

Referee Report for 25700, “Motivating Effort in Contributing to Public Goods Inside Organizations: Field Experimental Evidence”

This study is a field experiment that studies a pecuniary intervention and two other non-pecuniary interventions that appeal to intrinsic motivation, namely improving the workplace and improving patient care, in a health care setting where the “public good” is participation in a contest for proposals to improve the workplace. There is much to like about the paper. In particular the setting, taking place in a leading hospital’s “Healthcare Transformation Lab,” seems potentially cutting-edge from a perspective of health economics, as we seek to understand how to improve health care delivery in the US. This paper aims to address a general issue of how to motivate effort on public goods via pecuniary and non-pecuniary incentives.

Motivation

- I am sympathetic to the authors’ framing of their paper, but despite this general interest question, I was still left unsure about the paper’s contribution. In their own literature review, they mention that this topic has been studied in both laboratory and field experiments. With respect to field experiments, they mention Della Vigna et al (2016), Erev et al (1993), and Gibbs et al (2014). Although they note differences between their study and the other papers in the setup, I would still like to see more about how both their setup *and* their results together make this paper a major contribution with respect to their motivation.
- As it is currently written, I see no particular advantage of this high-technology, complex, and potentially inefficient setting of cardiac care in a leading hospital other than potentially greater intrinsic motivation, as the authors note, in a “mission-oriented” industry. Of note, the intervention featuring the workplace-oriented framing performed better than the intervention with a framing to improve patient care.

Control and Treatment Arms

- On page 13, it is unclear whether the treatment arms also were aware that they would receive project funding up to \$20,000 for their ideas. Reading that page literally, it seems that they might not. Also, the labeling of Figure 3 seems to denote the control arm as “FUND”. If so, then I think this could be a problem, because the control arm has some value that the treatment arms do not.
- Even if participants in all arms were equally aware of the potential for \$20,000 to fund their idea, I would like some commentary on how to think of this relative to the prize. That is, if an iPad mini with a value of \$400 had such a large effect either instead of or additional to a \$20K cash grant for the project idea, is this informative for the relative weight placed on public vs. private goods? And is this consistent (or does it need to be consistent) with the statement on page 30 that social preferences are equivalent to 45 percent of the private cost of contributing?

Model and Calibration Exercise

- They observe that females respond to the prize more than males and use previous literature that women are less likely to engage in competition to posit that women likely have larger fixed participation costs that can be overcome with cash rewards. I am not sure I see how this maps into their model.

- More specifically, this seems to be at odds with their exercise in Section 7, where they assume that the cost of participation is the same for everyone and moreover is equal to the median hourly wage. If however, they do not observe the costs of participation, which could be more than pecuniary, then it seems that their exercise in Section 7 would not be identified, where they now have three unknowns.
- Relatedly, reconciling that physicians have higher wages but equal participation would need more moving parts in the model.

Effect on Judging

- It seems to me that another interesting analysis would be the effect of the treatments on how many proposals a subject judged in the subsequent phase.
- Given that there is no pecuniary reward for judging, this analysis might be able to provide additional identification or a falsification test. That is, the iPad pecuniary award treatment arm should not induce more judging.

Selection and Quality of Proposals

- When comparing the quality of proposals conditional on submission, a natural question is selection. In a plausible Roy model, one might expect intrinsically higher-quality proposals (with higher chance of winning) to be submitted when the costs are higher. On the other hand, participants may exert more effort to make the same proposals higher quality when there is more reward. These two effects go in opposite directions, so it is unclear whether one would expect the quality proposals conditional on submission to be higher or lower with a treatment.

Exposition, Typos

- In general, I thought that the manuscript was well-written, but there were a few places that either seemed unjustified:
 - Page 6: "...these results are remarkable because they support the hypothesis of a complementarity between the use of pecuniary incentives and the degree of social preferences." I do not see how they reach the conclusion of complementarity. In fact, they impose a specific functional form in Equation (1) and do not (perhaps cannot) test that. This also seems directly contradictory to the thesis of crowding-out discussed briefly on page 33.
- Minor points of exposition
 - Is "voting" the same as scoring or judging the proposals? The two have different connotations usually but seem to be used interchangeably here.
 - I was confused by the units on proposals implied by Table A.5. That table says that the outcome variable is the number of proposals per submission, yet the coefficients are implausibly high if so.
 - "Sorting" seems to be an unclear word. "Selection" is probably more descriptive.
 - It would be clearer if the words "percentage points" were used instead of "percent" when talking about additive treatment effects.
 - Page 26: "liner regression"
 - Page 33: "cautions" should be "cautious"