Economics Bulletin

# Volume 41, Issue 4

## Creative solutions: Expertise versus Crowd Sourcing

Uri Gneezy
*University of California, San Diego*

Katharina Laske
*University of Cologne*

Marina Schröder
*Leibniz University Hannover*

## Abstract

We compare creativity by experts and non-experts. Experts are more expensive, and hence, for a given budget, one has to often choose between recruiting few experts or many non-experts to find a creative solution to a task. In our set-up, we find that the best solution offered by few non-experts was better than the one offered by an expert.

We thank the German Research Foundation (DFG) through the research unit "Design and Behavior – Economic Engineering of Firms and Markets" (FOR 1371) and the individual research grant 456592273 for financial support.

**Citation:** Uri Gneezy and Katharina Laske and Marina Schröder, (2021) "Creative solutions: Expertise versus Crowd Sourcing", *Economics Bulletin*, Vol. 41 No. 4 pp. 2580-2586.

**Contact:** Uri Gneezy - ugneezy@uscd.edu, Katharina Laske - laske@wiso.uni.koeln.de, Marina Schröder - schroeder@wipol.uni-hannover.de.

**Submitted:** March 30, 2021. **Published:** December 29, 2021.

# 1. Introduction

Imagine that you are about to publish a book and need to find a title. You have a budget to hire help with finding a good title. With this budget you can either hire one expert in the field, or a few non-experts. Your sole interest is to have the best possible title for your book. Who is more likely to help produce this: the expert or the non-experts? Think of a copywriting service as an example. Copywriting is an essential component of marketing activities and is often outsourced either to freelancers or to marketing agencies.

Firms spend large amounts of money on creativity, as evidenced by large research and development (R&D) budgets. For many of these creative developments, expertise and a deep understanding of science and technology is necessary. However, companies also expend large sums of money in areas such as copywriting, to acquire creative expertise from external providers. For instance, the world's ten biggest advertising agencies had a combined gross income of more than $25 billion in 2017.[1] Yet, it is less clear how important expertise is in such areas of idea generation.

One prevailing view is that creativity needs knowledge as a source of ideas from which novel products can then be generated (Cropley 2006). Amabile (1998) notes that domain-specific expertise is the "raw material" for creative ideas and people who are familiar with the focal field will be more successful in finding good solutions. A different prevailing view, however, argues experts rely on accustomed habits and think in familiar patterns. Likely bound to the current thinking in their field, divergent thinking may be blocked, reducing the potential for generating creative ideas. Thus, to return to our thought experiment, it is unclear whether experts will produce higher quality creative output than non-experts and whether employing experts in a domain such as marketing is efficient in terms of cost-benefit calculations.

Our argument is simple. Even if, on average, experts produce a more creative outcome, average outcomes are not the right measure. Since only the best ideas—the positive outliers—matter, it might be better to have more solutions, even if average output is lower. To this end, we conduct a lab-in-the-field experiment in which we use a real-effort task that allows us to derive an objective measure for the quality of ideas. In our experiment, experts consist of members from the platform upwork and non-experts were obtained via MTurk.

We add to the growing literature in experimental economics on creativity. Our paper is related to the literature on idea generation and incentives (Eckartz, Kirchkamp, and Schunk 2012; Erat and Gneezy 2016; Gibbs *et al.* 2017; Bradler, Neckermann, and Warnke 2019), the dimensions that should be incentivized (Kachelmeier, Reichert, and Williamson 2008; Laske and Schröder 2018), how incentives interact with the type of task (Charness and Grieco 2018), how the magnitude of the incentive effects the creative outcome (Ariely *et al.* 2009), how the time horizon of the incentive - short-term vs. long-term - influences individuals ideation performance (Azoulay, Zivin, and Manso 2011; Ederer and Manso 2013) and group work and ideation (Grözinger *et al.* 2020; Gneezy *et al.* 2021). Two relevant studies investigate the role of expertise in a context in which profound knowledge is indispensable. Jeppesen and Lakhani (2010) examine successful solutions in science problem-solving contests, finding that the provision of a winning solution is positively related to increasing distance between the solver's field of technical expertise and the focal field of the problem. Franke, Poetz, and Schreier (2014) study whether experts in the focal area provide better ideas for new product development than experts in a different but similar market. They observe that product ideas from the latter group have a lower potential for immediate use, but have substantially higher levels of novelty. Similar to Jeppesen, and Lakhani (2010), the authors find that this effect is particularly pronounced when the distance between the two similar markets increases. Building on this literature, we study the effectiveness of experts in the creative domain. Our results

[1] https://co.agencyspotter.com/50-largest-marketing-companies-in-the-world/

provide a rationale for why an increasing number of firms have chosen to utilize crowdsourcing for idea generation purposes.

# 2. Experimental Design and Procedure

We asked participants to come up with a creative title for a 70-second video clip on various forms of cheating. Only titles related to the content of the video were evaluated. We informed participants that we will collect several titles from different people and that the creator of the best title will be rewarded a bonus of $400. We specified that the quality of the title will be evaluated according to the number of clicks the title generated among laboratory participants. See Appendix for the instructions for experts. Non-experts received identical instructions except of the payment section.

For quality elicitation of the titles, we recruited 600 student raters via ORSEE (Greiner 2004). The student raters were blind to the source of the title ideas (experts vs. non-experts) and did not take part in any previous related experiments. We provided each rater with four randomly drawn titles and asked them to first click on the title of the video they want to watch and then watch the video. Thus, raters made real decisions involving opportunity cost of time.

Quality of a creative title was quantified by measuring click rate. On average, each title was seen by 20 raters, each time in a different combination with three other randomly selected titles. We derive the quality of each title by the fraction of raters who clicked on it, thus, quality ranges from 0 to 1. For example, if a title was clicked by 8 out of 20 raters, it would receive a quality score of 0.4. See Table I for examples of the three best and the three worst titles in our experiment. After watching the video, we asked raters to assess the fit between the video and the title on a 5-point scale (1 = very poor to 5 = very good).

TABLE I: EXAMPLES OF TITLES WITH DIFFERENT QUALITY SCORES

| **Best three titles** | **Quality score** | **Worst three titles** | **Quality score** |
|---|---|---|---|
| 1. Liars, Cheats & Thieves: The Truth about Human Character | 0.61 | 1. ALL KINDS OF CHEATERS | 0.05 |
| 2. Immoral Statistics: Breaking Society's Rules | 0.60 | 2. We know you cheat! | 0.05 |
| 3. Cheating, the new social norm? | 0.55 | 3. Video infographic on cheating | 0.05 |

In October 2015, we recruited 30 experts on the platform upwork. People who register as experts on this platform are thoroughly vetted and reviewed. Only when information and professional credentials are verified (e.g. by entering codes for test results such as the Cambridge Certificate) is the profile approved. Upwork is the world's largest freelancing website where independent professionals from all over the world offer their expertise to potential customers. The freelancers from upwork hired for our experiment all specialized in *copywriting*. To restrict the definition even further, we only recruited freelancers with an hourly wage of at least $60, with the assumption a higher hourly wage signals a greater degree of expertise.[2] To minimize heterogeneity between the two treatment samples, we only recruited US residents. In treatment *expert,* workers received a fixed amount of $60 per created title, independent of their performance or the time spent working on the task, plus the $400 bonus. Given, that we only recruit experts who require a minimum hourly wage of $60, the fixed payment we offer for performing the task is below their expected hourly wage. For treatment

[2] For each job category, upwork provides price ranges for entry, intermediate and expert level freelancers. For hiring people on expert level in the job category *copywriting,* the platform suggests a payment of more than $46.50/hour.

*non-expert*, we recruited 90 workers on Amazon's Mechanical Turk marketplace (Mturk), which is a standard source for experiments in psychology, marketing and economics. We announced the job on Mturk in the same way as we did on upwork, with a description stating that participants would earn a fixed amount of $1.5 for performing the task, plus additional bonus opportunities.

Both experts and non-experts were recruited to perform the task on a fixed payment schedule. Thus, payment did not depend on the time spent working on the task. While we calibrated incentives according to expected hourly wages, we cannot fully rule out any differences in performance arising due to the level of incentives provided.[3]

# 3. Experimental Results

Do experts perform better than non-experts in a task where prior knowledge is not necessarily required, and only the best entries are important? As explained above, we define *quality* by the fraction of raters who clicked on a particular title in order to watch the related video. This performance measure is highly relevant for copywriting tasks in practice. By design, the average overall quality is 0.25. The minimum quality in our experiment was 0.05 and maximum quality 0.61. We observe slightly higher quality in *expert* (0.268) compared to *non-expert* (0.244) titles. This difference, however, is not statistically significant on a conventional level (U-Test, p=0.28; we report two-sided p-values in the entire paper). Figure 1 illustrates the results, with the vertical lines representing the means. We also elicit the fit between the video and the title on a 5-point scale (1 = very poor to 5 = very good). The fit of titles generated by non-experts (3.87) and experts (3.80) are not statistically different from each other on a conventional level (U-Test, p=0.37).

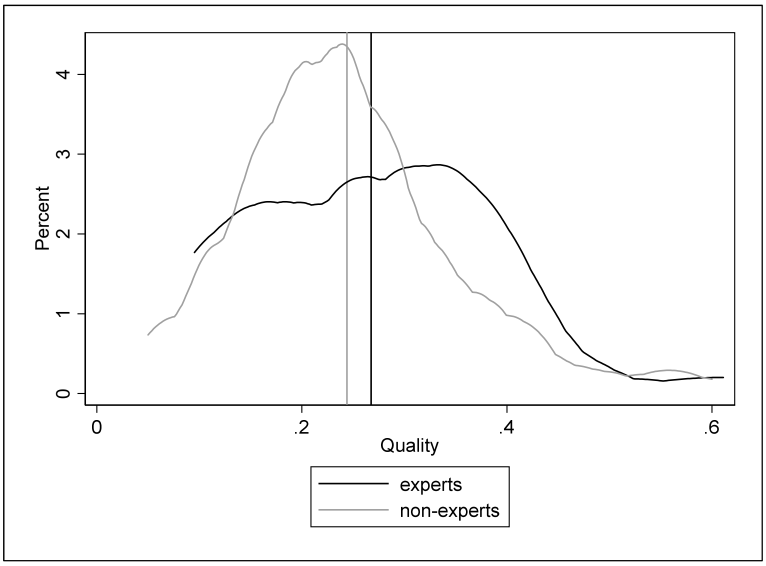


FIGURE 1: CLICK RATES BY TREATMENT

Table II reports the results of an OLS regression analysis with average quality in columns (1) and (2) as dependent variables. In line with the non-parametric analysis, we observe a slight but insignificant increase in the performance of experts.

Overall, we find no evidence that experts perform better than non-experts in this simple task. One reason for the lack of difference in performance may be differences in opportunity cost of

[3] Paying participants the same amount would have been an alternative approach to calibrating incentives. To do so, we would have had to pay $60 for workers on MTurk. Such high rates on MTurk would have been more than unusual and would have likely caused undesired discussions among participants.

time. Presumably, experts have higher opportunity cost of time as their outside options pay better. As a result, experts may spent less time on the task and may therefore be less likely to generate high quality titles. To analyze such differences, we compare submission time between groups. On average, participants took about half an hour to come up with a final title, with minimum 1.2 minutes and maximum 1380 minutes. However, the difference between non-experts and experts is large: non-experts took, on average, 4.5 minutes, while experts took 105 minutes on average. This difference is highly significant (pairwise U-test, $p<0.001$). As this result on average time is driven in part by outliers, we conduct a median regression analysis to remove this outlier bias. Columns (3) and (4) of Table II display the results of a median regression analysis with submission time as the dependent variable. In line with the results from the non-parametric analysis, column (3) reveals a significant increase in the median submission time whenever an expert created a title. It seems as if experts take longer to come up with creative titles, but that this endeavor does not translate into better performance.[4] This finding is consistent with Amabile's (1996) premise that creativity does not emerge from simply trying harder.

TABLE II: EXPERTISE ON QUALITY, TIME AND PRODUCTIVITY

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Average quality (in percent) | | Average submission time (in min) | |
| Expert | 0.024 | 0.020 | 13.950** | 14.287** |
| | (0.025) | (0.034) | (6.138) | (6.751) |
| Female | | -0.005 | | 0.195 |
| | | (0.020) | | (0.913) |
| Age | | 0.000 | | 0.027 |
| | | (0.001) | | (0.085) |
| Expertise (self-rated) | | 0.006 | | -0.513 |
| | | (0.010) | | (0.325) |
| Education | | yes | | yes |
| Ethnicity | | yes | | yes |
| Constant | 0.244*** | 0.202*** | 2.800*** | 3.760 |
| | (0.011) | (0.055) | (0.392) | (2.931) |
| Observations | 120 | 119 | 120 | 119 |
| R-squared | 0.009 | 0.080 | | |

*Notes:* Estimates in col. (1-2) are based on OLS regression. Estimates in col. (3-4) are based on LAD regression. Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variables: col. (1 and 2): title quality in percent; col. (3 and 4) time between reading the task and submitting a title in minutes. The dummy Expert is equal to 1 if a worker was recruited on the freelancer platform upwork.com and 0 if a worker was recruited on Mturk. Self-rated expertise is a variable ranging from 1 (very low) to 5 (very high). Education is a variable indicating the highest degree of education differentiating between high school diploma, bachelor's degree, master's degree or higher and other. Ethnicity differentiates between white, black, asian, hispanic and mixed.

Turning to the top performers, which is the relevant statistic for success, we observe that out of the ten best titles, only three were created by experts. Our initial conjecture was that experts would have lower variance. However, the standard deviations of experts (0.12) and non-experts (0.10) are not significantly different from each other (Variance ratio test, $p=0.38$). A more

[4] The platform upwork provides proxies for a worker's expertise. In addition to workers' self-reported field of expertise and hourly wage, we also collected information on number of projects completed and hours worked. However, we observe no significant differences for any of these additional measures (i.e., they were not a more powerful predictor for performance).

likely explanation is that of a statistical nature: the more ideas generated, the higher the probability that some of them will be very creative.

Lastly, we investigate whether it is more efficient to hire a few experts or many non-experts to work on a creative task. We account for the different fixed payments and calculate the cost per click as a productivity measure. Bonus aside, experts earned \$60 while non-experts earned \$1.5 per created title. Thus, a click for a title created by an expert on average cost \$11.39 while a click on a title created by a non-expert on average cost \$0.31. This difference is highly significant (U-test, p=0.00). To account for the fact that this result depends on the parametrization of the payments, we calculate the break-even point at which it becomes more efficient to hire experts. Assuming that the level of payment does not affect performance, it would be more efficient to hire experts in our experiment only if non-experts cost more than \$55.95. If performance of non-experts increases in payment, this number would be even larger.

# 4. Discussion and Conclusion

We study the effectiveness of expertise in finding creative solutions in an area in which prior knowledge is not necessarily required. In our setting, we observe that for a given budget, it is more efficient to hire many non-experts instead of a few experts and that the best solutions (i.e. titles) were created by non-experts. We also find that, while experts put significantly more effort (time) into the task, this effort does not translate to better performance, consistent with the literature on creativity (Amabile 1988; Erat and Gneezy 2016). Creative performance is likely a probabilistic function of quantity (Simonton 2003; 2004; Laske and Schröder 2018). We can only provide evidence for the quality of titles generated by workers recruited from the two prominent platforms. We calibrate incentives according to expected payments on these two platforms, where expected payments are substantially higher on upwork as compared to MTurk. Thus, payment for experts is substantially higher compared with payment to non-experts in our experiment.

Our results are in line with the trend of using the creative potential of the population to fuel new product development, adopted by many firms. One prominent example is McDonalds, which instead of contracting food or industry experts in 2014, invited the public to submit ideas for the types of burgers they wanted to be offered in store. People could create their individual burgers online and the rest of the country could vote for the best ones, which were ultimately sold in the branches. Similarly, Threadless, a t-shirt company based in Chicago, has relied on the creative potential of the crowd since it has been founded in 2000. Instead of employing professional designers, the company has asked people to submit designs via their crowdsourcing platform. Once an idea is posted, people start voting for it and leaving comments. Based on the average score and user feedback, about 10 designs are selected each week, printed on clothes and other products and sold worldwide.

## References


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

Amabile, T.M. (1996) *Creativity in context: Update to the social Psychology of Creativity*, Westview Press: New York.

Amabile, T.M. (1998) "How to kill creativity" *Harvard Business Review* **40**(1), 39-58.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009) "Large Stakes and Big Mistakes" *The Review of Economic Studies* **76**(2), 451-469.

Azoulay, P., J.S Graff Zivin, and G. Manso (2011) "Incentives and creativity: evidence from the academic life sciences" *The RAND Journal of Economics* **42**(3), 527-554.

Bradler, C., S. Neckermann, and A. Warnke (2019) "Incentivizing Creativity: a Large-Scale Experiment with Performance Bonuses and Gifts" *Journal of Labor Economics* **37**(3), 793-851.

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

Cropley, A. (2006) "In Praise of Convergent Thinking" *Creativity Research Journal* **18**(3), 391-404.

Eckartz, K, O. Kirchkamp, and D. Schunk. (2012) "How do Incentives Affect Creativity?" CESifo working paper series number 4049.

Ederer, F and G. Manso (2013) "Is pay for performance detrimental to innovation?" *Management Science* **59**(7), 1496-1513.

Erat, S and U. Gneezy (2016) "Incentives for Creativity" *Experimental Economics* **19**, 269-280.

Franke, N., M.K Poetz and M. Schreier (2014) "Integrating problem solvers from analogous markets in new product ideation" *Management Science* **60**(4), 1063-1081.

Gibbs, M., S. Neckermann, and C. Siemroth (2017) "A field experiment in motivating employee ideas" *Review of Economics and Statistics* **99**(4), 577-590.

Gneezy, U., K. Laske, and M. Schröder (2021) "Teams Do Not Outperform Individuals in a Simple Creative Task*" Applied Economics Letters*.

Greiner, B. (2004) An online recruitment system for economic experiments.

Grözinger, N., B. Irlenbusch, K. Laske, and M. Schröder (2020) Innovation and communication media in virtual teams–An experimental study. *Journal of Economic Behavior & Organization* **180**, 201-218.

Jeppesen, L.B and K.R. Lakhani. (2010) "Marginality and problem solving effectiveness in broadcast search." *Organizational Science* **21**(5), 1016-1033.

Kachelmeier, S.J., B.E. Reichert, and M.G. Williamson (2008) "Measuring and Motivating Quantity, Creativity, or Both." *Journal of Accounting Research* **46**, 341-374.

Laske, K and M. Schröder. (2018) *Quality through Quantity - The Effects of Piece-Rate Incentives on Creative Performance*, Mimeo.

Simonton, D.K. (2003) "Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives." *Psychological Bulletin* **129**(4), 475.

Simonton, D.K. (2004) *Creativity in science: Chance, logic, genius, and zeitgeist*, Cambridge University Press.

# Appendix: Instructions Experts

0% completed

In the following you will see a short video (72 seconds). We ask you to come up with a title, which we can use to upload this video on our homepage. The aim is to find a creative title that generates high click rates and is related to the video. We will collect several titles from different people. The best title is the title that generates the highest click rates among people from the internet marketplace MTurk.

You will receive $60 for performing this task. Additionally, the person who comes up with the best title will receive a reward of $400.

Next

Contact – 2015

25% completed


**Please watch the video and come up with a creative title for this video:**

Next