NBER WORKING PAPER SERIES

WHAT POLICIES INCREASE PROSOCIAL BEHAVIOR? AN EXPERIMENT WITH REFEREES AT THE JOURNAL OF PUBLIC ECONOMICS

Raj Chetty
Emmanuel Saez
László Sándor

Working Paper 20290
http://www.nber.org/papers/w20290

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2014

This paper is in preparation for publication at the Journal of Economic Perspectives. We thank Nava Ashraf, David Autor, Stefano DellaVigna, Hilary Hoynes, Damon Jones, Emir Kamenica, Lawrence Katz, Henrik Kleven, Ulrike Malmendier, Monica Singhal, Timothy Taylor, and numerous seminar participants for helpful discussions and comments. We thank Jenny Henzen and the Elsevier staff for supporting the project and providing data and Liz Anderson for implementing the experiment. Greg Bruich, Jessica Laird, Keli Liu, Alex Olssen, and Heather Sarsons provided outstanding research assistance. Financial support from the Lab for Economic Applications and Policy at Harvard and the National Science Foundation is gratefully acknowledged. A de-identified version of the referee-level dataset used in this paper can be downloaded at http://obs.rc.fas.harvard.edu/chetty/jpube_experiment.zip The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Policies Increase Prosocial Behavior? An Experiment with Referees at the Journal of Public Economics

Raj Chetty, Emmanuel Saez, and László Sándor

NBER Working Paper No. 20290
July 2014
JEL No. H23,H41

ABSTRACT

We evaluate policies to increase prosocial behavior using a field experiment with 1,500 referees at the Journal of Public Economics. We randomly assign referees to four groups: a control group with a six week deadline to submit a referee report, a group with a four week deadline, a cash incentive group rewarded with $100 for meeting the four week deadline, and a social incentive group in which referees were told that their turnaround times would be publicly posted. We obtain four sets of results. First, shorter deadlines reduce the time referees take to submit reports substantially. Second, cash incentives significantly improve speed, especially in the week before the deadline. Cash payments do not crowd out intrinsic motivation: after the cash treatment ends, referees who received cash incentives are no slower than those in the four-week deadline group. Third, social incentives have smaller but significant effects on review times and are especially effective among tenured professors, who are less sensitive to deadlines and cash incentives. Fourth, all the treatments have little or no effect on agreement rates, quality of reports, or review times at other journals. We conclude that small changes in journals' policies could substantially expedite peer review at little cost. More generally, price incentives, nudges, and social pressure are effective and complementary methods of increasing prosocial behavior.

Raj Chetty
Department of Economics
Harvard University
1805 Cambridge St.
Cambridge, MA 02138
and NBER
chetty@fas.harvard.edu

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

László Sándor
Department of Economics
Harvard University
1805 Cambridge St.
Cambridge, MA 02138
sandorl@gmail.com

The peer review process familiar to all academic researchers offers a classic example of the positive externalities from prosocial behavior: the reviewer bears the costs from submitting a high-quality referee report quickly, while the gains to the authors of the paper and to society from the knowledge produced are potentially large. We evaluate the impacts of economic and social incentives on peer review using an experiment with 1,500 referees at the *Journal of Public Economics*. The specific aim of the experiment is to understand how to improve the speed and quality of peer review, an issue of particular importance to the economics profession given the slowdown of the publishing process (Ellison 2002). Our broader objective is to evaluate commonly used methods of increasing prosocial behavior and to test the predictions of competing theories.

In our experiment, we randomly assign referees to four groups: a control group with a six-week (45 day) deadline to submit a referee report, a group with a four week (28 day) deadline, a cash incentive group rewarded with $100 for meeting a four week deadline, and a social incentive group in which referees were told that their turnaround times would be publicly posted. The experiment yields four sets of results.

First, shortening the deadline from 6 weeks to 4 weeks reduces median review times from 48 days to 36 days. Because missing the deadline has no direct consequence, we believe the shorter deadline acts primarily as a "nudge" (Thaler and Sunstein 2008) that changes the default date at which referees submit reports. Second, providing a $100 cash incentive for submitting a report within four weeks reduces median review times by an additional eight days. Third, the social incentive treatment reduces median review times by approximately 2.5 days – which is intriguing given that the degree of social pressure applied here is relatively light. We also find that that social incentives have much larger effects on tenured professors, but in contrast, tenured professors are less sensitive to deadlines and cash incentives than untenured referees.

Finally, we evaluate whether the treatments have an impact on other outcomes besides review time.[1] Economic models of multi-tasking (e.g., Holmstrom and Milgrom 1991) predict that referees will prioritize the incentivized task (i.e., submitting a report quickly) at the expense of other aspects of performance (e.g., the quality of reviews). We find that the shorter deadline has no effect on the quality of the reports that referees submit, as measured by whether the editor follows their recommendation or the length of referee reports. The cash and social incentives induce referees to

[1] The cash incentive increases the fraction of referees who agree to review a manuscript. The social incentive reduces agreement rates, while the shorter deadline has no impact. We show that the selection effects induced by these changes in agreement rates are modest and are unlikely to explain the observed changes in review times.

write slightly shorter referee reports, but do not affect the probability that the editor follows the referee's advice. We also find little evidence of negative spillovers across journals: the treatments have no detectable effects on referees' willingness to review manuscripts and review times at other Elsevier journals.

We conclude that small changes in journals' policies could substantially improve the peer review process at little cost. Shorter deadlines appear to be an essentially costless means of expediting reviews. Cash and social incentives are also effective, but have monetary and psychic costs that must be weighed against their benefits.

A large body of evidence from the lab has considered the determinants of prosocial behavior and altruism (for example, Ledyard 1995; Fehr and Fischbacher 2003; Vesterlund 2014). Our study provides evidence from the field, which has been considerably more limited. Prior work concerning prosocial behavior has often debated whether extrinsic incentives such as cash payments are effective in increasing prosocial behavior because they may crowd out intrinsic motivation (Titmuss 1971; Bénabou and Tirole 2003). In our application, if referees submit reviews to be recognized for their service to the profession by editors, the provision of monetary incentives could potentially erode this signal and have a negative impact on review times. However, our analysis shows that, at least in this context, price incentives, nudges, and social pressure are all effective and complementary methods of increasing prosocial behavior.

Experimental Design

We conducted the experiment over a 20 month period, from February 15, 2010 to October 26, 2011. All referees for the *Journal of Public Economics* during this period were randomly assigned to one of four groups. For simplicity, only referee requests for new submissions were included in the experiment. These assignments were permanent for the duration of the experiment: referees never switched groups. The co-editors in charge of handling each new submission chose referees to review the paper without seeing the group to which the referee was assigned.

Some key features of the four groups are shown in Table 1.[2] All deadlines for the differing groups were defined relative to the date at which the invitation was sent – not the date at which the referee accepted the invitation – to eliminate incentives to delay agreement.

The *control* or what we will refer to as the *six-week* group actually faced a 45 day deadline for submitting a referee report, the deadline that was in place at the journal before the experiment began. The deadline was described using the following language in the invitation letter: "If you accept this invitation, I would be very grateful if you would return your review on or before **July 21, 2010** (6 weeks from now)."

The *four-week* group faced a 28 day deadline for submitting a report. The email they received was identical to that sent to referees in the control group, except for the due date.

The *cash incentive* group faced a 28 day deadline and received a $100 Amazon gift card for submitting a report before the deadline. In addition to the standard text describing the deadline, the invitation letters in the cash incentive group included the following text: "As a token of appreciation for timely reviews, you will receive a **$100** Amazon.com® Gift Card if you submit your report on or before the due date. The Journal of Public Economics will automatically email you a gift card code within a day after we get your report (no paperwork required)."

Finally, the *social incentive* group faced a six-week (45 day) deadline and was told that referee times would be publicly posted by name at the end of the calendar year. In addition to the standard text describing the deadline, the invitation letters in the social incentive group included the following text: "In the interest of improving transparency and efficiency in the review process, Elsevier will **publish referee times by referee name**, as currently done by the Journal of Financial Economics at this website. The referee times for reports received in 2010 will be posted on the Journal of Public Economics website in January 2011. Note that referee anonymity will be preserved as authors only know the total time from submission to decision (and not individual referee's times)."

One week prior to their deadlines, referees who had not yet submitted reports received emails reminding them that their reports were due in a week. For the social and cash incentive groups, these

[2] An on-line appendix available with this paper includes the details of the experiment. Appendix Figure 1 presents a flow chart for the entire experiment. Appendix A shows our invitation emails. Appendix B shows our reminder and thank-you emails. Appendix C includes more detail on data sources and variable definitions. Appendix Table 1 presents summary statistics for the primary experimental period (referee invitations between February 15, 2010 and May 9, 2011). Appendix D describes the reweighting methodology behind Figure 2b. Appendix E presents the hazard model estimates of treatment effects on review times. Appendix F provides a list of other journals used to assess spillover effects. Appendix G presents a summary of all the appendix tables and figures. A de-identified version of the 3,397 observation dataset is available at http://obs.rc.fas.harvard.edu/chetty/jpube_experiment.zip.

emails included language reminding referees of the treatments they faced. We also sent overdue reminders 5 days, 19 days, and 33 days after the due date. Referees in the cash, four-week, and six-week groups were simply informed their reports were past due. Referees in the social incentive group were again reminded that their referee times would be publicly posted. After the referees submitted reports, they received a thank you email. Referees in the cash incentive group received an Amazon gift card code in this thank you email if they submitted before the 28 day deadline. Those in the social incentive group received information on the number of days it took for them to submit the report.

To study the impact of monetary payments on intrinsic motivation after cash incentives are withdrawn, we stopped cash payments on May 9, 2011, roughly six months before we ended the other treatments. Referees in the cash incentive group continued to face a four-week deadline after this point, and received the same invitation and reminder emails as those in the four-week group. All other treatments continued until the end of the experiment on October 26, 2011, at which point all referees were reverted back to the six-week (45 day) deadline.

We analyze the effects of the experiment using information from two sources. We obtain information on referee assignments, review times, and other related outcomes at the *Journal of Public Economics*, as well as other Elsevier journals from Elsevier's editorial database. We obtain information on referee characteristics – an indicator for holding an academic position, tenure status, gender, and an indicator for working in the United States – from curricula vitae posted online.

Each observation in our analysis dataset corresponds to a single referee invitation sent between February 15, 2010 and October 26, 2011. During this period, 3,397 invitations were sent out to 2,061 distinct referees. We include all observations in the referee report level dataset in our analysis, so that referees who are invited multiple times contribute multiple observations.

In our baseline analysis, we restrict attention to referee invitations sent between February 15, 2010 and May 9, 2011, the period when the cash reward was offered. We term this period the *primary* experimental period. During this period we sent 2,423 invitations, of which 66.2 percent were accepted. Among these referees, 93.7 percent submitted a report before the editor made a decision. The median turnaround time for those who submitted reports was 41.0 days. Among the 1,157 referees who agreed to review a manuscript during the primary experimental period, 74.9 percent of referees agreed to review one manuscript during the experiment, 16.4 percent agreed to review two manuscripts, and the rest agreed to review three or more manuscripts.

To verify the validity of our experimental design, we calculated these summary statistics by treatment group for referee assignments from November 1, 2005 to February 15, 2010, *before* the

experiment began. As expected, given randomization, we find no statistically significant differences across the control group or the three treatment groups in these pre-determined characteristics (details in Appendix Table 2a). Hence, differences in performance across the four groups during the experimental period can be interpreted as causal effects of the treatments.

Four Sets of Outcomes

We analyze four sets of outcomes: 1) agreement to submit a review, 2) time taken to submit the review, 3) report quality, and 4) performance at other journals.

Outcome 1: Acceptance of Referee Invitation

Table 2 shows the percentage of referee invitations accepted by treatment group. We structure this and all subsequent tables as follows. The four columns correspond to the four experimental groups: six-week, social, four-week, and cash. For each group, we report the point estimate and associated standard error in parentheses. We cluster standard errors by referee to account for the fact that some referees review multiple papers. We also report p-values for the null hypothesis that agreement rates are the same in each treatment group and its corresponding control group. For the social incentive and four-week deadline groups, the control group is defined as the six-week deadline group. For the cash incentive group, the control group is defined as the four-week deadline group, which is the relevant comparison because the cash incentive group also faced a four-week deadline.

Table 2 shows that 67.6 percent of the referee invitations are accepted in the six-week group. The acceptance rate is slightly lower at 61.1 percent in the social incentive group, a difference that is marginally statistically significant ($p = 0.045$). The acceptance rate in the four-week deadline group is 64.1 percent, not significantly different from the acceptance rate in the six-week group. Lastly, the acceptance rate in the cash incentive group is 72.0 percent, which is significantly higher than the acceptance rate in the four-week deadline group ($p = 0.010$).

Consistent with this statistical evidence, the journal received a few emails showing that the treatments influenced the decisions by some referees to review papers. For example, a referee assigned to the social incentive group wrote, "I was surprised to receive an email stating the journal is posting referee times by names… I would like to withdraw my agreement to referee this paper. Sorry about that. I would have been happy to send in a report on time under a different policy."

Other referees' emails explain why cash incentives increase acceptance rates. For instance, a referee in the control group wrote, "I am sorry to have to decline this "invitation" to work for free… Can't Elsevier offer a better reward for the time they ask to devote to this screening?"

Overall, these results allay the concern that pushing referees to submit reviews quickly will make it difficult to find referees who are willing to submit reviews.

Outcome 2: Review Time

We now turn to the central outcome our treatments were designed to change: the time that referees take to submit their reviews. Naturally, we can only observe review times for referees who agree to submit reviews. Because the referees who accept invitations may differ across the treatment groups, differences in review times across groups reflect a combination of selection effects (changes in the composition of referees) and behavioral responses (changes in a given referee's behavior). For instance, referees who expect to be unable to submit a review quickly might be less likely to agree to review a paper under the shorter four-week deadline. This would reduce average review times in the four-week group via a selection effect even if referee behavior did not change.

Distinguishing between selection and changes in behavior is not critical for a journal editor seeking to reduce average review times, because it does not matter whether improvements come from getting faster referees or inducing a given set of referees to work faster. For the broader objective of learning about how incentives affect prosocial behavior, however, it is important to separate selection from behavioral responses. We therefore begin by assessing selection and then present estimates of treatment effects on review times both with and without adjustments for selection.

We evaluate the magnitude of selection effects in two ways. First, we compare pre-determined referee characteristics, such as tenure status and nationality, across the four groups. We find that these characteristics are generally quite similar across referees who accept invitations in the four groups (details available in Appendix Table 2b).

Second, we compare the *pre-experiment* review times of referees who agreed to review papers in each of the four experimental groups. For this analysis, we focus on the 67 percent of referees in our primary experimental sample who reviewed a manuscript for the journal before the experiment began (from November 2005 to February 15, 2010). All of these pre-experiment reviews were subject to a six week deadline. Figure 1 plots survival curves for review times according to the treatment group to which the referees were later assigned, using data from the most recent review

before the experiment began. These survival curves show the fraction of reviews that are still pending after a given number of days.[3]

The survival curves in the cash, four-week deadline, and six-week deadline groups are all very similar. Referees who agreed to submit a review under a shorter deadline or cash incentive treatment are no faster than those in the control group based on historical data. Non-parametric (Wilcoxon) tests for equality of the survival curves uncover no differences in review times across these three groups. We find marginally significant evidence ($p = 0.068$) that referees who agree to review papers in the social incentive group are slightly *slower* than those in the six-week control group. Hence, if anything, the social incentive treatment appears to induce slightly unfavorable selection in terms of referee speed. One explanation may be that diligent referees tend to be more concerned and anxious about their reputation and are hence less likely to accept the invitation with the social treatment. Overall, this evidence indicates that selection effects are modest and that differences in outcomes across the groups during the experiment are likely to be driven primarily by changes in referee behavior, with the possible exception of the social incentive group.

Figure 2 presents our main results on the impact of the treatments on review times during the primary experimental period. Panel A plots raw survival curves for reviews by treatment group. In Panel B, we adjust for selection using propensity score reweighting as in DiNardo, Fortin, and Lemieux (1996). We reweight the four-week, cash, and social incentive groups to match the six-week group on pre-experiment review times (including an indicator for having no pre-experiment data) using the procedure described in Appendix D. We report median survival times (the point at which 50% of reports have been submitted) and non-parametric Wilcoxon tests for the equality of the survival curves in each figure (see Appendix Table 3 for details).

In contrast with the survival curves in Figure 1, the survival curves in Figure 2 diverge sharply, showing that the treatments induced substantial changes in review times. Adjusting for differences in prior review times (Panel B) does not affect the results substantially, indicating that most of the change in review times is driven by changes in referee behavior rather than selection effects. We discuss next the impacts of each of the treatments in detail, starting with the shorter deadline and then turning to the cash and social incentives.

Shortening the deadline from six weeks (45 days) to four weeks (28 days) reduces median review times by 12.3 days, based on the baseline estimates in Panel A of Figure 2. Hence, we

[3] We include reviewers who do not submit reviews in these and all subsequent survival curves by censoring their spells at the point when editors make a decision on the paper.

estimate that shortening the deadline by one day reduces median review times by 12.3/(45-28) = 0.72 days. The effect is so large because nearly 25 percent of referee reports are submitted in the week between the reminder email and the deadline, and the shorter deadline simply shifts these reports forward. Before week three (shown by the first dashed line in Figure 2), the number of pending reports in the four-week and six-week groups is not very different; however, in week four, the survival curve for the four-week deadline group drops sharply relative to the six-week group. The four-week deadline thus appears to act as a nudge that makes referees work on their reports in the fourth week rather than the sixth week.

Providing a $100 cash incentive for submitting a report within four weeks reduces median review times by an additional eight days relative to the four week deadline. The cash incentive has powerful effects especially after referees receive the reminder email: nearly 50 percent of referees submit a report in the window between the reminder email and the deadline for receiving the cash payment. Missing the four week deadline simply postpones writing the report by a few weeks but costs $100. Consistent with what one would predict based on a standard model of intertemporal optimization, the survival curve is much flatter immediately after the four week deadline, as very few referees submit reports immediately after the cutoff for the cash payment. Nevertheless, because so many referees make an effort to meet the four week deadline, there are fewer reports pending even 10 weeks after the initial invitation in the cash incentive group relative to all the other groups.

The strong response to the cash incentive in the week before the deadline also supports the view that the cash incentive changes referee behavior, rather than the selection of referees who agree to review, as selection effects would be unlikely to generate such non-linear responses. Indeed, the response to the cash treatment is so large that one can show that selection effects account for very little of the impact using a non-parametric bounding approach, as in Lee (2009). Recall from Table 2 that referees in the cash group are 12.3=(72.0/64.1-1) percent more likely to accept review invitations than referees in the four-week group. Assuming that referees who accept the four week invitation would also have accepted the (more attractive) cash invitation, we can bound the selection effect by considering the worst case scenario in which the additional referees who accept the cash invitation have the shortest spells. For example, 66 percent of referees in the cash group submit their report within 28 days. If we exclude the 12.3 percent fastest referees in the cash group, we obtain a selection-adjusted lower bound of (66-12.3)/(100-12.3) = 61 percent submitting within 28 days. This remains well above the 36 percent of referees who submit a report within 28 days in the four-week group, showing that the difference in review times between the two groups cannot be caused by

selection. A similar bounding exercise implies that the difference in review times between the four-week and six-week groups also cannot be due to selection.

Figure 2 demonstrates that the direct incentive effect of money outweighs any crowd-out of intrinsic motivation to submit referee reports in a timely manner. To investigate the impact of monetary incentives on intrinsic motivation more directly, we study the behavior of referees for the six months *after* the cash incentive ended on May 9, 2011. A long literature in social psychology starting with the classic work of Deci (1971) predicts that cash rewards have negative long-run effects on prosocial behavior by eroding intrinsic motivation. Existing evidence for this effect is based primarily on lab experiments (Deci et al. 1999; Frey and Jegen 2001; Kamenica 2012). Our experiment offers a new test of this hypothesis in the field that complements earlier work on economic incentives and prosocial behavior in other settings (for example, Gneezy and Rustichini 2000; Gneezy, Meier, and Rely-Biel, 2011; Lacetera, Macis, and Slonim 2013).

In our application, the prediction from theories in which monetary payments crowd-out intrinsic motivation is that referees who had previously received cash incentives should become slower after they stop receiving cash payments—at least relative to referees in the four-week deadline group, who never received cash payments. We test this hypothesis in Figure 3, which plots survival curves for referees assigned to the four-week and cash incentive groups using data before May 9 vs. after May 9, when cash payments ended.[4] The survival curves for the four-week group are similar for invitations before and after May 9, indicating that review times do not vary significantly by invitation date. Referees assigned to the cash incentive group are much less likely to meet the 28 day deadline after May 9 than before May 9, when they were receiving cash rewards. However, there is no evidence that these referees become *slower* than those in the four-week comparison group, which is what one would expect if intrinsic motivation had been eroded. If anything, it appears that the cash treatment leads to some persistent improvements even after the incentive is removed, perhaps because referees have gotten in the habit of submitting reports slightly sooner.[5] We

[4] Of the referees who were assigned to the cash incentive group and accepted a review invitation after May 9 (after the cash rewards had ended), 47 percent did not receive an invitation to review a manuscript before May 9. To minimize selection effects, we include these referees in Figure 3 even though they never received the cash incentive treatment. The estimates in Figure 3 should therefore be interpreted as intent-to-treat estimates. Restricting the sample to the selected subset of referees who received prior invitations yields very similar results.

[5] One might be concerned that referees did not recognize that the cash incentive had stopped after May 9, biasing our comparisons in Figure 3. Two facts allay this concern. First, if referees mistakenly thought the cash reward was still in place after May 9, one would expect to see the post-May cash survival curve in Figure 3 to drop steeply in the week before the four-week deadline. This does not occur: the post-May cash survival curve tracks the four-survival curves almost perfectly prior to the deadline. Second, the cash incentive increased agreement rates from 64.1 percent (in the four-week group) to 72.0 percent prior to May 9, as shown in Table 2. This difference also disappears after

conclude that the temporary provision of monetary incentives does not have detrimental subsequent effects in the case of peer review.

Next, we turn to the social incentive treatment. We find a significant difference between the social incentive and control group survival curves when reweighting on pre-experiment durations in Figure 2b. The difference between the unweighted social and control survival curves in Figure 2a is smaller and statistically insignificant. This is because the social incentive treatment appears to induce slightly slower referees to accept review invitations, as shown in Figure 1. Once we adjust for this selection effect, we find that the social incentive treatment induces referees to work significantly faster, although the magnitude of the impact remains small. Based on the reweighted survival curves, we estimate that the social incentive reduces the median review time by 2.3 days.[6]

Finally, we explore the heterogeneity of the treatment effects by referee characteristics. We find no significant heterogeneity in treatment effects by several of the referee characteristics we collected: an indicator for holding an academic position, gender, and an indicator for working in the United States. However, we do find substantial heterogeneity in treatment effects between tenured and untenured referees, as shown in Figure 4. This figure replicates Figure 2a, dividing the sample into referees who had tenure at the time they were invited to review the manuscript (Panel A) and those who were not tenured at that time (Panel B). The shorter deadline has a significantly larger effect on untenured referees than tenured referees. Untenured referees make a clear effort to submit reports before the deadline, as evident from the sharp drop in the survival curve in Figure 4b just before the deadline for the four-week group. In contrast, tenured referees are not very sensitive to the shorter deadline.

The cash incentive improves performance substantially in both groups, but again the impact is larger among untenured referees: 78 percent of untenured referees submit reports before the deadline to receive the cash reward, whereas only 58 percent of tenured referees do so. While the cash incentive and shorter deadline have smaller effects on tenured referees, the social incentive has

---

May 9: 64.1 percent of referees previously assigned to the cash incentive group agree to do the review after May 9, compared with 65.4 percent in the four-week group during the same period.

[6] We evaluate the robustness of the treatment effect estimates using semi-parametric Cox hazard models in Appendix E. Consistent with the graphical evidence in Figure 2, we find that the cash incentive and 4 week deadlines substantially increase hazard rates of report submission, particularly in the week before the deadline. The social incentive treatment reduces review times significantly when controlling for differences in pre-experiment review times. These results, which are reported in Appendix Table 4, are robust to changes in the control vector and sample specifications. In Appendix Figure 2, we use all the data through the end of the experiment (October 26, 2011) rather than restricting the sample to the point at which cash treatments were stopped (May 9, 2011). The point estimates remain similar, but we obtain more precise estimates when using all the data as expected.

larger effects on tenured referees. Figure 4b shows that review times are almost identical in the social incentive and control groups for untenured referees. In contrast, tenured referees in the social incentive group submit reports significantly earlier than those in the control group, as shown in Figure 4a.

One explanation for why the social incentive treatment is more effective among tenured referees is that untenured referees are already concerned about their reputation with co-editors, who are typically senior colleagues in their field. In contrast, tenured referees might become more concerned about their professional reputation when they face social pressure.[7] Regardless of whether the heterogeneous effects are driven by this mechanism, the findings in Figure 4 suggest that social incentives can usefully complement other policy instruments by improving behavior among groups who are less responsive to cash incentives and nudges.

Outcome 3: Review Quality

Models of multi-tasking predict that if an agent is given an incentive to perform better in one aspect of a job (such as production speed), performance in other aspects of the job (such as quality) might deteriorate. Might the treatments that induce referees to submit reports more quickly also lead referees to submit lower-quality reviews?

We measure the quality of reviews in two ways. The first is an indicator for whether the editor follows the referee's recommendation with regard to whether the manuscript should be accepted, rejected, or revised and resubmitted. The second is the length of the referee report. While length is not equivalent to quality, one natural way in which referees might submit a report more quickly is by providing less detailed comments to authors, especially since only the editor knows the referee's identity.

Table 3, which is constructed in the same way as Table 2, shows the fraction of cases in which the editor follows the referee's recommendation (Panel A) and the median length of the referee report (Panel B) by treatment group. We find no statistically significant differences across the groups in the rate at which editors follow the referee's advice. We do, however, find that referees write shorter reports to authors under the social and cash incentive treatments. The median report is approximately 100 words (11 percent) shorter in the social and cash groups relative to the six-week

[7] Consistent with this explanation, we find that tenured referees are considerably slower than untenured referees in the control group, but behave like untenured referees in the social incentive group, as shown in Appendix Figure 3.

and four-week groups. These findings suggest that referees who rush to submit a report earlier because of explicit cash or social incentives might cut back slightly on the level of detail in their comments to authors. Interestingly, referees do not write shorter reports to meet the four-week deadline, consistent with the view that many referees begin writing reports only in the week after they receive a reminder.

Overall, we conclude that one can induce referees to submit reviews more quickly without reducing the quality of reviews significantly. Shorter deadlines have no adverse effect on either measure of quality, while cash and social incentives induce referees to write slightly shorter reports but do not affect the quality of the review as judged by the editor's ultimate decision.

Outcome 4: Spillover Effects on Other Journals

A natural concern with interventions that improve referee performance at one journal is that they may have negative spillover effects at other journals. Do referees who submit reviews more quickly at the *Journal of Public Economics* prioritize them over other referee reports? In this case, changes in journal policies might not improve the overall efficiency of the review process.

We test for such spillover effects using data from 20 other Elsevier journals in related subfields, such as the *Journal of Health Economics* and the *Journal of Development Economics* (see Appendix F for a complete list). We analyze referee invitations from other journals that are received (1) *after* referees have received an invitation from the *Journal of Public Economics* during the primary experimental period and (2) before December 31, 2011.

Specifically, we test whether referees' propensities to review manuscripts and their review times at other journals vary across our four treatment groups. Each observation in this analysis is a referee invitation at another journal. The mean agreement rate is approximately 60% in all four groups, with no statistically significant differences across the groups (see Appendix Table 5). Median review times are approximately 56 days in all four groups, again with no statistically significant differences across the groups (see Appendix Figure 4).[8]

Of course, referees must postpone some activity to prioritize submitting referee reports. The social welfare impacts of our treatments depend on what activities get postponed. If referees postpone activities with pure private benefits such as leisure, social welfare may increase because

[8] The similarity across the four groups in performance at other journals supports the view that the treatment effects at the *Journal of Public Economics* during the experimental period are driven by changes in referee behavior rather than selection effects.

referee reports have positive externalities. If on the other hand referees postpone working on their research or on other prosocial tasks, expediting referee reports could reduce welfare. If small delays in these other activities have little social cost, the welfare costs from such delays would be modest. Understanding the nature of crowd-out across different forms of prosocial behavior is an interesting question that we defer to future research.

Lessons for the Peer Review Process

Our results offer three lessons for the design of the peer review process at academic journals.

First, shorter deadlines are extremely effective in improving the speed of the review process. Moreover, shorter deadlines generate little adverse effect on referees' agreement rates, the quality of referee reports, or performance at other journals. Indeed, based on the results of the experiment, the *Journal of Public Economics* now uses a four week deadline for all referees.

Second, cash incentives can generate significant improvements in review times and also increase referees' willingness to submit reviews.[9] However, it is important to pair cash incentives with reminders shortly before the deadline. Some journals, such as the *American Economic Review*, have been offering cash incentives without providing referees reminders about the incentives; in this situation, sending reminders would improve referee performance at little additional cost.

Third, social incentives can also improve referee performance, especially among subgroups such as tenured professors who are less responsive to deadlines and cash payments. Light social incentives, such as the *Journal of Financial Economics* policy of posting referee times by referee name, have small effects on review times. Stronger forms of social pressure – such as active management by editors during the review process in the form of personalized letters and reminders – could potentially be highly effective in improving efficiency. It would be useful to test this hypothesis in future work using an experiment in which editors are prompted to send personalized reminders to referees at randomly chosen times.

More generally, our findings show that it is possible to substantially improve the efficiency of the peer review process with relatively low-cost interventions, demonstrating the value of studying the peer review process empirically (as in Card and DellaVigna 2012). Our results reject the view

[9] These findings contrast with the results of Squazzoni, Bravo, and Takacs (2013), who argue that monetary rewards decrease the quality and efficiency of the review process based on a lab experiment designed to simulate peer review. Our results might differ because the peer review process requires referees to invest considerable time to read papers and write referee reports, unlike the investment game studied in this lab experiment.

that the review process in economics is much slower than in other fields, such as the natural sciences, purely because economics papers are more complex or difficult to review.

Lessons for Increasing Prosocial Behavior

Beyond the peer review process, our results also offer some insights into the determinants of prosocial behavior more broadly.

First, attention matters: reminders and deadlines have significant impacts on behavior. Nudges that bring the behavior of interest to the top of individuals' minds are a low-cost way to increase prosocial behavior, consistent with a large literature in behavioral economics (Thaler and Sunstein 2008).

Second, monetary incentives can be effective in increasing some forms of prosocial behavior. We find no evidence that intrinsic motivation is crowded out by financial incentives in the case of peer review, mirroring the results of Lacetera, Macis, and Slonim (2013) in the case of blood donations. While crowd-out of intrinsic motivation could be larger in other settings, these results show that one should not dismiss corrective taxes or subsidies as a policy instrument simply because the behavior one seeks to change has an important prosocial element.

Finally, social incentives can be effective even when other policy instruments are ineffective. This result echoes findings in other settings – such as voting (Gerber, Green, and Larimer, 2008), campaign contributions (Perez-Truglia and Cruces 2013), and energy conservation (Allcott 2011) – and suggests that social incentives are a useful complement to price incentives and behavioral nudges.

Acknowledgements

We thank Nava Ashraf, David Autor, Stefano DellaVigna, Hilary Hoynes, Damon Jones, Emir Kamenica, Lawrence Katz, Henrik Kleven, Ulrike Malmendier, Monica Singhal, Timothy Taylor, and numerous seminar participants for helpful discussions and comments. We thank Jenny Henzen and the Elsevier staff for supporting the project and providing data and Liz Anderson for implementing the experiment. Greg Bruich, Jessica Laird, Keli Liu, Alex Olssen, and Heather Sarsons provided outstanding research assistance. Financial support from the Lab for Economic Applications and Policy at Harvard and the National Science Foundation is gratefully acknowledged. A de-identified version of the referee-level dataset used in this paper can be downloaded from this website.

References

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics*, 95(9-10): 1082-1095.

Andreoni, James, and Douglas B. Bernheim. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77(5): 1607-1636.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652-1678.

Card, David and Stefano DellaVigna. 2012. "Revealed Preferences for Journals: Evidence from Page Limits," NBER Working Paper No. 18663.

Deci, Edward L. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology*, 18(1): 105-15.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125(6), 627-668.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127(1): 1-56.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5): 1001-1044.

Ellison, Glenn. 2002. "The Slowdown of the Economics Publishing Process." *Journal of Political Economy*, 110(5): 947-993.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism". *Nature*, 425(6960): 785-791.

Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory: A Survey of Empirical Evidence." *Journal of Economic Surveys*, 15(5): 589-611.

Gerber, Alan S., Donald P. Green, and Cristopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment." *American Political Science Review*, 102(1) : 33-48.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives*, 25(4): 191-210.

Gneezy, Uri, and Aldo Rustichini. 2000. "A Fine is a Price." *Journal of Legal Studies*, 29(1): 1-17.

Holmstrom, Bengt., and Paul Milgrom. 1991. "Multitask Principal-agent Analyses: Incentives Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization*, 7: 24-52.

Kamenica, Emir. 2012. "Behavioral Economics and Psychology of Incentives." *Annual Review of Economics*, 4(1) : 427-452.

Lacetera, Nicola, Mario Macis, and Robert Slonim. 2013. "Economic Rewards to Motivate Blood Donations." *Science*, 340(6135): 927-928.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics*. John H. Kagel and Alvin E. Roth, eds., Princeton: Princeton University Press: 111-194.

Lee, David. S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *Review of Economic Studies*, 76(3): 1071-1102.

Perez-Truglia, Ricardo and Guillermo Cruces. 2013. "Social Incentives in Contributions: Field Experiment Evidence from the 2012 U.S. Presidential Campaigns." Harvard University working paper.

Squazzoni, Flaminio, Giangiacomo Bravo, Károly Takács. 2013. "Does incentive provision increase the quality of peer review? An experimental study." *Research Policy*, 42(1), 287-294.

Thaler, Richard H. and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Titmuss, Richard M. 1971: *The Gift Relationship*. London: George Allen and Unwin.

Vesterlund, Lise. 2014. "Voluntary Giving to Public Goods: Moving Beyond the Linear VCM," in *Handbook of Experimental Economics*. John H. Kagel and Alvin E. Roth, eds., Volume 2, Princeton: Princeton University Press.

Online Appendices

Appendix A. Invitation Emails

Control Invitation Email

Subject: Reviewer Invitation from JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS

Dear REFEREE,

You are invited to review the above-mentioned manuscript for publication in the Journal of Public Economics. The manuscript's abstract is at the end of this email.

If you accept this invitation, I would be very grateful if you would return your review by **July 21, 2010** (6 weeks from now).

Please choose one of the following options to proceed:

1) If you are willing to review this manuscript, please click: Agree to Review

2) If you are not able to review this manuscript, please click: Decline to Review

3) If you would like to view the manuscript before making a decision, please click: View Manuscript

To assist you in the reviewing process, I am delighted to offer you full access to Scopus (the largest abstract and citation database of research information) for 30 days. With Scopus you can search for related articles, references and papers by the same author. You may also use Scopus for your own purposes at any time during the 30-day period. If you already use Scopus at your institute, having this 30 day full access means that you will also be able to access Scopus from home. Access instructions will follow once you have accepted this invitation to review.

Yours sincerely,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Four-Week Invitation Email

Subject: Reviewer Invitation from JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS

Dear REFEREE,

You are invited to review the above-mentioned manuscript for publication in the Journal of Public Economics. The manuscript's abstract is at the end of this email.

If you accept this invitation, I would be very grateful if you would return your review by **July 4, 2010** (4 weeks from now).

Please choose one of the following options to proceed:

1) If you are willing to review this manuscript, please click: Agree to Review

2) If you are not able to review this manuscript, please click: Decline to Review

3) If you would like to view the manuscript before making a decision, please click: View Manuscript

To assist you in the reviewing process, I am delighted to offer you full access to Scopus (the largest abstract and citation database of research information) for 30 days. With Scopus you can search for related articles, references and papers by the same author. You may also use Scopus for your own purposes at any time during the 30-day period. If you already use Scopus at your institute, having this 30 day full access means that you will also be able to access Scopus from home. Access instructions will follow once you have accepted this invitation to review.

Yours sincerely,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Cash Invitation Email

Subject: Reviewer Invitation from JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS

Dear REFEREE,

You are invited to review the above-mentioned manuscript for publication in the Journal of Public Economics. The manuscript's abstract is at the end of this email.

If you accept this invitation, I would be very grateful if you would return your review by **July 4, 2010** (4 weeks from now). As a token of appreciation for timely reviews, you will receive a **$100** Amazon.com® Gift Card* if you submit your report before the due date. The Journal of Public Economics will automatically email you a gift card code within a day after we get your report (no paperwork required).

Please choose one of the following options to proceed:

1) If you are willing to review this manuscript, please click: Agree to Review

2) If you are not able to review this manuscript, please click: Decline to Review

3) If you would like to view the manuscript before making a decision, please click: View Manuscript

To assist you in the reviewing process, I am delighted to offer you full access to Scopus (the largest abstract and citation database of research information) for 30 days. With Scopus you can search for related articles, references and papers by the same author. You may also use Scopus for your own purposes at any time during the 30-day period. If you already use Scopus at your institute, having this 30 day full access means that you will also be able to access Scopus from home. Access instructions will follow once you have accepted this invitation to review.

Yours sincerely,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Social Invitation Email

Subject: Reviewer Invitation from JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS

Dear REFEREE,

You are invited to review the above-mentioned manuscript for publication in the Journal of Public Economics. The manuscript's abstract is at the end of this email.

If you accept this invitation, I would be very grateful if you would return your review by **July 21, 2010** (6 weeks from now). In the interest of improving transparency and efficiency in the review process, Elsevier will **publish referee times by referee name**, as currently done by the Journal of Financial Economics at this website. The referee times for reports received between Jan 1, 2010 and Dec 31, 2010 will be posted on the Journal of Public Economics website in January 2011. Note that referee anonymity will be preserved as authors only know the total time from submission to decision (and not individual referee's times).

Please choose one of the following options to proceed:

1) If you are willing to review this manuscript, please click: Agree to Review

2) If you are not able to review this manuscript, please click: Decline to Review

3) If you would like to view the manuscript before making a decision, please click: View Manuscript

To assist you in the reviewing process, I am delighted to offer you full access to Scopus (the largest abstract and citation database of research information) for 30 days. With Scopus you can search for related articles, references and papers by the same author. You may also use Scopus for your own purposes at any time during the 30-day period. If you already use Scopus at your institute, having this 30 day full access means that you will also be able to access Scopus from home. Access instructions will follow once you have accepted this invitation to review.

Yours sincerely,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Appendix B. Reminder and Thank You Emails

Control Group Reminder Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for agreeing to review this manuscript for the JPubE. I am writing to remind you that I would appreciate receiving your review **by July 21, 2010**, in a week.

You may submit your comments online in our editorial system by clicking here. Please login as a Reviewer using the username and password I sent you in my first email.

You may access the manuscript by selecting the "Pending Assignments" link on your Main Menu page. To submit your comments, please click on the "Submit Reviewer Recommendation" link.

With kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Four-Week Deadline Reminder Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for agreeing to review this manuscript for the JPubE. I am writing to remind you that I would appreciate receiving your review **by July 4, 2010**, in a week.

You may submit your comments online in our editorial system by clicking here. Please login as a Reviewer using the username and password I sent you in my first email.

You may access the manuscript by selecting the "Pending Assignments" link on your Main Menu page. To submit your comments, please click on the "Submit Reviewer Recommendation" link.

With kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Cash Incentive Reminder Email

Dear REFEREE,

Thank you for agreeing to review this manuscript for the JPubE. I am writing to remind you that I would appreciate receiving your review **by July 4, 2010**, in a week. As a token of gratitude for timely reviews, you will receive a **$100** Amazon.com® Gift Card* if you submit your report before the due date. The Journal of Public Economics will automatically email you a gift card code within a day after we get your report (no paperwork required).

You may submit your comments online in our editorial system by clicking here. Please login as a Reviewer using the username and password I sent you in my first email.

You may access the manuscript by selecting the "Pending Assignments" link on your Main Menu page. To submit your comments, please click on the "Submit Reviewer Recommendation" link.

With kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Social Incentive Reminder Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for agreeing to review this manuscript for the JPubE. I am writing to remind you that I would appreciate receiving your review **by July 21, 2010**, in a week. In the interest of improving transparency and efficiency in the review process, Elsevier will **publish referee times by referee name**, as currently done by the Journal of Financial Economics at this website. The referee times for reports received between Jan 1, 2010 and Dec 31, 2010 will be posted on the Journal of Public Economics website in January 2011. Note that referee anonymity will be preserved as authors only know the total time from submission to decision (and not individual referee's times).

You may submit your comments online in our editorial system by clicking here. Please login as a Reviewer using the username and password I sent you in my first email.

You may access the manuscript by selecting the "Pending Assignments" link on your Main Menu page. To submit your comments, please click on the "Submit Reviewer Recommendation" link.

With kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Control Group Thank You Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for your review of this manuscript.

You may access your review comments and the decision letter (when available) by logging onto the Elsevier Editorial System. Please login as a Reviewer.

Kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Four Week Deadline Thank You Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for your review of this manuscript.

You may access your review comments and the decision letter (when available) by logging onto the Elsevier Editorial System. Please login as a Reviewer.

Kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Cash Incentive Thank You Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for your review of this manuscript. As a token of appreciation for timely reviews, here is your $100 Amazon.com® Gift Card* code: Claim Code. You are able to use it any time to make purchases at Amazon.com without any paperwork. If you experience any problems with it, please do not hesitate to contact me at jpubec@gmail.com.

You may access your review comments and the decision letter (when available) by logging onto the Elsevier Editorial System. Please login as a Reviewer.

Kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Social Incentive Thank You Email

Subject: Reminder to review for JPubE
Ref. No.: JPUBE-D-10-00001
Title: TITLE
Editor: CO-EDITOR
Author(s): AUTHORS
Journal of Public Economics

Dear REFEREE,

Thank you for your review of this manuscript.

You may access your review comments and the decision letter (when available) by logging onto the Elsevier Editorial System. Please login as a Reviewer.

As you may remember, Elsevier will **publish referee times by referee name**, as currently done by the Journal of Financial Economics at this website. Your time of 27 days for this review will be posted on the Journal of Public Economics website in January 2011. Note that referee anonymity will be preserved as authors only know the total time from submission to decision (and not individual referee's times).

Kind regards,

Liz Anderson
Senior Editorial Assistant
Journal of Public Economics

Appendix C. Data Sources and Variable Definitions

Data from the Journal of Public Economics. Our primary source of data is the Elsevier online editorial system. We downloaded data from this system on July 22, 2012 for the analysis reported in the paper. We use data on all referees invited to review a new submission between February 15, 2010 and October 26, 2011. We exclude 15 observations that were contaminated (e.g. by letters with errors) and 5 observations in which the referee did not receive the email invitation. Referee requests for revisions are excluded from the experiment and are always subject to the 6 week deadline.

The Elsevier data system records time of invitation, agreement and submission of the report. Using these data, we generate an indicator for accepting the invitation, the turnaround time in days, and month of invitation. We obtain data on referees' review invitations and turnaround times prior to the start of the experiment from the same database, which contains information going back to November 1, 2005. The online system uses a single numerical identifier for a referee; we consolidate a few cases where referees have multiple ID's manually using the reviewer's name and affiliation.

The editorial system also stores all referee reports, which are submitted either as file attachments or as plain text via an online form. We define word counts for the referee reports as the sum of the words in the online text forms and any attachments. We use a similar procedure to measure the word count of manuscripts as well as the number of tables and equations in each manuscript. Note that these automated counting procedures do not always deliver accurate counts, but we expect such measurement error to be balanced across the treatment groups.

Each referee must select a recommendation for the manuscript on an online menu (accept, revise-and-resubmit, or reject). We use this information to define an indicator for whether the editor follows the referee's recommendation on whether or not to reject the submission, grouping the accept and revise-and-resubmit categories into a single category.

Demographics. We collected demographic information by locating referees' CVs online. We downloaded these CVs during Fall 2010, with an update for new referees in November 2011. We use these CVs to define indicators for gender, tenure status, working in the U.S., and working in an academic position. Tenure status is defined as being a full professor at a university or mentioning tenure on the CV for any other position. Working in the U.S. is based on the employer's address and an academic position is defined as having an affiliation with a university. We code these variables as missing for referees for whom we were unable to locate CVs online or whose CVs did not contain the relevant information. We located CVs for 92.9% of the 1,606 referee reports in our primary (February 15, 2010 to May 9, 2011) sample.

Data from Other Journals. We obtain data from other Elsevier journals (listed in Appendix F) from the Elsevier editorial system. We compiled the longest histories available in the system for each journal. The available data vary across the journals, with the earliest records going back to November 2005. We use data up to December 31, 2011 from other journals. Elsevier does not use a unique identifier for referees across journals. We therefore linked referees to their performance at other journals based on email addresses (after extensive manual cleaning to match text fields).

Appendix D. Reweighting Methodology

This appendix describes the reweighting procedure used to construct Figure 2b. We first discretize each referee's most recent pre-experiment review time into eight bins, $b=1,\ldots,8$: seven monthly indicators for the pre-experiment review time if available (<30, 30-59, 60-89, 90-119, 120-149, 150-179, and ≥180 days) and an indicator for having no pre-experiment data.

To reweight the social incentive group to match the six-week control group, we take the referees assigned to those two groups and calculate the fraction of observations in bin b in the social incentive group, which we denote by p_b. The fraction of observations in bin b in the six-week control group is $1-p_b$. We weight each observation i by $(1-p_{b(i)})/p_{b(i)}$ when estimating the survival curve for the social group, where $b(i)$ denotes the bin to which observation i belongs.

We reweight the cash and four-week groups to match the control group on pre-experiment durations using the same approach. The survival curve for the six-week (control) group is unchanged by definition.

To adjust for differences in pre-experiment durations when testing for the equality of the survival curves, we conduct unweighted Wilcoxon tests that are stratified by the bin variable b.

Appendix E. Hazard Model Estimates of Treatment Effects on Review Times

This appendix presents estimates of the impacts of the treatments on review times using Cox hazard models. Let h_{it} denote the hazard rate of submitting a referee report t days after the invitation (i.e., the probability of submitting a report on day t conditional on not submitting prior to day t). Let α_{mt} denote the baseline hazard rate for referees in the six-week control group who receive an invitation to review a paper in month m of the experiment. We stratify the baseline hazards by invitation month to account for any differences over time in referee behavior. The Cox hazard model specification is

$$h_{it} = \alpha_{mt}\exp(\beta_1 fourweek_i + \beta_2 cash_i + \beta_3 social_i + \beta_4 postcash_i + \gamma X_i)$$

In this specification, the *fourweek* indicator is 1 for both the four-week and cash incentive groups, who face four-week deadlines. Hence, the coefficients on the cash variables represent the effect of the cash treatment over and above the four-week deadline effect. The *cash* variable is an indicator for being in the cash incentive group while cash rewards were offered (prior to May 9); it is defined as 0 for all review invitations after May 9. The *social* variable is an indicator for being in the social incentive group. The *postcash* variable is an indicator for previously being in the cash incentive group; it is defined as 0 for all review invitations before May 9. The vector X_i is a set of controls that we vary across specifications. We censor spells that last for more than 20 weeks at 140 days to reduce the influence of outliers and we cluster standard errors by referee.

We report estimates from variants of this model in Appendix Table 4. We begin in Column 1 by estimating the hazard model with no additional controls (no X vector). We use the extended sample, which includes all invitations from February 15, 2010 to October 26, 2011, in this specification. Consistent with the results in Figure 2, we find that both the four week deadline and the cash incentive substantially increase hazard rates of submitting reports, i.e. reduce review times. The estimates β_j can be interpreted as the percentage impact of the variable on the baseline hazard rate. For example, the coefficient of 0.266 on the four-week indicator implies that the hazard rate is 26.6% higher on average for referees facing a four-week deadline relative to those facing a six-week deadline. The point estimate on the post-cash indicator is positive and marginally significant, supporting the view that there is no crowd-out of intrinsic motivation for referees who previously received cash incentives. The estimated impact of social incentives is small and statistically insignificant. This is consistent with Figure 2a, which shows that we do not detect significant differences between the social incentive and control groups when comparing raw distributions of review times.

Column 2 adds a rich set of controls for referee and manuscript characteristics to the specification in Column 1. We control for a referee's pre-experiment review times by including bi-monthly indicator variables (up to 6 months) for the review time for each of the previous three

referee reports. We also include controls for tenure, working in the U.S., working in academia, and referees' agreement rates to invitations in the available history of the *Journal of Public Economics*, as well as the number of words, tables, and equations in the article reviewed by the referee. The covariates are set to 0 if they are missing and all specifications include indicators for the observation having a missing value of the covariate. Hence, the sample is exactly the same as in column 1.

The inclusion of the controls increases the estimated impact of the social incentive treatment significantly. This result confirms the pattern in Figure 2b, showing that referees who agree to review manuscripts under the social incentive treatment are slightly negatively selected in terms of review times. Adjusting for these differences in pre-experiment turnaround times and other observables, we find that the social incentive treatment increases hazard rates by approximately 18% relative to the six-week deadline. The cash and four-week treatments continue to have highly significant impacts on hazard rates with controls.

Column 3 replicates Column 2 restricting the sample to the primary experimental period from February 15, 2010 to May 9, 2011, when the cash reward was offered, as in the main text. We find that the impacts of the shorter deadline, cash incentives, and social incentives are all very similar when we restrict to this subset of referee reports.

The preceding specifications all assume that the treatments have a constant percentage impact on hazard rates throughout the spell. However, the non-parametric survival curves in Figure 2 show that this proportional hazards assumption is not a good approximation. In particular, the four week deadline and cash incentives have much greater effects before the deadline than after the deadline, as one would expect. To account for these responses, in Column 4 we estimate a Cox model that extends Column 2 to permit time-varying covariates. We include indicators for being near the deadline and past the deadline, which represent the period one week before and after the due date, respectively. We also interact these indicators with the cash and post-cash indicators to capture the greater impacts of the cash treatment before the deadline.

Consistent with the patterns in Figure 2, the time varying covariates are highly significant: hazard rates are 86% higher in the week before the deadline and 119% higher in the week after the deadline. The cash treatment increases hazard rates by 100% in the week before the deadline but does not have a statistically significant effect in the week after the deadline. The post-cash treatment has no time-varying effect, as one would expect. The estimated impact of the social incentive remains similar to the other specifications. Overall, the model with time-varying covariates confirms the results in Figure 2 and shows that all three treatments have significant effects on referee behavior.

Appendix F. List of Other Journals Used to Assess Spillover Effects

Economics & Human Biology
Economics Letters
Energy Economics
European Economic Review
European Journal of Political Economy
Games and Economic Behavior
Journal of Banking & Finance
Journal of Comparative Economics
Journal of Corporate Finance
Journal of Development Economics
Journal of Economic Behavior & Organization
Journal of Economic Psychology
Journal of Environmental Economics and Management

Journal of Health Economics
Journal of International Economics
Journal of Monetary Economics
Journal of Urban Economics
Labour Economics
Regional Science and Urban Economics
Resource and Energy Economics

Appendix G. Summary of Appendix Tables and Figures.

Appendix Figure 1 depicts the timeline of the refereeing process during the experiment.

Appendix Figure 2 replicates Figure 2 from the text using the full experimental period from February 15, 2010 to October 26, 2011. This figure includes the period after May 9, when the cash reward was stopped, for the four week, six week, and social incentive groups. For the cash group, we continue to use data only up to May 9.

Appendix Figure 3 plots a subset of the survival curves reported in Figure 4 in the main text on a single figure to show that tenured referees have longer turnaround times than untenured referees in the control group, but behave like untenured referees when facing social pressure.

Appendix Figure 4 shows survival curves for review times at other Elsevier journals by the treatment group to which referees were assigned at the Journal of Public Economics.

Appendix Table 1 presents the summary statistics for referee invitations sent between February 15, 2010 and May 9, 2011, the time period when the cash reward was offered.

Panel A of Appendix Table 2 presents randomization tests for the set of referees invited during the primary experimental period. Panel B replicates Panel A in the subsample of referees who accept the invitations to test for selection effects.

Appendix Table 3 presents estimates of treatment effects on median review times.

Appendix Table 4 presents Cox hazard model estimates of the effects of the treatments on review times.

Appendix Table 5 reports estimates of the effects of the treatments at the Journal of Public Economics on acceptance rates and review times at other Elsevier journals during the experimental period.

Table 1

Description of Treatment Groups

Group:	6 Week (1)	Social (2)	4 Week (3)	Cash (4)
Deadline	6 weeks (45 days)	6 weeks (45 days)	4 weeks (28 days)	4 weeks (28 days)
Incentives	None	Review time posted online at end of year	None	$100 Amazon gift card if deadline met
Duration of Intervention	Feb 15, 2010 to Oct 26, 2011	Feb 15, 2010 to Oct 26, 2011	Feb 15, 2010 to Oct 26, 2011	Feb 15, 2010 to May 9, 2011

Notes: This table describes the four treatment groups to which referees were randomly assigned. Every referee was assigned permanently to one group; referees never changed groups. Referees were notified about the conditions of the review request upon invitation and were sent a reminder 1 week before the deadline. Examples of these invitation and reminder emails are shown in Appendices A and B. Cash incentives were stopped for invitations after May 9, 2011; after that point, referees assigned to the cash incentive group simply faced a 4 week deadline, with no incentives. The other treatments were implemented without any changes for the full duration of the experiment, from February 15, 2010 to October 26, 2011.

Table 2

Fraction of Referees who Accept Review Invitation by Treatment Group

Group:	6 Week (1)	Social (2)	4 Week (3)	Cash (4)
Percent who accept invitation	67.6%	61.1%	64.1%	72.0%
	(2.14)	(2.43)	(2.23)	(2.17)
p-value for equality with control		0.045	0.252	0.010
Observations	639	568	626	590

Notes: This table shows the percentage of referees who accept invitations to review in each treatment group. We restrict the sample to invitations sent between February 15, 2010 and May 9, 2011, the time period when the cash reward was offered. Standard errors, clustered at the referee level, are reported in parentheses. We also report p-values for the null hypothesis that agreement rates are the same in each treatment group and its corresponding control group. For the social and 4 week groups, the control group is defined as the 6 week deadline group. For the cash incentive group, the control group is defined as the 4 week deadline group, which is the relevant comparison because the cash incentive group also faced a 4 week deadline. The number of observations (referee report invitations) is reported in the last row.

Table 3

Measures of Review Quality by Treatment Group

Group:	6 Week (1)	Social (2)	4 Week (3)	Cash (4)
A. Agreement between Editor Decision and Referee's Recommendation				
Editor Follows Referee's Recommendation	77.9%	76.2%	77.5%	76.2%
	(2.00)	(2.34)	(2.20)	(2.15)
p-value for equality with control		0.585	0.884	0.686
Observations	403	324	373	404
B. Length of Referee Report				
Median Number of Words in Referee Report	877	757	864	786
	(29.1)	(32.5)	(30.3)	(29.2)
p-value for equality with control		0.006	0.757	0.064
Observations	401	321	369	399

Notes: This table shows the effects of the treatments on review quality. The sample includes all referees who received invitations sent between February 15, 2010 and May 9, 2011 (the period when the cash reward was offered) and submitted a report. In Panel A, the outcome is the fraction of reports in which the editor's decision (reject vs. accept/revise-and-resubmit) matches the referee's recommendation. We report standard errors in parentheses. Standard errors are clustered by referee in Panel A (but not Panel B). We also report p-values for the null hypothesis that the percentages are the same in each treatment group and its corresponding control group. For the social and 4 week groups, the control group is defined as the 6 week deadline group. For the cash incentive group, the control group is defined as the 4 week deadline group, which is the relevant comparison because the cash incentive group also faced a 4 week deadline. The number of observations (referee reports submitted) is reported in the last row. In Panel B, the outcome is the median number of words in the referee report. Standard errors are reported in parentheses and the p-values are for hypothesis tests analogous to those in Panel A. The number of observations is the number of submitted reports for which we were able to obtain automated word counts of report length.

Appendix Table 1

Summary Statistics for Experimental Sample

Variable	Mean (1)	Std. Dev. (2)	Median (3)
Invitation to Referee (N = 2,423)			
Agreed to submit review	66.2%	47.3%	
Refereeing statistics conditional on agreement (N = 1,605)			
Reviews censored (not submitted)	6.3%	24.3%	
Review time conditional on submitting review (days)	44.9	28.6	41.0
New referee (no historical data)	32.7%	46.9%	
Referee Characteristics (N = 1,157)			
Agreed to do 1 job during experiment	74.9%	43.4%	
Agreed to do 2 jobs during experiment	16.4%	37.1%	
Agreed to do 3+ jobs during experiment	8.6%	28.1%	
Tenured	54.6%	49.8%	
Academic	92.4%	26.5%	
American	52.5%	50.0%	
Female	12.3%	32.9%	

Notes: This table reports summary statistics for referee invitations sent between February 15, 2010 and May 9, 2011, the time period when the cash reward was offered. The first section of the table shows the fraction of referee requests that were accepted. The second section reports statistics for the subsample of referee requests that were accepted. A review is defined as censored if it is not submitted before the editor makes a decision on the paper. The summary statistics for review times are based on the subsample of submitted reviews. The third section of the table reports statistics on the referees who accepted the invitation and for whom the relevant information is available. See Appendix C for the definitions of the variables used in this table.

Appendix Table 2
Randomization and Selection Tests

Group:	6 Week	Social	4 Week	Cash	Equality test p-value
	(1)	(2)	(3)	(4)	(5)
A. Randomization Tests: Full Sample of All Invited Referees					
Has pre-experiment data	58.2%	63.6%	66.0%	66.6%	0.07
Prior agreement rate	73.8%	70.3%	77.4%	73.8%	0.17
Prior median turnaround time	54.1	57.1	55.2	58.6	0.24
Tenured	60.2%	68.4%	59.8%	65.9%	0.07
Academic	90.2%	93.4%	93.0%	93.4%	0.51
American	53.4%	58.6%	53.8%	51.2%	0.30
Female	12.2%	8.3%	13.4%	11.8%	0.20
Observations	639	568	626	590	
B. Selection Tests: Sample of Referees who Accepted Invitations					
Has pre-experiment data	64.1%	65.1%	71.6%	68.2%	0.25
Prior agreement rate	82.5%	79.2%	87.3%	81.5%	0.03
Prior median turnaround time	52.1	57.1	53.8	57.0	0.19
Tenured	50.8%	59.9%	50.9%	59.4%	0.09
Academic	91.0%	96.2%	91.8%	93.0%	0.09
American	56.5%	57.9%	55.9%	51.1%	0.51
Female	14.1%	9.9%	16.1%	12.6%	0.30
Observations	432	347	401	425	

Notes: This table reports summary statistics of pre-experiment variables by treatment group. Panel A uses all referees invited to review a paper between February 15, 2010 and May 9, 2011 (the period when the cash reward was offered). Panel B replicates Panel A for the selected sample of referees who accepted the invitation to review. Column (5) reports the p-value for a test of equality of the coefficients across all four groups, clustering standard errors by referee (except for median review times). Has pre-experiment data is an indicator for having information in the editorial system at some point between November 1, 2005 and February 15, 2010, when the experiment began. Prior agreement rate is the fraction of reviews that the referee accepted during that period. Prior median review time is the median review time for the three most recent manuscripts reviewed before the experiment (among referees who reviewed manuscripts before the experiment). Tenured is an indicator for having tenure (based on CV's posted online) when the referee received the invitation; academic is an indicator for being in an academic position. American is an indicator for a US-based employer, and Female is a gender indicator from data collected manually. The number of observations in Panel A is the number of referee report invitations; in Panel B, it is the number of accepted invitations.

Appendix Table 3

Median Review Times by Treatment Group

Group: Sample	6 Week (1)	Social (2)	4 Week (3)	Cash (4)
Full Sample	47.8	45.9	35.5	27.5
	(1.02)	(0.84)	(1.60)	(0.24)
	432	347	401	425
Tenured Referees	50.4	46.8	44.1	27.7
	(1.58)	(1.62)	(2.67)	(0.49)
	203	199	189	236
Untenured Referees	45.9	45.5	31.7	27.3
	(0.83)	(0.75)	(1.70)	(0.27)
	197	133	182	161

Notes: This table shows the effects of the treatments on median review times. These estimates are reported in Figure 2a and Figure 4 and are reproduced here with standard errors as a reference. The sample includes all referees who accepted invitations sent between Feb. 15, 2010 and May 9, 2011 (the period when the cash reward was offered). Standard errors and number of observations are reported below each estimate. The first row of estimates uses the full-sample; the second and third rows restrict the sample to referees who were tenured vs. untenured at the time of the experiment. Tenure status was collected from CV's posted online and hence is not available for all referees. See Appendix C for further details.

Appendix Table 4

Cox Hazard Model Estimates of Treatment Effects on Review Times

	Extended Sample No Controls (1)	Extended Sample With Controls (2)	Primary Sample (3)	Time-Varying Covariates (4)
4 week deadline	0.266***	0.393***	0.418***	0.391***
	(0.0720)	(0.0720)	(0.0783)	(0.0738)
Cash	0.388***	0.502***	0.485***	0.161
	(0.0969)	(0.0953)	(0.0968)	(0.113)
Social	0.0769	0.179**	0.152*	0.187**
	(0.0637)	(0.0657)	(0.0746)	(0.0693)
Post-cash	0.185	0.255*		0.242
	(0.111)	(0.114)		(0.139)
Near deadline				0.864***
				(0.0921)
Past deadline				1.188***
				(0.0972)
Cash near deadline				1.007***
				(0.169)
Cash past deadline				0.362
				(0.216)
Post-cash near deadline				0.00595
				(0.231)
Post-cash past deadline				-0.109
				(0.269)
Controls		X	X	X
Number of spells	2,212	2,212	1,605	2,212

Notes: This table reports coefficients from Cox proportional hazard models, with standard errors clustered by referee in parentheses. The asterisks represent statistical significance: * p<0.05, ** p<0.01, *** p<0.001. The point estimates can be interpreted as the percentage impact of the variable on the baseline hazard rate (which measures hazards in the 6 week control group). Columns 1, 2 and 4 report estimates from the extended sample, which includes all invitations from February 15, 2010 to October 26, 2011. Column 3 reports estimates from the baseline sample, which includes invitations from February 15, 2010 to May 9, 2011 (the period during which the cash reward was offered). In all four columns, baseline hazards are stratified by invitation month and spells that last for more than 20 weeks (140 days) are censored at 140 days. The 4 week deadline indicator is 1 for both the four-week and cash incentive groups, who face four-week deadlines. The cash variable is an indicator for being in the cash incentive group while cash rewards were offered (prior to May 9); it is defined as 0 for all review invitations after May 9. The post-cash variable is an indicator for previously being in the cash incentive group; it is defined as 0 for all review invitations before May 9. The social variable is an indicator for being in the social incentive group. Columns 2-4 control for a referee's pre-experiment review times by including bimonthly indicator variables (up to 6 months) for the review time for each of the previous three referee reports. They also include controls for tenure, U.S. residence, working in academia, and the fraction of reviews the referee accepted prior to the start of the experiment at the *Journal of Public Economics*, as well as the number of words, tables, and equations in the article reviewed by the referee. See Appendix C for definitions of all of these variables. Covariates are set to 0 if they are missing and all specifications include indicators for the observation having a missing value of the covariate. Column 4 includes terms allowing for time-varying hazard rates. Near and past deadline represent the period one week before and after the due date, respectively. These indicators are also interacted with the cash and post-cash indicators.

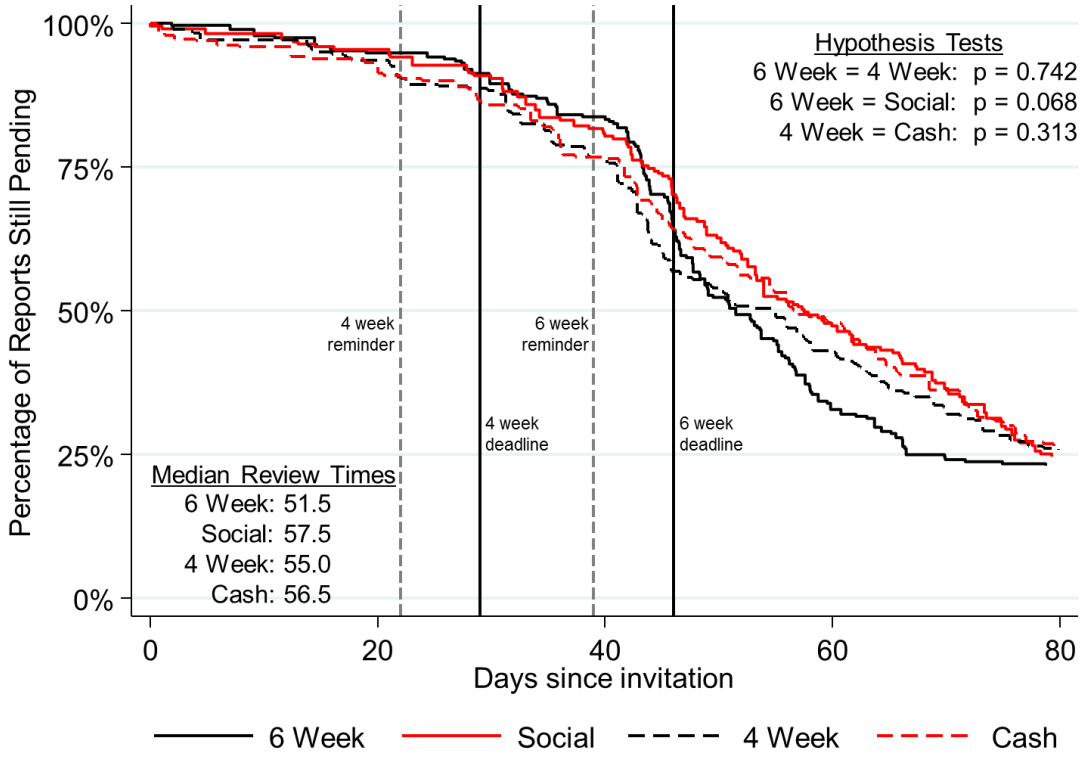
Appendix Table 5

Spillover Effects on Other Journals

Group:	6 Week (1)	Social (2)	4 Week (3)	Cash (4)
A. Reviewer Invitation Acceptance Rate at Other Journals				
Percent accepting invitation	62.1%	58.8%	60.6%	61.8%
	(2.31)	(2.55)	(2.38)	(2.26)
p-value for equality with control		0.344	0.654	0.702
Observations	999	806	969	993
B. Review Times at Other Journals				
Median review time (days)	56.2	54.0	56.5	57.0
	(1.39)	(1.72)	(1.81)	(1.71)
p-value for equality with control		0.562	0.596	0.894
Observations	620	474	587	614

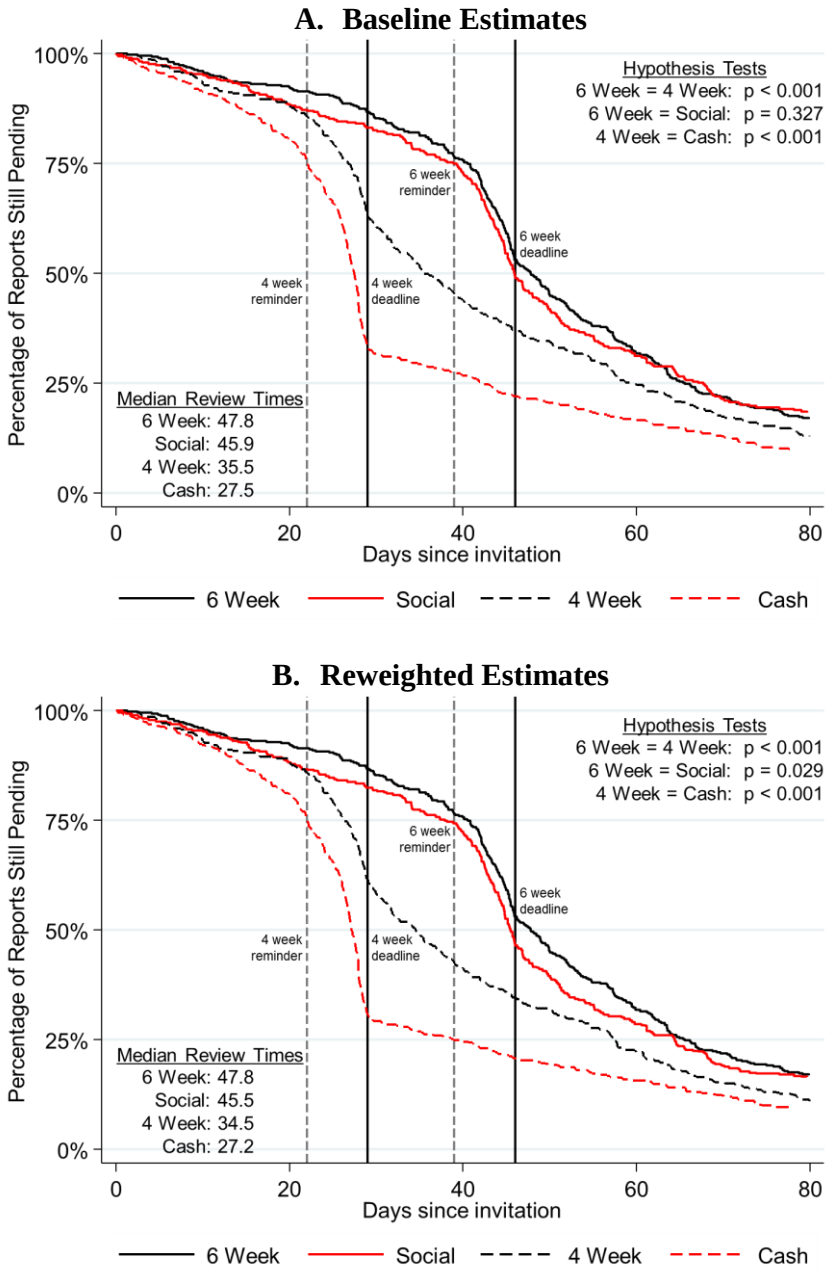
Notes: This table reports estimates of spillover effects of the treatments on referee behavior at other Elsevier journals during the experimental period. The sample includes all referees who accepted invitations to review papers for the Journal of Public Economics between February 15, 2010 and May 9, 2011 (the period when the cash reward was offered). We use data from other Elsevier journals in related fields (see Appendix F for a list) in this table, restricting attention to reviewer invitations received after the first invitation during the experimental period at the Journal of Public Economics and before December 31, 2011. In Panel A, the outcome is the percentage of referees who accept invitations to review papers at other journals. We report standard errors, clustered by referee, in parentheses. We also report p-values for the null hypothesis that the percentages are the same in each treatment group and its corresponding control group. For the social and 4 week groups, the control group is defined as the 6 week deadline group. For the cash incentive group, the control group is defined as the 4 week deadline group, which is the relevant comparison because the cash incentive group also faced a 4 week deadline. There is one observation for each review invitation that referees received from other Elsevier journals. In Panel B, the outcome is the median number of days taken to submit a review conditional on accepting the invitation to referee. Standard errors are reported in parentheses and the p-values are for hypothesis tests analogous to those in Panel A. The number of observations is the number of referee reports submitted to the other journals.

Figure 1: Pre-Experiment Review Times for Referees who Accept Invitations During Experiment



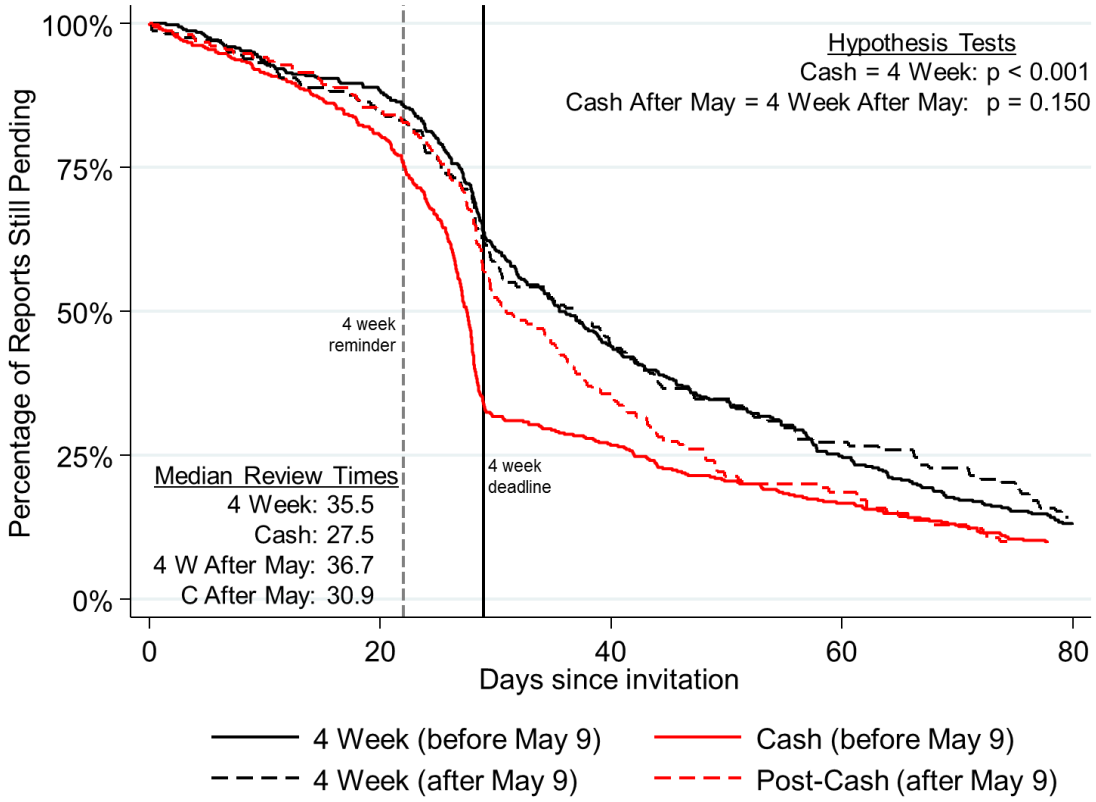
Notes: This figure plots survival curves that show the distribution of *pre-experiment* review times by treatment group. The sample consists of referees who accepted invitations between February 15, 2010 and May 9, 2011, the period when the cash reward was offered. Among these referees, 67.3 percent accepted a review invitation before the experiment began (from November 2005 to February 15, 2010); we use their data to construct this figure. For referees who reviewed multiple papers, we use the most recent pre-experimental review. Each survival curve plots the percentage of reports still pending vs. the number of days elapsed since the referee received the invitation. The solid vertical lines depict the six week deadline (45 days) and the four week deadline (28 days) that were used during the experiment. The dashed vertical lines depict the reminders sent one week before each deadline. Before the experiment, all referees faced the six week deadline and reminders were not sent systematically. We report median review times, defined as the point at which the fraction of reports pending is 50 percent, for each group. We also report p-values from non-parametric Wilcoxon tests for the hypothesis that the pre-experiment review times are the same in each treatment group and its corresponding control group. We compare the four-week and social incentive groups to the six-week group. We compare the cash group to the four-week group because the cash group also faced a four week deadline. We truncate the x-axis at 80 days in the figure for scaling purposes, but use all available data for the hypothesis tests.

Figure 2: Review Times by Treatment Group During Experiment



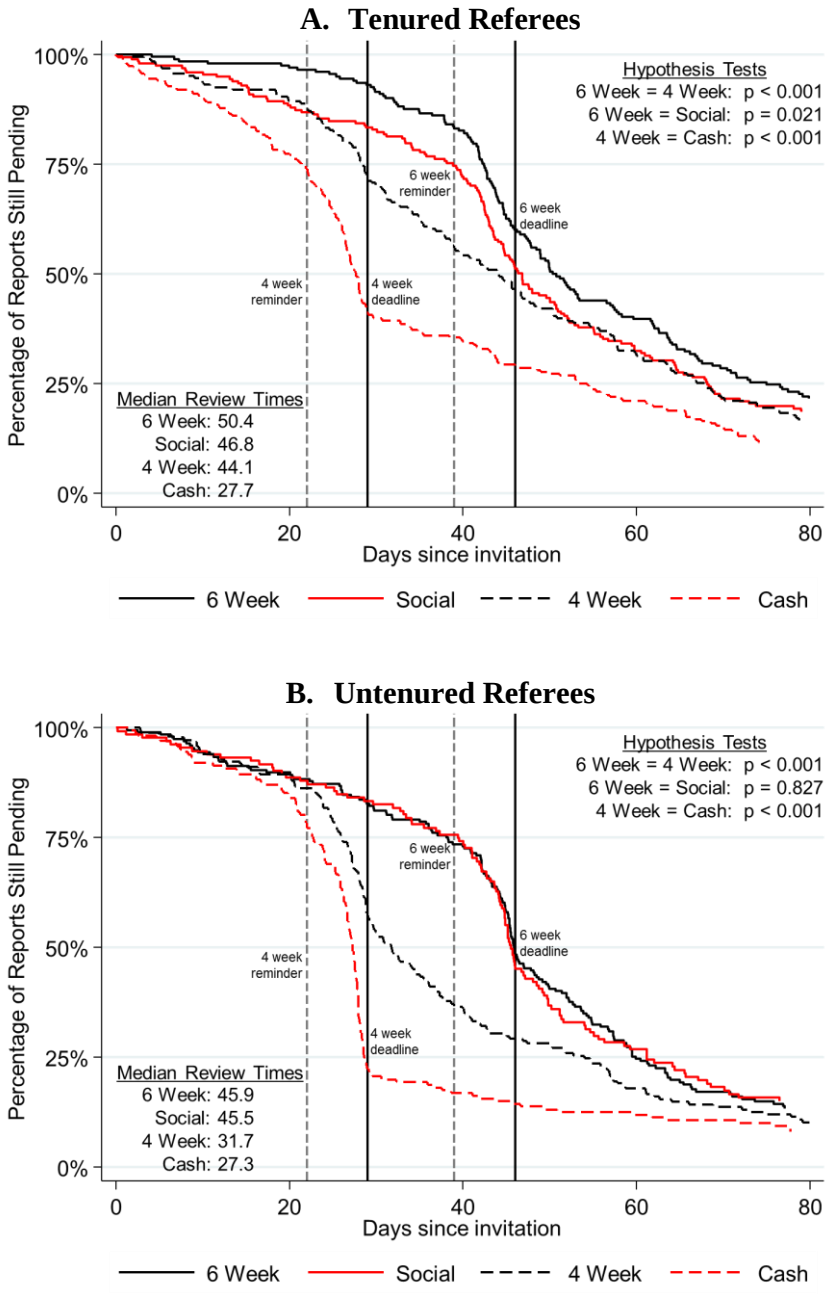
Notes: This figure plots survival curves showing the distribution of review times by treatment group during the primary experimental period, February 15, 2010 to May 9, 2011 (when the cash reward was offered). In Panel A, each survival curve plots the percentage of reports still pending vs. the number of days elapsed since the referee received the invitation. Panel B replicates Panel A, reweighting the observations in the three treatment groups to match the distribution of pre-experiment review times in the six-week group (see Appendix D for details). The solid vertical lines depict the six week deadline (45 days) and the four week deadline (28 days). The dashed vertical lines depict the reminders sent one week before each deadline. We report median review times, defined as the point at which the fraction of reports pending is 50 percent, for each group. We also report p-values from non-parametric Wilcoxon tests for the hypothesis that review times are the same in each treatment group and its corresponding control group. We compare the four-week and social groups to the six-week group. We compare the cash group to the four-week group because the cash group also faced a four week deadline. We truncate the x-axis at 80 days in the figures, but use all available data for the hypothesis tests.

Figure 3: Review Times Before vs. After End of Cash Reward



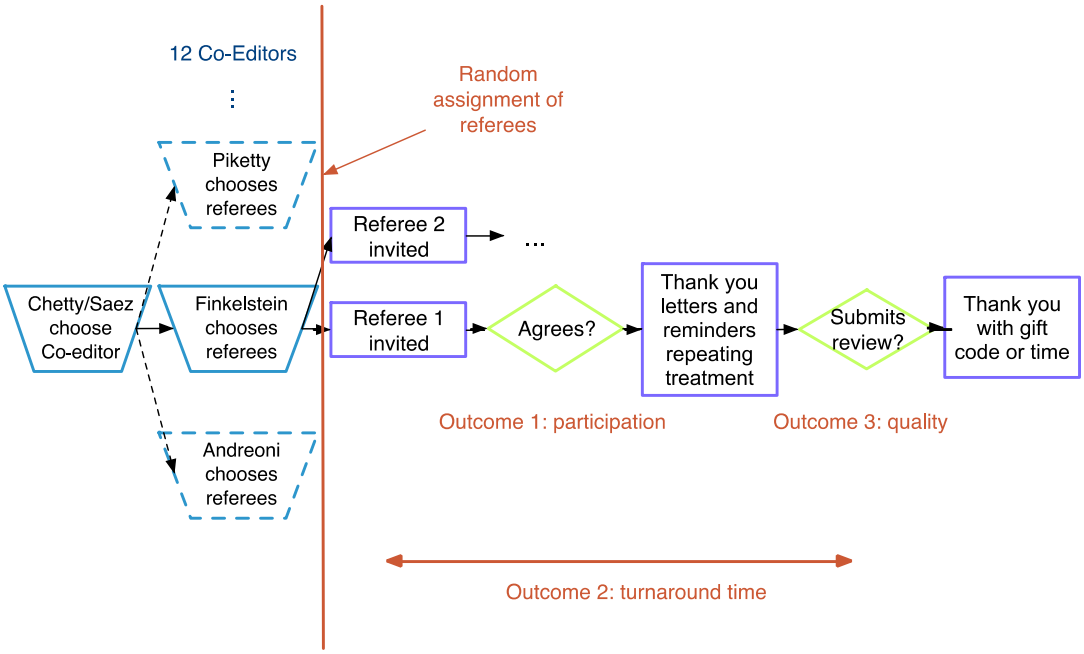
Notes: This figure plots survival curves showing the distribution of review times in the four-week and cash treatment groups before vs. after May 9, 2011. On May 9, cash rewards were stopped for those in the cash treatment group and referees in this group were subsequently treated identically to those in the four-week group. Hence, the cash (after May 9) group includes referees who previously received cash rewards but no longer do, while the cash (before May 9) group includes referees receiving cash incentives. The four-week group faced the same treatment both before and after May 9. Each survival curve plots the percentage of reports still pending vs. the number of days elapsed since the referee received the invitation. The solid vertical line depicts the four week deadline (28 days). The dashed vertical line depicts the reminder sent one week before the deadline. We report median review times, defined as the point at which the fraction of reports pending is 50 percent, for each group. We also report p-values from non-parametric Wilcoxon tests for the hypothesis that review times are the same in the cash and four week groups before and after May 9.

Figure 4: Heterogeneity in Treatment Effects by Tenure Status



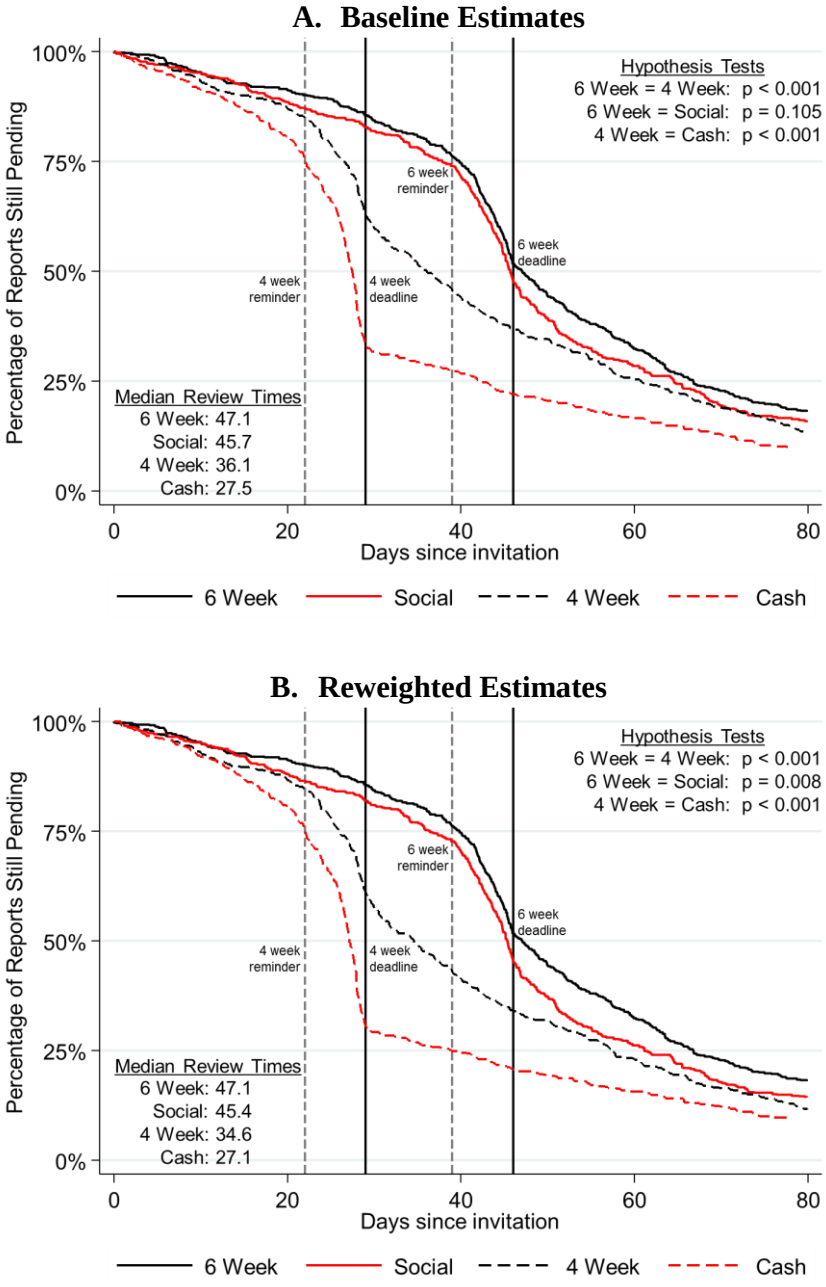
Notes: This figure replicates Figure 2a, splitting the sample between tenured referees (Panel A) and untenured referees (Panel B). Tenure status is measured during the experiment based on information from CVs posted online (see Appendix C for details); referees whose tenure status could not be identified are excluded from this figure. In both panels, the sample consists of referees who accepted invitations between February 15, 2010 and May 9, 2011, the period when the cash reward was offered. Each survival curve plots the percentage of reports still pending vs. the number of days elapsed since the referee received the invitation. See notes to Figure 2 for further details.

Appendix Figure 1: Timeline of Interventions and Outcomes



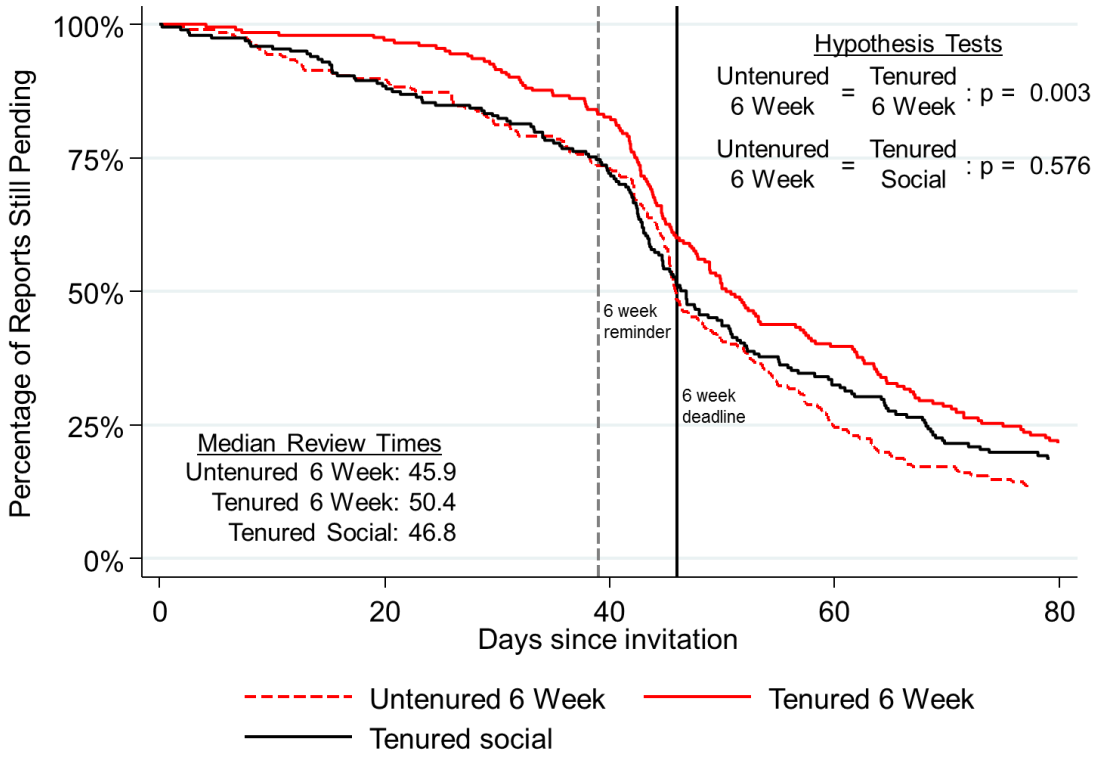
Notes: The figure depicts the timeline of the refereeing process during the experiment. Once a submission is received, editors assign a co-editor in charge who then chooses referees. Invited referees are randomly assigned to one of the four groups (six-week, four-week, cash, social) and receive an email invitation tailored to their group (shown in Appendix A). Referees accept or decline the invitation, which is the first outcome we study. If they accept, we send group-specific reminders one week before the deadline (shown in Appendix B). We then measure the time taken to submit a review, the second outcome we study. If a review is submitted, we send a thank you letter with the cash reward (to eligible referees) and measure the quality of the report, the third outcome we study.

Appendix Figure 2: Review Times by Treatment Group in Extended Sample



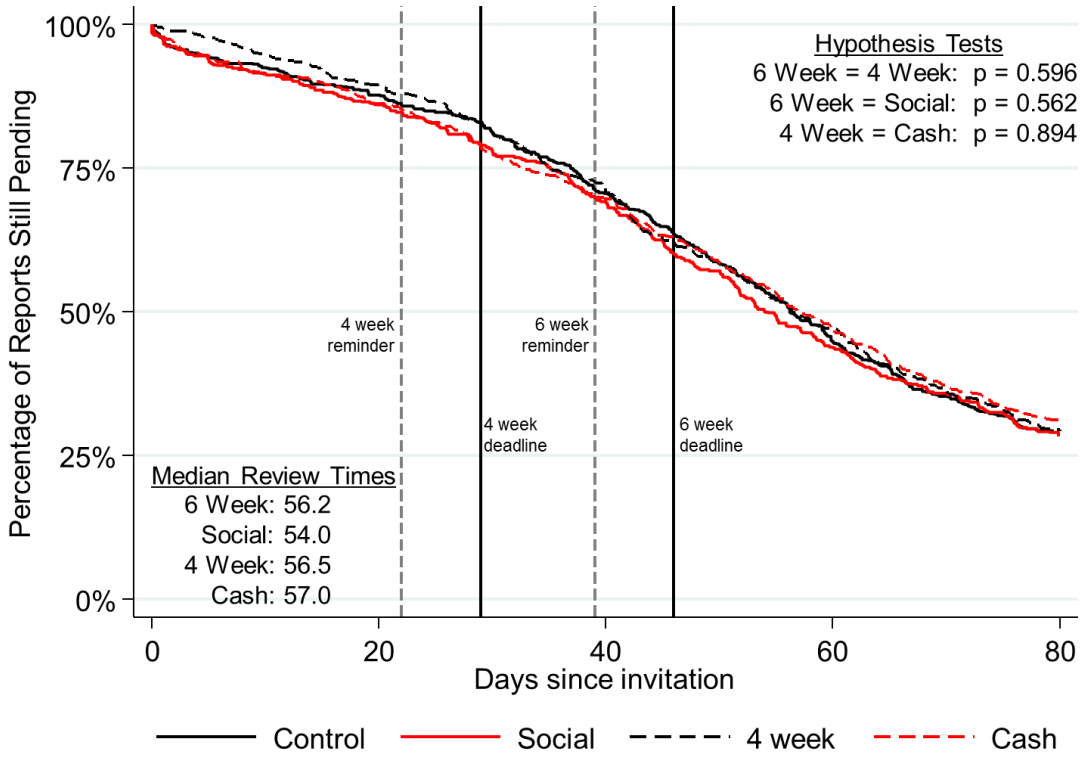
Notes: This figure replicates Figure 2 using the full experimental period from February 15, 2010 to October 26, 2011, including the period after May 9, when the cash reward was stopped. The cash group in this figure still includes only referee invitations up to May 9, 2011. The other groups include all invitations during the full experiment. See notes to Figure 2 for details on the construction of this figure and Appendix Table 4 for Cox hazard model estimates corresponding to these survival curves.

Appendix Figure 3: Social Incentives and Tenured vs. Untenured Referees



Notes: This figure plots a subset of the survival curves reported in Figure 4 on a single figure to show that tenured referees have longer turnaround times than untenured referees in the control group, but behave like untenured referees when facing social pressure. We replicate the series in Figure 4 for (a) untenured referees in the six-week group, (b) tenured referees in the six-week group, and (c) tenured referees in the social group. The solid vertical line depicts the six week deadline relevant for these groups. The dashed vertical line depicts the deadline reminders sent one week before this deadline. We report median review times, defined as the point at which the fraction of reports pending is 50 percent, for each group. We also report p-values from non-parametric Wilcoxon tests for the hypothesis that review times are the same in the untenured six-week group and the two tenured groups. We truncate the x-axis at 80 days in the figure for scaling purposes, but use all available data for the hypothesis tests.

Appendix Figure 4: Spillover Effects: Review Times at Other Journals



Notes: This figure shows the effects of our experimental interventions at the *Journal of Public Economics* on referees' review times at other Elsevier journals (listed in Appendix F). The sample includes all referees who accept a refereeing invitation at another Elsevier journal (before December 31, 2011) *after* receiving an invitation to referee at the *Journal of Public Economics* during our primary experimental period, February 15, 2010 to May 9, 2011. Each survival curve plots the percentage of reports still pending vs. the number of days elapsed since the referee received the invitation from the other journal. As a reference, the solid vertical lines depict the six week deadline (45 days) and the four week deadline (28 days) used at the *Journal of Public Economics* during the experiment. The dashed vertical lines depict the reminders sent one week before each deadline. Other journals have different deadlines and reminder policies. We report median review times, defined as the point at which the fraction of reports pending is 50 percent, for each group. We also report p-values from non-parametric Wilcoxon tests for the hypothesis that review times at other journals are the same in each treatment group and its corresponding control group. We compare the four-week and social incentive groups to the six week group. We compare the cash group to the four-week group because the cash group also faced a four week deadline. We truncate the x-axis at 80 days in the figure for scaling purposes, but use all available data for the hypothesis tests.