Economic Games on the Internet: The Effect of $1 Stakes

Ofra Amir[1]*, David G. Rand[2], Ya'akov Kobi Gal [1,3]

1 Department of Information Systems Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel, 2 Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts, United States of America, 3 School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Online labor markets such as Amazon Mechanical Turk (MTurk) offer an unprecedented opportunity to run economic game experiments quickly and inexpensively. Using Mturk, we recruited 756 subjects and examined their behavior in four canonical economic games, with two payoff conditions each: a stakes condition, in which subjects' earnings were based on the outcome of the game (maximum earnings of $1); and a no-stakes condition, in which subjects' earnings are unaffected by the outcome of the game. Our results demonstrate that economic game experiments run on MTurk are comparable to those run in laboratory settings, even when using very low stakes.

Introduction

Online labor markets such as Amazon Mechanical Turk (MTurk) are internet marketplaces in which people can complete short tasks (typically 5 minutes or less) in exchange for small amounts of money (typically $1 or less). MTurk is becoming increasingly popular as a platform for conducting experiments across the social sciences [1–7]. In particular, MTurk offers an unprecedented opportunity to run incentivized economic game experiments quickly and inexpensively. Recent work has replicated classical findings such as framing and priming on MTurk [8–10], found a high level of test-retest reliability on Mturk [10–12], and shown quantitative agreement in behavior between MTurk and the physical laboratory [6,8]. Yet concerns remain regarding the low stakes typically used in MTurk experiments.

In this study, we directly examine the effect of such stakes by comparing un-incentivized play with play involving typical MTurk sized stakes (up to $1) in four canonical economic games - the dictator game, ultimatum game, trust game and public goods game. Our results are consistent with previous research conducted in the physical laboratory using an order of magnitude higher stakes.

Prior work on the dictator game found that subjects became significantly less generous when going from no stakes to low stakes [13], but that going from low stakes to high stakes did not affect donations [13,14]. Consistent with these results, we find that the average donation on MTurk decreases from 44% with no stakes to 33% with $1 stakes.

Prior work on the ultimatum game has found that adding stakes does not affect the average proposal but may increase the variance in proposals [13,15], while the results for responder behavior are more mixed, with one study finding no effect [13] and another finding a significant decrease in rejection rates [15]. It has also been found that increasing from low to high stakes has little effect on either proposals or rejection rates, unless the stakes are

extremely large [13–17]. Our results when comparing no stakes with $1 stakes on MTurk are broadly consistent with these previous findings. In particular, we see no difference in Player 1 proposals, or the minimal amount accepted by Player 2 s when excluding 'inconsistent' players (people who accepted some offer X while also rejecting one or more offers greater than X). However we do find that adding stakes decreases the fraction of such inconsistent Player 2 s, and decreases rejection rates of some Player 1 offers when including inconsistent Player 2 s.

There has been less study of the role of stakes in other social dilemma games. To our knowledge, comparisons between no stakes and low stakes have not been performed in the public goods game or the trust game. Considering the increase of stake size, Kocher, Martinsson and Visser [18] found no significant difference in subjects' contributions in the public goods game when going from low to high stakes, and Johansson-Stenman, Mahmud and Martinsson [19] found that in the trust game, the amount sent by investors decreased when using very high stakes but the fraction returned by trustees was not affected by the changes in stakes. We find no difference in cooperation in the public goods game or trust or trustworthiness in the trust game when comparing no stakes with $1 stakes on MTurk.

Materials and Methods

This research was approved by the committee on the use of human subjects in research of Harvard University, application number F17468-103. Informed consent was obtained from all subjects.

We recruited 756 subjects using MTurk and randomly assigned each subject to play one of four canonical games - the dictator game, ultimatum game, trust game and public goods game - either with or without stakes. In all eight conditions, subjects received a $0.40 show up fee. In the four stakes conditions, subjects had the opportunity to earn up to an additional $1.00 based on their score

in the game (at an exchange rate of 1 point = 1 cent). In the four no-stakes conditions, subjects were informed of the outcome of the game, but the score in the game did not affect subjects' earnings. In all conditions, subjects had to complete a series of comprehension questions about the rules of the game and their compensation, and only subjects that answered all questions correctly were allowed to participate. We now explain the implementation details of each of the four games.

In the Dictator game (DG), Player 1 (the dictator) chose an amount x ($x < = 100$) to transfer to Player 2, resulting in Player 1 receiving a score of $100 - x$ and Player 2 receiving a score of x.

In the Ultimatum Game (UG), Player 1 (the proposer) chose an amount x ($x \in [0,10,20,30,40,50,60,70,80,90,100]$) to offer to Player 2 (the responder). Player 2 could then accept, resulting in Player 1 receiving a score of $100 - x$ and Player 2 receiving a score of x; or reject, resulting in both players receiving a score of 0. We used the strategy method to elicit Player 2 decisions (i.e., Player 2 indicated whether she would accept or reject each possible Player 1 offer). For each Player 2 we then calculated her Minimum Acceptable Offer (MAO) as the smallest offer she was willing to accept. As in the physical lab, some subjects were 'inconsistent' in that they were willing to accept some of the lower offers, but rejected higher offers (that is, they did not have a threshold for acceptance) [20]. When calculating MAOs, we did not include such inconsistent players. We also examined how the addition of stakes changed the fraction of inconsistent players, as well as the rejection rates for each possible Player 1 offer when including all Player 2 s (consistent and inconsistent).

In the Trust Game (TG), Player 1 (the investor) chose an amount x ($x < = 40$) to transfer to Player 2 (the trustee). The transferred amount was multiplied by 3 and given to the trustee, who then chose a fraction y (where $y < = 1$) to return to Player 1. As a result, Player 1 received a score of $(40 - x) + 3yx$ and Player

2 received a score of $3x(1 - y)$. We used the strategy method to elicit Player 2 decisions (i.e., Player 2 indicated the fraction she would return for each possible Player 1 transfer).

In the Public Goods Game (PGG), four players each received an initial endowment of 40 units, and simultaneously choose an amount x ($x_i < = 40$) to contribute to a public pool. The total amount in the pot was multiplied by 2 and then divided equally by all group members. As a result, player i received the score

$$(40 - x_i) + 2 \frac{\sum_{j=1}^{4} x_j}{4}.$$

In the DG, UG and TG experiments, each subject played both roles, first making a decision as Player 1, and then making a decision as Player 2. Subjects were not informed that they would subsequently play as Player 2 when making their Player 1 decisions. Unless otherwise noted, all statistical tests use the Wilcoxon Rank-sum test.

Results and Discussion

As can be seen in Figure 1, introducing stakes altered the distribution of offers in the DG, significantly reducing the average offer (no-stakes = 43.8%, stakes = 33.2%, $p = 0.022$). In the UG, we found a marginally significant positive effect of stakes on Player 1 proposals (no-stakes = 46.1%, stakes = 49.7%, $p = 0.097$). Given the small effect size and borderline significant p-value, we conclude that stakes have little effect on P1 offers in the UG. We also find no significant effect on Player 2 MAOs in the UG (excluding inconsistent players) ($p = 0.285$). However, we do find a significantly higher proportion of inconsistent Player 2's in the no-stakes condition compared to the stakes condition (Chi^2 test, $p = 0.015$). As a result, we also find a significant effect of stakes on Player 2 rejection rates for some Player 1 offers in the UG when including inconsistent players ($p = 0.037$ for the 30% offer, and $p < 0.004$ for
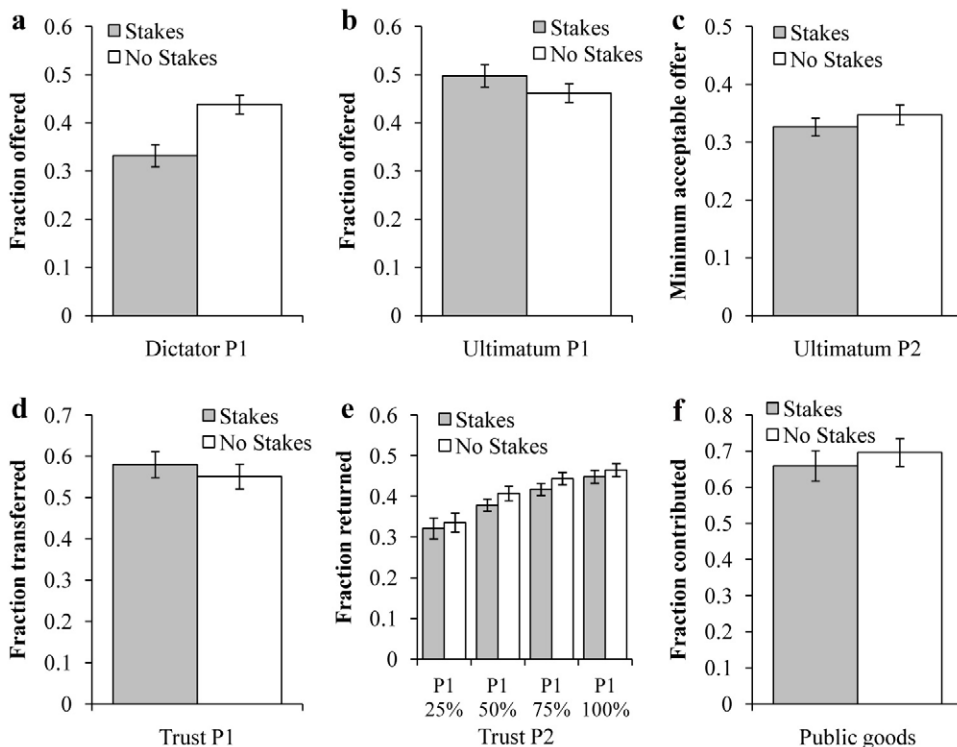
Figure 1. The effect of stakes on average behavior across games.
doi:10.1371/journal.pone.0031461.g001

all offers above 60%). There was no significant effect of stakes on transfers in the TG ($p = 0.386$), back-transfers in the TG ($p > 0.40$ for all possible Player 1 transfers), and contributions in the PGG ($p = 0.656$). We also test whether the variance in behavior differs between the stakes and no-stakes conditions using Levene's F-test. Consistent with our results above, we find that the variance in DG donations is significantly smaller in the stakes condition compared to the no-stakes condition ($p = 0.008$), but that adding stakes did not have an effect on the variance of offers ($p = 0.889$) and MAOs in the UG ($p = 0.678$), transfers ($p = 0.113$) and back-transfers in the TG ($p > 0.600$ for back-transfers on all Player 1 transfers, except for the transfer of 25% where the variance of Player 2 back-transfers in the stakes condition was marginally higher, $p = 0.098$), and contributions in the PGG ($p = 0.517$).

Furthermore, we find that the average behavior on MTurk is in line with behavior observed previously in the physical laboratory with higher stakes. The average donation of 33.2% in our $1 stakes DG is close to the average donation of 28.4% aggregated over more than 616 DG treatments as reviewed in a recent meta-analysis [21].

Since there was little difference in behavior between the stakes and no-stakes conditions in the UG, TG and PGG, we compare the aggregated averages from both conditions in these games to prior work. Considering the UG, it has been shown that using the strategy method significantly affects behavior [22], and to our knowledge no meta-analyses exist which focus on strategy-method UGs. Therefore, we examine behavior in various previous UG experiments that used the strategy method [23–29], and compare the range of outcomes to what we observe in our data. The average Player 1 offer of 48.1% in our experiment is within the range of behavior observed in those studies (35.4%–48.4%), as is our average Player 2 MAO of 33.7% (compared to the range of previous MAOs of 19.2%–36.0%).

Turning now to the TG, we find that the average percentage sent by Player 1 in our experiment (56.6%) is quite close to the

average value of 50.9% reported in a recent trust game meta-analysis aggregating over approximately 80 experiments [30]. The fraction returned by Player 2 of 40.1% in our experiment is also close to the average returned fraction of 36.5% from the same meta-analysis.

For the public goods game, it is important to compare our results to those obtained in previous experiments using the same Marginal Per-Capita Return value (MPCR = 0.5 in our study). In the absence of a meta-analysis that breaks contributions down by MPCR, we compare the average contribution level in our experiment to the range of average contributions observed in various previous studies using the same MPCR [31–36]. The average fraction of the endowment contributed to the public good in our study of 67.7% is within the range observed in these studies (40%–70.4%).

To conclude, we have assessed the effect of $1 stakes compared to no stakes in economic games run in the online labor market Amazon Mechanical Turk. The results are generally consistent with what is observed in the physical laboratory, both in terms of the effect of adding stakes, and the average behavior in the stakes conditions. These experiments help alleviate concerns about the validity of economic game experiments conducted on MTurk and demonstrate the applicability of this framework for conducting large scale scientific studies.

Author Contributions

Conceived and designed the experiments: DR OA. Performed the experiments: OA. Analyzed the data: OA. Contributed reagents/materials/analysis tools: OA DR. Wrote the paper: OA DR YG.

References

1. Dreber A, Gerdes C, Gransmark P (2010) Beauty queens and battling knights: Risk taking and attractiveness in chess.
2. Horton J (2010) Employer expectations, peer effects and productivity: Evidence from a series of field experiments. SSRN eLibrary.
3. Lawson C, Lenz G, Baker A, Myers M (2010) Looking like a winner: Candidate appearance and electoral success in new democracies. World Politics 62: 561–593.
4. Rand D, Nowak M (2011) The evolution of antisocial punishment in optional public goods games. Nature Communications 2: 434.
5. Rand D, Arbesman S, Christakis N (2011) Dynamic social networks promote cooperation in experiments with humans. Proceedings of the National Academy of Sciences 108: 19193–19198.
6. Suri S, Watts D (2011) Cooperation and contagion in web-based, networked public goods experiments. PLoS One 6: e16836.
7. Chandler D, Kapelner A (2010) Breaking monotony with meaning: Motivation in crowdsourcing markets. University of Chicago mimeo.
8. Horton J, Rand D, Zeckhauser R (2010) The online laboratory: Conducting experiments in a real labor market. Technical report, National Bureau of Economic Research.
9. Paolacci G, Chandler J, Ipeirotis P (2010) Running experiments on amazon mechanical turk. Judgment and Decision Making 5: 411–419.
10. Buhrmester M, Kwang T, Gosling S (2011) Amazons mechanical turk: A new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science 6: 3.
11. Mason W, Suri S (2010) Conducting behavioral research on amazons mechanical turk. Behavior Research Methods. pp 1–23.
12. Rand D (2011) The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. Journal of Theoretical Biology.
13. Forsythe R, Horowitz J, Savin N, Sefton M (1994) Fairness in simple bargaining experiments. Games and Economic behavior 6: 347–369.
14. Carpenter J, Verhoogen E, Burks S (2005) The effect of stakes in distribution experiments. Economics Letters 86: 393–398.
15. Cameron L (1999) Raising the stakes in the ultimatum game: Experimental evidence from indonesia. Economic Inquiry 37: 47–59.
16. Hoffman E, McCabe K, Smith V (1996) On expectations and the monetary stakes in ultimatum games. International Journal of Game Theory 25: 289–301.
17. Andersen S, Ertac S, Gneezy U, Hoffman M, List J (2011) Stakes matter in ultimatum games. American Economic Review 101: 3427–39.
18. Kocher M, Martinsson P, Visser M (2008) Does stake size matter for cooperation and punishment? Economics Letters 99: 508–511.
19. Johansson-Stenman O, Mahmud M, Martinsson P (2005) Does stake size matter in trust games? Economics Letters 88: 365–369.
20. Bahry D, Wilson R (2006) Confusion or fairness in the field? rejections in the ultimatum game under the strategy method. Journal of Economic Behavior & Organization 60: 37–54.
21. Engel C (2010) Dictator games: A meta study. Experimental Economics. pp 1–28.
22. Oosterbeek H, Sloof R, Van De Kuilen G (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. Experimental Economics 7: 171–188.
23. Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization 3: 367–388.
24. Kahneman D, Knetsch J, Thaler R (1986) Fairness and the assumptions of economics. Journal of business. pp 285–300.
25. Larrick R, Blount S (1997) The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. Journal of Personality and Social Psychology 72: 810.
26. Solnick S (2001) Gender differences in the ultimatum game. Economic Inquiry 39: 189–200.
27. Straub P, Murnighan J (1995) An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. Journal of Economic Behavior & Organization 27: 345–364.
28. Wallace B, Cesarini D, Lichtenstein P, Johannesson M (2007) Heritability of ultimatum game responder behavior. Proceedings of the National Academy of Sciences 104: 15631.
29. Zak P, Stanton A, Ahmadi S (2007) Oxytocin increases generosity in humans. PLoS One 2: e1128.

30. Johnson N, Mislin A (2008) Cultures of kindness: A meta-analysis of trust game experiments. Available at Social Science Research Network: http://ssrn com/abstract 1315325.

31. Andreoni J (1988) Why free ride?: Strategies and learning in public goods experiments. Journal of Public Economics 37: 291–304.

32. Andreoni J (1995) Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. The Quarterly Journal of Economics 110: 1.

33. Andreoni J (1995) Cooperation in public-goods experiments: kindness or confusion? The American Economic Review. pp 891–904.

34. Brandts J, Schram A, Institute T (1996) Cooperative gains or noise in public goods experiments. Tinbergen Institute.

35. Keser C, van Winden F (1997) Partners contribute more to public goods than strangers.

36. Weimann J (1994) Individual behaviour in a free riding experiment. Journal of Public Economics 54: 185–200.