

► Summary of the rules

Evaluation Validity

The target word was **wind**. Thank you for participating in **Word Task 1**.

Evaluation: Now we ask you to check the clue pairs submitted by another group.

Please mark all clue pairs as invalid if you cannot **see any connection** to the target word.

Important: This step is not about how useful or original the pairs are, but only about whether there is a basic connection to the target word. Please check carefully so as not to rule out any potentially valid pairs.

The clue pairs of one group were:

Selection	Invalid
hello + world	<input type="checkbox"/>
test + yes	<input type="checkbox"/>

Next

► Summary of the rules

^a Following the first of 32 sessions (control condition), the instruction “Important: This step is not about how useful or original the pairs are, but only about whether there is a basic connection to the target word. Please check carefully so as not to exclude any potentially valid pairs.” was added to prevent quality assessors from incorrectly flagging too many clue pairs as invalid.

Figure A29: CRAT (translated from German)

Word Task 2

You are given **three words**.

Your task is **to find a fourth word** that is related to all three given words.

Example: *Screen - Darkness - Flower*

Solution: *Sun*

In total, you have **5 minutes** for 5 terms. A timer will show you the time remaining.

Next

Word Task 2

Time remaining for this page: **41:53**

1. Wall - Bracelet - Time

2. Film - Sheet - Spray

3. Worm - Shelf - Broadcast

4. Worker - Fire - Ant

5. Flower - Rag - Cake

Appendix D: Detailed Variable Descriptions

The dataset includes demographic characteristics, individual traits, creativity measures, and task-related variables, separately collected for ideators and quality assessors. All variables are described below in full detail.

Ideators

Demographic Variables

Age is measured in years. Female is coded as 0 for male, 1 for female, and 0.5 for diverse participants. Participants reported their field of study from a comprehensive list of disciplines. For analysis, the three largest categories—Law and Economics, Humanities, and STEM—are represented as separate binary dummy variables, coded 1 if the participant belongs to the respective field and 0 otherwise. Native Speaker indicates whether the participant is a native speaker of the study language, German (binary).

Individual Traits

Risk Tolerance is assessed on a scale from 0 to 10, with higher values indicating greater willingness to take risks (cf. [Dohmen et al., 2011](#)). Pro-sociality is measured using one round of a Dictator Game, with scores ranging from 0 to 10; higher values reflect more pro-social behavior. Pro-sociality II, assessed on a 5-point Likert scale from -2 to 2, captures the willingness to make sacrifices for the group’s well-being. Competitiveness is a standardized mean index (Cronbach’s α : 0.81) derived from two items evaluating competition-driven motivation on a -2 to 2 scale: ”I enjoy competing with others” and ”I perform better when I compete against others”.

Group-related attitudes

Group Work, measured on a -2 to 2 scale, indicates the willingness to work collaboratively rather than alone. Group Attitude (-2 to 2) captures the belief that groups are more productive than individuals. Leadership Attitude (-2 to 2) measures the extent to which participants endorse the necessity of a designated leader for effective group performance. Conformity, measured on a -2 to 2 scale, reflects the tendency to align with majority opinions to avoid social conflict.

Creativity Measures

The Alternate Uses Task (AUT) Score, ranging from 1 to 5, captures originality in divergent thinking, with higher scores indicating more original responses. The scores are automatically calculated using Ocsai (Open Creativity Scoring with Artificial Intelligence), a fine-tuned set of GPT-based large language models ([Organisciak et al., 2023](#)). Creativity Self-Assessment is a standardized mean index (Cronbach’s α : 0.78) based on three items rated on a 5-point Likert scale from -2 to 2, measuring perceived creativity: ”I think that I am an above-average creative person”, ”It is important to me to be a creative person”, and ”I am good at finding creative solutions to problems” (cf. [Karwowski et al., 2018](#)). Creative Style (-2 to 2) reflects the endorsement of a trial-and-error approach to creative work, indicating a preference for unstructured, exploratory processes without predetermined plans. Novelty Seeking (-2 to 2) measures the propensity to seek out new and unfamiliar experiences. Creative Purpose (-2

to 2) captures the degree to which individuals associate creative activity with purposeful or goal-directed intentions. Creative Exploration (-2 to 2) represents the willingness to engage in open-ended creative processes, accepting the inherent risk of losing direction. The four items on creative style are inspired by [Charness and Grieco \(2023\)](#).

Task-Related Measures

All behavioral measures are recorded for each round of the task, whereas survey-based measures are assessed only at the end of the experiment, with the exception of the communication assessment, which is recorded after each round. Effort is operationalized through task duration (in seconds) across the three phases—idea generation, feedback, and revision—and the intensity of engagement, measured by the number of generated ideas, modified ideas, and suggestions provided per minute. A timeout occurs after three minutes in the idea generation phase and after 90 seconds in the revision and feedback phases. The share of changed/removed pairs captures the proportion of clue pairs modified or deleted during the feedback process, while the share of invalid pairs measures the proportion of invalid pairs among all generated pairs before and after feedback and revision. The share of given positive ratings reflects the proportion of positive ratings among all provided evaluations.

The number of suggestions represents the raw count of received suggestions, while the share of used suggestions reflects the proportion of implemented suggestions among those received. Number of Chat Messages captures the total chat messages sent during the selection phase. The Chat Sentiment variable is derived using a sentiment analysis model applied to the chat data. The analysis employs the German Sentiment Classification model ([Guhr et al., 2020](#)), based on Google’s BERT architecture and trained on approximately 1.83 million German-language samples from various domains (e.g., X (Twitter), Facebook, reviews). Each message is preprocessed and classified as positive, negative, or neutral, with sentiment scores ranging continuously between -1 and +1, capturing both polarity and emotional intensity.

Heterogeneity Originality and Heterogeneity Quantity measure within-group variability (standard deviation) in the average originality and the number of generated pairs per individual before feedback and revision, respectively.

All survey-based measures are assessed on a five-point Likert scale from -2 (“strongly disagree”) to +2 (“strongly agree”). Enjoyment measures the extent to which ideators enjoy the task. Leadership captures the degree to which one person is perceived as dominant in leading the group’s ideation process, and Peer Pressure measures the extent to which participants feel motivated to contribute equally to others to avoid underperforming. Difficulty Quality, Difficulty Originality, and Difficulty Quantity assess perceived difficulty in generating high-quality, original, or numerous ideas, respectively.

Communication is a standardized mean index (Cronbach’s $\alpha = 0.76$) derived from two items evaluating coordination after each round: “Communication within my group was effective” and “In the end, we selected the best clue pair”. Group Effort is a standardized mean index (Cronbach’s $\alpha = 0.73$) based on two items measuring the perceived effort of other group members: “All group members made an effort” and “All group members contributed equally.” Teamwork is a standardized mean index (Cronbach’s $\alpha = 0.74$) derived from three items assessing collaboration: “My group developed a real team spirit”, “Our collaboration in the group was productive”, and “Our individual skills and perspectives complemented each other well”. Feedback Assessment is a standardized mean index (Cronbach’s $\alpha = 0.71$) derived from four items evaluating the feedback process: “I consider the feedback I gave to be

helpful,” “My group members were very willing to accept feedback,” “I felt that my ideas and I were valued,” and “My group members were very willing to accept feedback.”

The following variables are assessed only in the treatment conditions. Pro-social Group is a standardized mean index (Cronbach’s $\alpha = 0.76$) derived from two items measuring pro-social attitudes and perceived group norms: “Since our pay depended on the group’s success, we supported each other strongly”, and “Since our pay depended on the group’s success, we all felt obligated to do our best”. Motivation Competition and Motivation Payment (-2 to 2) capture motivation driven by competition and monetary incentives, respectively. Tradeoff measures the perceived focus on one creative dimension (quantity, quality, or originality) at the expense of the others, while Self-efficacy captures participants’ belief in their ability to improve their ideas through additional effort in terms of quantity, quality, or originality.

Quality Assessors

Demographic Variables

Age, gender, academic background, and native speaker status are measured identically to ideators.

Individual Traits

Risk tolerance and competitiveness are also assessed identically to ideators.

Creativity Measures

CRAT (Compound Remote Associates Task) performance, ranging from 1 to 5, assesses convergent creative thinking, with higher scores reflecting stronger performance. The score equals the number of problems that the assessor solves correctly, with all problems drawn from [Landmann et al. \(2014\)](#). Creativity Self-Assessment, Creative Style, Novelty Seeking, Creative Purpose, and Creative Exploration are measured identically to ideators.

Task-Related Measures

Quality assessors’ strategies in guessing the target words are recorded across five distinct measures (Strategy I–V) on a 5-point Likert scale. Strategy I indicates whether the assessor first considered the two clue words separately before searching for a connection. Strategy II measures the extent to which assessors intuitively searched for associations between the clue words. Strategy III captures attempts to find an overarching concept or theme uniting both clue words. Strategy IV reflects deliberate attempts to think outside the box to discover unusual or non-obvious connections. Strategy V measures the tendency to select the first word that comes to mind. Taboo Words records whether assessors took into account the existence of taboo words when forming responses. Difficulty Generation and Difficulty Selection, both measured on a -2 to 2 scale, capture perceived difficulty in generating responses from the clue pairs and making the final choice, respectively. Enjoyment and motivation derived from competition and monetary incentives are measured identically to ideators.

Appendix E: Mediation Analysis

Mediation analysis provides a framework to decompose the total effect of a treatment into components that operate directly and indirectly through intermediate variables (mediators). Conceptually, a treatment T influences an outcome Y both directly and indirectly through a mediator M , such that the causal structure can be represented as $T \rightarrow Y$ and $T \rightarrow M \rightarrow Y$ (Baron and Kenny, 1986). The potential outcomes framework allows this structure to be expressed rigorously in counterfactual terms. Let $t \in 0, 1$ denote the treatment status (presence versus absence of an incentive treatment), $M(t)$ the mediator under treatment t , and $Y(t, m)$ the outcome that would be observed under treatment t and mediator m . Within this framework, the total effect (TE) of the treatment is defined as the difference in outcomes under treatment and control, $TE = Y(1) - Y(0)$. This total effect can be decomposed into the natural direct effect (NDE) and the natural indirect effect (NIE): the natural direct effect, $NDE = Y(1, M(0)) - Y(0, M(0))$, captures the portion of the treatment effect that operates independently of the mediator M , while the natural indirect effect, $NIE = Y(0, M(1)) - Y(0, M(0))$, quantifies the effect transmitted through the mediator. By construction, the total effect is $TE = NDE + NIE$. The proportion mediated, defined as NIE/TE , summarizes the share of the total effect that is transmitted through the mediator. This framework accommodates a wide range of variable types and modeling approaches and enables causal interpretation under standard identification assumptions (no unmeasured confounding of the treatment–mediator, treatment–outcome, and mediator–outcome relationships, together with consistency and positivity; see Nguyen et al. (2022)). As treatment randomization does not resolve unobserved mediator–outcome confounding, this may bias direct and indirect effect estimates in my setting, whereas the total effect remains unbiased.