Supplementary Text and Figures: A laboratory study of open and closed peer review reveals differences in accuracy and social dynamics

Authors: Jeffrey T. Leek, Margaret A. Taub, and Fernando J. Pineda

Theoretical Model

Definition of the Peer Review Game

In our model there are K players participating in a game for a total of T units of time. Each player in the game participates in two activities: (1) solving problems and (2) reviewing solutions of their peers. For player k, the total time spent reviewing T_k^r and solving T_k^s must be less than the total time allocated for playing the game $T_k^r + T_k^s \leq T$. Over the course of the game, player k submits N_k^s solutions and reviews N_k^r solutions for other players. Let s_{ikj} indicate the ith solution for player k, which is reviewed by player j. For each solution there is a corresponding time the solution was submitted t_{ikj}^s and time that the reviewer completed the review t_{ikj}^r. For player k the number of accepted papers at time t is the sum of the indicators that each of their submitted solutions is accepted up to that point: $A_k(t) = \sum_{\{i:t_{ijk}^r < t\}} 1(s_{ikj} \text{ accepted})$.

The payoff is proportional to the number of accepted solutions, which reflects the commonly held belief of "publish or perish" in academia. So the expected payoff for player k at time t is:

$$\begin{aligned} \mathrm{E}[A_k(t)] &= \mathrm{E}\left[\sum_{\{i:t_{ijk}^r < t\}} 1(s_{ikj} \text{ accepted})\right] \\ &= \sum_{\{i:t_{ijk}^r < t\}} p_{ikj} \end{aligned}$$

where p_{ikj} is the probability that solution i for player k is accepted by player j. The payoff is a function of the number of submitted solutions and the probability that each solution is accepted. The probability a solution is accepted is a function of the submitter, the reviewer, the time the solution is reviewed, and the solution itself.

$$p_{ikj} = f(s_{ikj}, t_{ikj}^r, j, k)$$

where $f(\cdot)$ is a non-negative function mapping the solution, the review time, the solver, and the reviewer

onto $[0, 1]$. Player k can increase their payoff by increasing the number of solutions they submit or increasing the probability each solution is accepted.

An alternative is a competitive payoff where the payoff function is proportional to the difference between a player's number of accepted solutions and the maximum of all the other player's payoffs. In this case, the expected payoff is:

$$\begin{aligned} \mathrm{E}[A_k(t)] - \mathrm{E}[\max_k\{A_k(t)\}] &= \mathrm{E}\left[\sum_{\{i:t^r_{ijk}<t\}} 1(s_{ikj} \text{ accepted})\right] \\ &= \sum_{\{i:t^r_{ijk}<t\}} p_{ikj} - \mathrm{E}[\max_k\{A_k(t)\}] \end{aligned}$$

Closed Peer Review (CPR)

Under closed peer review, the model for the acceptance probability for solution s_{ikj} is modeled as:

$$f^{-1}(p_{ikj}) = \alpha(s_{ikj}) + \beta(k) + \gamma(j) + \kappa(A(t^r_{ikj}))$$

Here there is an effect for the solution itself $\alpha(\cdot)$ which may reflect a large number of factors about the solution, including the type of problem or the time spent on the solution. There is also an effect for the solver $\beta(\cdot)$ since some solvers are more likely to submit correct solutions than others. Each reviewer may choose to accept or reject problems at a different rate which we model by $\gamma(\cdot)$. Under CPR the public information is the number of solutions that each player has submitted and had accepted by another player. $A(t^r_{ikj}))$ is a vector of the cumulative number of accepted solutions for each player at time t^r_{ijk}. The function $\kappa(\cdot)$ quantifies the influence of this information on the probability solution s_{ikj} is accepted.

At any given time point a player can choose between three different strategies: (1) solve and submit a problem, (2) review a problem and reject, or (3) review a problem and accept. The first strategy has the potential to improve a player's payoff, by increasing the number of submitted solutions. If a player chooses either of the first two strategies, no other player's score will increase. If the player chooses strategy (3), then another player's score will increase. However, that person will not know who accepted their solution. Under CPR, if a player chooses strategy (2) or (3) they will reduce the amount of time they spend solving a problem and will reduce their expected payoff. However, no other player will be

aware of this choice since reviews are anonymous and only the cumulative accepted solutions for each player is known. In this game, there is no increase to the payoff function for reviewing. Therefore, each player maximizes their expected payoff by always choosing strategy (1) and never reviewing, so this solution is the Nash equilibrium [1].

Open Peer Review (OPR)

Under OPR the model for the acceptance probability for solution s_{ikj} includes the same terms as CPR, along with terms that encode the influence of the current public and private information available to each player.

$$f^{-1}(p_{ikj}) = \alpha(s_{ikj}) + \beta(k) + \gamma(j) + \kappa(A(t^r_{ikj})) + \eta(R^a(t^r_{ikj})) + \xi(R_{kj}(t^r_{ikj}), R^a_{kj}(t^r_{ikj}))$$

The model includes a term, $\eta()$, that is a function of vector of the cumulative number of solutions reviewed and accepted by each player. The functions $\kappa(\cdot)$ and $\eta(\cdot)$ encode the public information available to each player. Under the open system, player k also knows the cumulative number of times player j has reviewed their solutions, $R_{kj}(t^r_{ikj})$, and accepted their solutions $R^a_{kj}(t^r_{ikj})$ at the time of the review. The function ξ quantifies the effect of this information on the probability of acceptance.

Under the OPR it is possible that a player may incur some benefit by reviewing for other players. Specifically if a player has previously accepted solutions for player j, they may improve the probability their solution is accepted through the function $\xi(\cdot)$. Similarly, if they are a generous reviewer to all the other players, player j may again be more sympathetic and the probability of acceptance may be increased through the function $\eta(\cdot)$. The residual benefit of reviewing may carry over to future times, so the functions η and ξ are functions of the cumulative reviews and acceptances to time point t^r_{ijk}.

Under OPR, a player still has the same three strategy choices at any given time point: (1) solve and submit a problem, (2) review a problem and reject, or (3) review a problem and accept. However, under OPR a player may incur some increase in their probability of acceptance if they choose strategy (2) or (3). They are particularly likely to incur increases in their acceptance probability when choosing strategy (3). Under this mode, additional Nash equilibria may be possible. To calculate these equilibria, substantial additional assumptions are required about the benefit of reviewing, the time it costs to perform

a review, and the timing of additional reviews. Since the payoffs functions now depend continuously on the number of accepted and reviewed at each time point, the game must be modeled as a continuous game. Theoretical analysis of OPR represents a potentially fruitful area for future research.

Relative Payoff of Reviewing and Solving

It is not difficult to argue that in science, the payoff for solving problems is significantly greater than the payoff of reviewing submissions. The only way to change this ordering is to decrease the payoff for solving problems or to increase the payoff for reviewing problems, or both. The former might be achieved in situations where the information available to the community causes the community to punish a player by reducing the acceptance rate of the player's submissions. The latter might be achieved by increasing the time spent reviewing and rejecting the submissions of other players. For example if a player could somehow reject all the submissions of a strong competitor, without knowledge of these actions being provided to the community.

Experimental Setup - The Peer Review Game

The Peer Review Game

We developed a peer review game that can be played by two or more players. The game was developed as an Amazon Machine Image (AMI) that can be launched from the Amazon Elastic Compute Cloud (`http://aws.amazon.com/`). The game was developed using the vWorker online development platform (`http://www.vworker.com/`). Players were directed to a website of a temporary web-server and logged on with a user name and password. When the investigator initiated the game, the players were shown a task selection page (Supplementary Figure 2). They could choose to solve a problem or choose to review a problem from their list of pending reviews. If a player chose to solve a problem, then a GRE-like problem was selected from a database for them to solve and displayed to their screen (Supplementary Figure 3). The GRE problems used for the experiment were based on problems from the website `http://majortests.com/`. If they chose to review a problem, then they were shown a solution to a problem submitted by one of their peers (Supplementary Figure 4). They could choose to either accept or reject the solution to the problem. The program acted as editor, randomly assigning

problems to players for peer review.

In both the open and closed games reviewers were shown the identity of the player who solved the problem. Under the open system, solvers were also shown the identity of the player who acted as peer reviewer for their solution. During the course of the game, information was projected onto a screen at the front of the room. In the closed mode, the number of solutions each player had submitted and had accepted was displayed (Supplementary Figure 5a). In the open mode, the number of solutions each player had reviewed and accepted for one of their peers was also displayed (Supplementary Figure 5b).

At the beginning of each game, the players were read the instructions for the appropriate mode (closed or open) as described in the following sections. The investigator then initiated a session of the Peer Review Game that lasted for T=40 minutes in each case. Nametags were given to each subject with their anonymous subject ID at the beginning of the experiment and players were permitted to speak to one another during the course of the experiment.

Recruitment

Six laboratories at the Johns Hopkins Medical School and Johns Hopkins Bloomberg School of Public Health were recruited to participate in the peer review experiment. Laboratories consisted of graduate students, postdoctoral fellows, research scientists, and principal investigators. Each laboratory participated in one replication of the Peer Review Game; the goal was to mimic the small and relatively tight-knit communities of scientists who act as peer reviewers for each other's papers. Experiments were performed on laboratories of laboratories of size K=8, 8, and 9 players for the closed game and K=7,10, and 8 players for the open game. Participating laboratories were offered $50 for each 10 participating members of the lab, a complimentary lunch, and the potential for two lab members to earn $5 each. Written informed consent was obtained from all participants in the study. Recruitment was performed with approval from the Johns Hopkins Bloomberg School of Public Health IRB, project number 3316.

Instructions for the Closed Peer Review Games

Purpose of research project This research is being done to evaluate open and closed peer review systems experimentally. Peer review is the process by which scientific research is evaluated for publica-

tion in journals. The goal of this study is to determine whether anonymous (closed) or non-anonymous (open) peer review results in more correct research being accepted.

Why you are being asked to participate You are being asked to participate in the study because you are a graduate student, postdoctoral research fellow, scientist, or faculty member at Johns Hopkins University and are representative of the population of individuals who will participate in the peer review process.

Procedures Once the experiment begins, you will be asked to answer multiple choice questions similar to questions on the graduate record exam (GRE). After you submit your answer, the solution will be randomly assigned to another participant in the study for review. The reviewer can either choose to accept or reject the solution. The reviewer will know your subject ID. However subjects who submit solutions will not know the ID of the reviewer of their solution. Throughout the course of the experiment you will act as both a reviewer and a problem solver. You may spend as much time as you like on either task. The experiment will last for forty minutes. I will now show you example screens from the experiment website and you may ask questions about the study procedure.

Risks/discomforts You may experience some stress since you will be asked to answer GRE like problems and review the solutions of your peers. However, the only interaction you will have with other participants will be through the anonymous subject IDs.

Payment The two individuals with the most accepted answers at the conclusion of the experiment will receive $5. The payment will be in cash immediately following the experiment. If you leave the study early you will lose your opportunity to win the cash prizes distributed at the end of the experiment.

Protecting data confidentiality All research projects carry some risk that information about you may become known to people outside of a study. We minimize these risks by not connecting your responses to any information that could be used to identify you. All data collected during this experiment will only be connected with the anonymous subject ID you have been assigned.

Protecting subject privacy during data collection Your responses and reviews will not be personally associated with you. All interaction will be performed based on the anonymous subject IDs you have been assigned.

What happens if you leave the study early? You may leave the study at any time without penalty.

Instructions for the Open Peer Review Games

Purpose of research project This research is being done to test open and closed peer review systems experimentally. Peer review is the process by which scientific research is evaluated for publication in journals. The goal of this study is to determine whether anonymous (closed) or non-anonymous (open) peer review results in more correct research being accepted.

Why you are being asked to participate You are being asked to participate in the study because you are a graduate student, postdoctoral research fellow, scientist, or faculty member at Johns Hopkins University and are representative of the population of individuals who will participate in the peer review process.

Procedures Once the experiment begins, you will be asked to answer multiple choice questions similar to questions on the graduate record exam (GRE). After you submit your answer, the solution will be randomly assigned to another participant in the study for review. The reviewer can either choose to accept or reject the solution. The reviewer will know your subject ID and you will know the reviewer ID for each solution after it is reviewed. Throughout the course of the experiment you will act as both a reviewer and a problem solver. You may spend as much time as you like on either task. The experiment will last for one forty minutes. I will now show you example screens from the experiment website and you may ask questions about the study procedure.

Risks/discomforts You may experience some stress since you will be asked to answer GRE like problems and review the solutions of your peers. However, the only interaction you will have with other participants will be through the anonymous subject IDs.

Payment The two individuals with the most accepted answers at the conclusion of the experiment will receive $5. The payment will be in cash immediately following the experiment. If you leave the study early you will lose your opportunity to win the cash prizes distributed at the end of the experiment.

Protecting data confidentiality All research projects carry some risk that information about you may become known to people outside of a study. We minimize these risks by not connecting your responses to any information that could be used to identify you. All data collected during this experiment will only be connected with the anonymous subject ID you have been assigned.

Protecting subject privacy during data collection Your responses and reviews will not be personally associated with you. All interaction will be performed based on the anonymous subject IDs you

have been assigned.

What happens if you leave the study early? You may leave the study at any time without penalty.

Statistical methods

Group dynamics measurement

Next we estimated a measure of cooperation or obstruction between subjects i and j. The baseline observed acceptance probability for subject i, P_i was calculated as A_i/N_i where A_i is the number of solutions accepted by subject i and N_i is the number of solutions reviewed by subject i. We computed the observed probability that subject i accepts a solution submitted by subject j, P_{ij}, as A_{ij}/N_{ij}, where A_{ij} is the number of solutions accepted by subject i which were submitted by subject j, and N_{ij} is the number of solutions reviewed by subject i which were submitted by subject j. The difference $d_{ij} = P_{ij} - P_i$ gives a measure of the change in the probability subject i accepts a solution from subject j relative to their overall acceptance rate. Similarly, we can calculate d_{ji} as a symmetric measurement. If d_{ij} and d_{ji} are both positive, then the interaction between the two subjects is cooperative. Similarly, if both values are negative, the interaction between the two subjects is obstructive. We calculated the total number of possible interactions under both the open and closed peer review experiments. Among these, we identified the number that were cooperative. We then performed a two-sample test of proportions to evaluate whether there was more cooperation under OPR or CPR.

Outcome modeling

In all outcome modeling, the unit of observation is one reviewed problem. Each reviewed problem has a solver and a reviewer and is associated with a particular study type, either open or closed.

To control for differences in behavior between individual participants, the models described below were fit using a mixed-model framework, with all models including separate random effects for solvers and reviewers. Model fitting was done in the statistical programming language R (`http://www.r-project.org`) using the function `glmer` from the package `lme4` with a linear link assuming Gaussian distribution of random effects [2]. In a general form the random effects model can be written

as

$$y_{ijt} = \mu + \sum_{k=1}^{p} \beta_k x_{ijtk} + u_i + v_j + \epsilon_{ijt}$$

where y_{ijt} is the outcome of interest related to a review at time t by subject j, for a solution submitted by subject i; μ is the mean outcome over the whole data set; x_{ijtk} is the k^{th} covariate of interest which has effect size β_k, u_i is a random effect associated with subject i and v_j is a random effect associated with subject j. We assume that u_i, v_j and ϵ_{ijt} are mean zero Normal random variables with variances σ_u^2, σ_v^2 and σ_ϵ^2, respectively.

To assess the impact of previous review performance by a subject on the chance that solutions submitted by that subject will be accepted, we associated to each reviewed problem the number of solutions accepted by the problem submitter, up to the time the problem was reviewed. In the open framework, this value was known to all study participants, including the reviewer; in the closed framework, this value was unknown.

Modeling the acceptance probability of a submission as a function of this covariate and the study type, and their interaction, we assessed the change in acceptance probability for each solution accepted by the submitter, in either the open or closed review setting. Define a_{ijt} to be the indicator that solution s_{ijt} is accepted. The model is then:

$$a_{ijt} = \mu + \beta_1 R_{it}^a + \beta_2 S_{ij} + \beta_3 R_{it}^a S_{ij} + u_i + v_j + \epsilon_{ijt}$$

where R_{it}^a is the number of reviewed and accepted solutions by subject i by time t, S_{ij} is an indicator of the study type that subjects i and j participated in (taking a value of 0 for closed review and 1 for open review). In this model u_i is a random effect representing the solver, v_j is a random effect representing the reviewer and ϵ_{ijt} represents residual variation not due to reviewer or solver effects.

To assess the impact of the open or closed scenarios on review quality, we associated to each reviewed problem an indicator of whether the review was accurate, given that we know the correctness of the submitted solution.

We defined the variable c_{ijt} to be an indicator of whether solution s_{ijt} was correctly reviewed (e.g. accepted if correct, rejected if incorrect). To assess the impact of cooperation on review accuracy, for each reviewed problem, we defined a 0-1 indicator O_{ijt} which takes a values of 1 if subjects i and j

have a cooperative interaction. We then fit the model

$$c_{ijt} = \mu + \beta_1 O_{ijt} + u_i + v_j + \epsilon_{ijt}$$

where all terms are as defined above.

To ensure the effect observed in this model is not due only to the increased accuracy of the solution submitted by the problem solver, for each reviewed problem we defined a three-level factor, with level 0 indicating that neither the solver nor the reviewer was part of a cooperative pair, 1 indicating that only the solver was part of a cooperative pair, and 2 indicating that both the solver and the reviewer are part of a cooperative pair. Calling this variable Q_{ijt} we then fit the model

$$c_{ijt} = \mu + \beta_1 Q_{ijt} + u_i + v_j + \epsilon_{ijt}$$

where all terms are as defined above.

We also modeled this accuracy as a function of study type alone to determine whether one scenario produced more accurate reviews. We fit the model

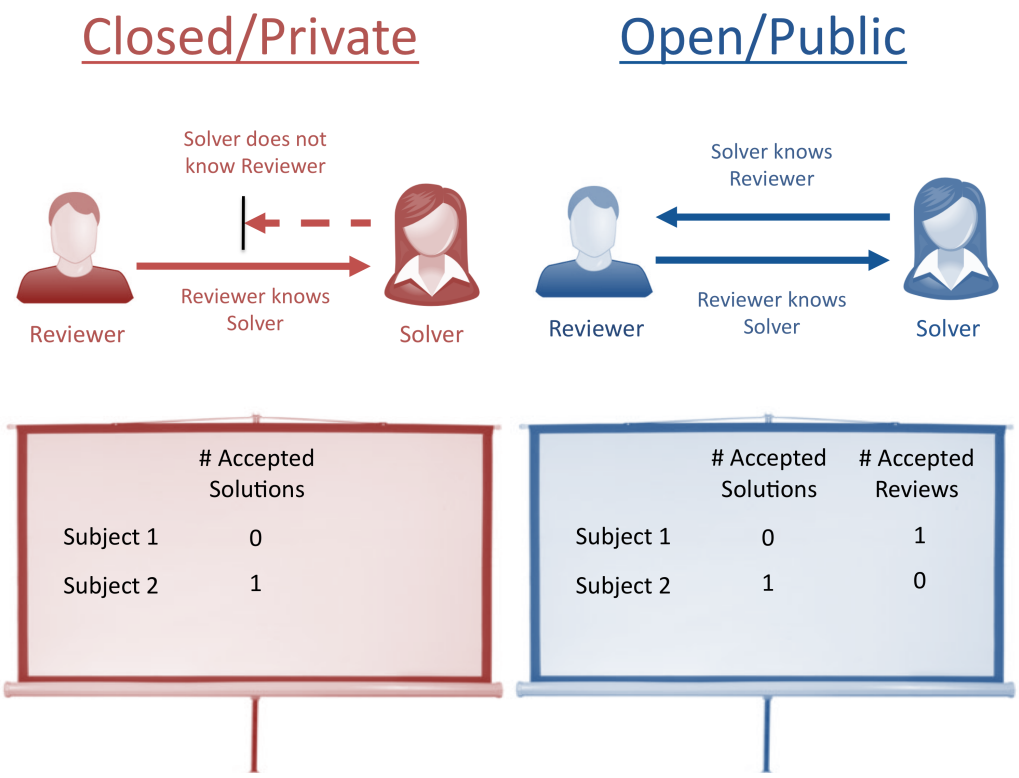
$$c_{ijt} = \mu + \beta_1 S_{ij} + u_i + v_j + \epsilon_{ijt}$$

where all terms are as defined above.

Reproducible Research

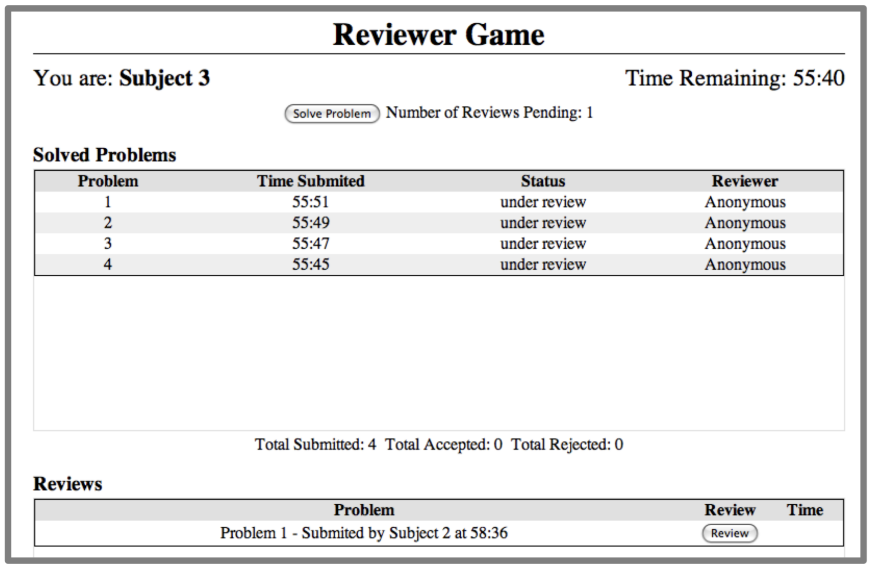
To conform with the standards of reproducible research, we have included R (`http://www.r-project.org`) scripts and R data object that can be used to reproduce all of the results and figures in our manuscript.

Supplementary Figures

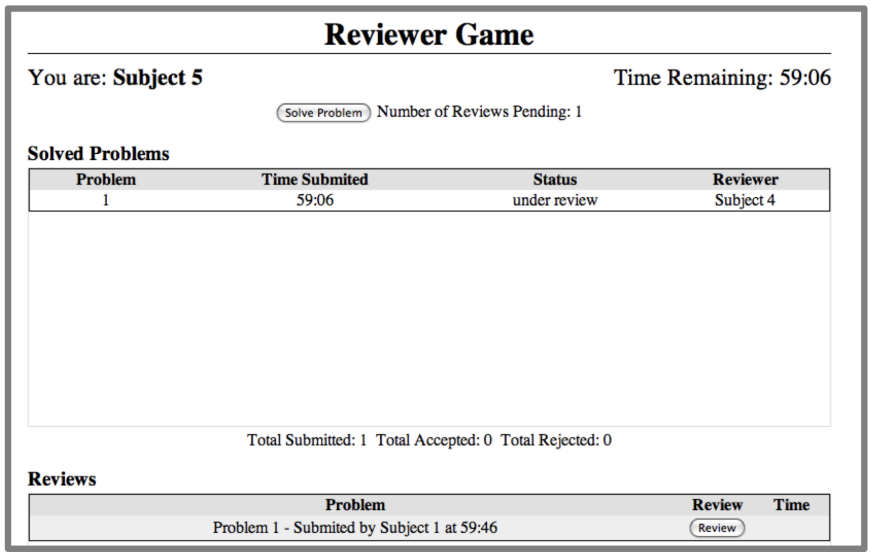


Supplementary Figure 1. Open versus closed peer review systems for the peer review game Under the closed system of peer review (left column), reviewers know the identify of problem solvers, but problem solvers do not know the identity of the reviewers. Public information is limited to the number of accepted solutions for each player. Under the open system of peer review (right column) solvers and reviewers are known to each other, and both the number of accepted solutions and accepted reviews for each player are known publicly.

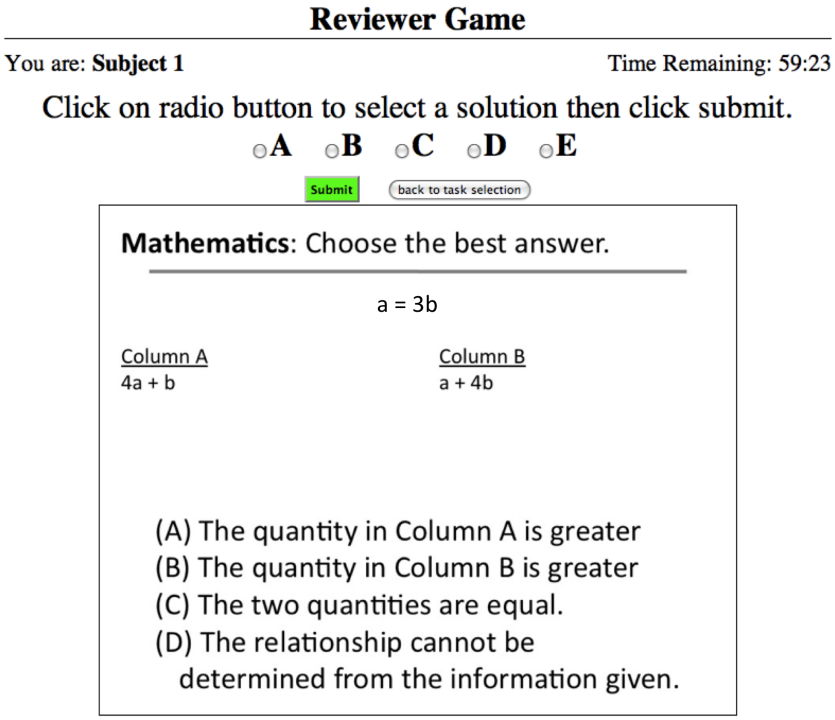
a



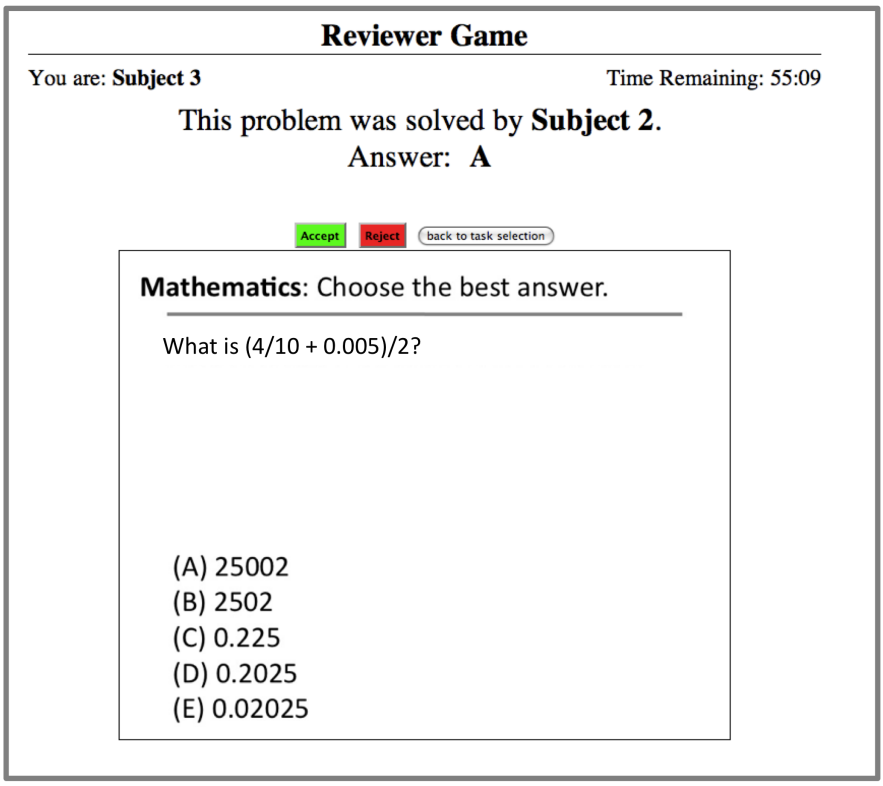
b



Supplementary Figure 2. The task selection screens for the Peer Review Game under the **(a)** closed and **(b)** open modes. In each case a player may elect to solve or review a problem. In the open peer review mode, players know the identity of the players reviewing their solutions.

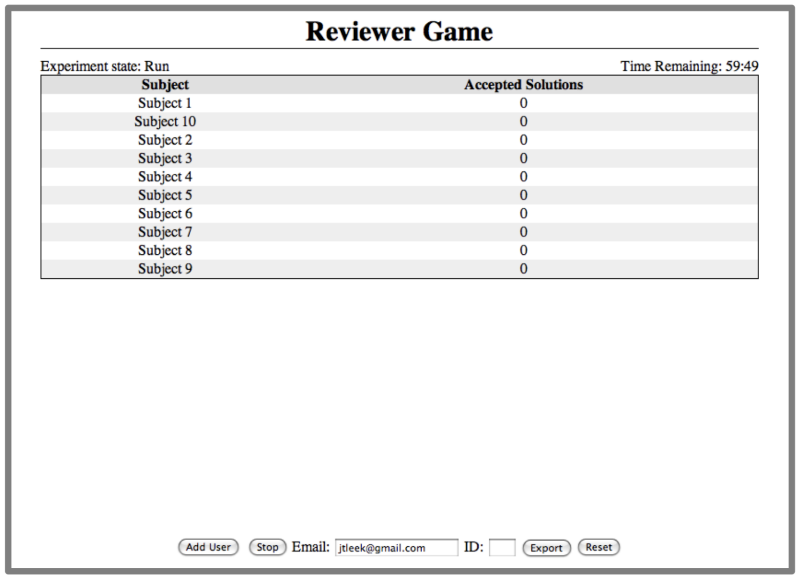


Supplementary Figure 3. The solving screen for both the open and closed modules.

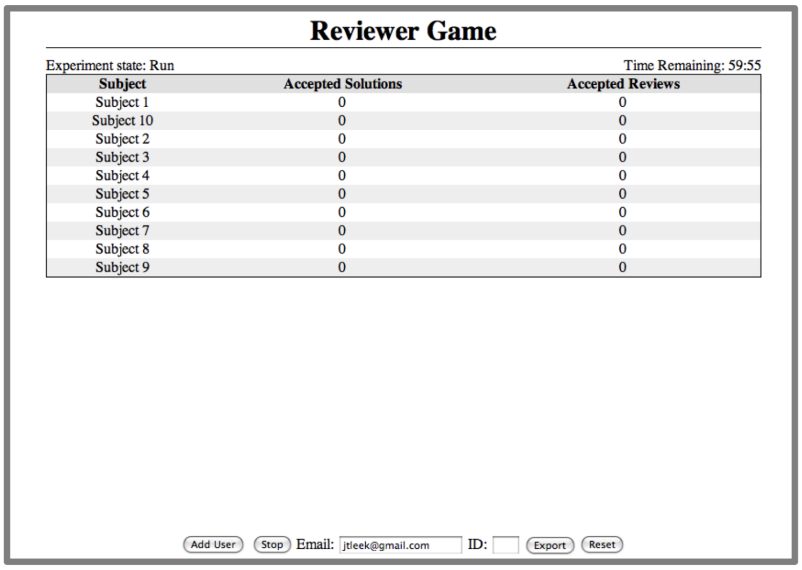


Supplementary Figure 4. The reviewing screen for both the open and closed modules.

a



b



Supplementary Figure 5. The information projected for all players to see in the Peer Review Game under the **(a)** closed and **(b)** open modes. In each case the number of solutions each player has had accepted are displayed. In the open review system, the number of solutions reviewed and accepted by each player is also displayed.

References

[1] R. B. Myerson, *Game theory: analysis of conflict* (Harvard University Press, 1997).

[2] P. Diggle, P. Heagerty, K.-Y. Liang, S. L. Zeger, *Analysis of longitudinal data* (Oxford University Press, Oxford, 2002).