

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

Raj Chetty, Emmanuel Saez, and László Sándor

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 six weeks to four weeks reduces median review times from 48 days to 36 days. Because missing the deadline has no direct

■ *Raj Chetty is Professor of Economics, Harvard University, Cambridge, Massachusetts. Emmanuel Saez is Professor of Economics at the University of California, Berkeley, California. László Sándor is a PhD candidate in Economics, Harvard University, Cambridge, Massachusetts. Their email addresses are chetty@fas.harvard.edu, saez@econ.berkeley.edu, and sandor@fas.harvard.edu.*

[†]To access the Appendix, data Appendix, and disclosure statements, visit <http://dx.doi.org/10.1257/jep.28.3.169>

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 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 multitasking (for example, Holmstrom and Milgrom 1991) predict that referees will prioritize the incentivized task (in this application, submitting a report quickly) at the expense of other aspects of performance (in this case, potentially 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 the recommendation of the referee or the length of referee reports. The cash and social incentives induce referees to 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 forthcoming). 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 2006). In our application, if referees submit reviews to be recognized by editors for their service to the profession, 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.

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.

Table 1
Description of Treatment Groups

	<i>Group</i>			
	<i>6 Week</i>	<i>Social</i>	<i>4 Week</i>	<i>Cash</i>
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 one week before the deadline. Examples of these invitation and reminder emails are shown in online 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 four-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.

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. These assignments were permanent for the duration of the experiment: referees never switched groups. The coeditors 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

²An online Appendix available with this paper at <http://e-jep.org> includes the details of the experiment. Appendix Figure 1 presents a flow chart for the entire experiment. Appendix A shows our invitation emails. Online 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 the *JEP* website: <http://e-jep.org>.

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 incentive and cash incentive groups, these emails included language reminding referees of the treatments they faced. We also sent overdue reminders five 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 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 also 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 predetermined characteristics (details in online 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

Fraction of Referees Who Accept Review Invitation by Treatment Group

	<i>Group:</i>			
	<i>6 Week</i>	<i>Social</i>	<i>4 Week</i>	<i>Cash</i>
Percent who accept invitation	67.6% (2.14)	61.1% (2.43)	64.1% (2.23)	72.0% (2.17)
<i>p</i> -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 four-week 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. The number of observations (referee report invitations) is reported in the last row.

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, our results on acceptance rates 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 predetermined 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 online Appendix Table 2b).

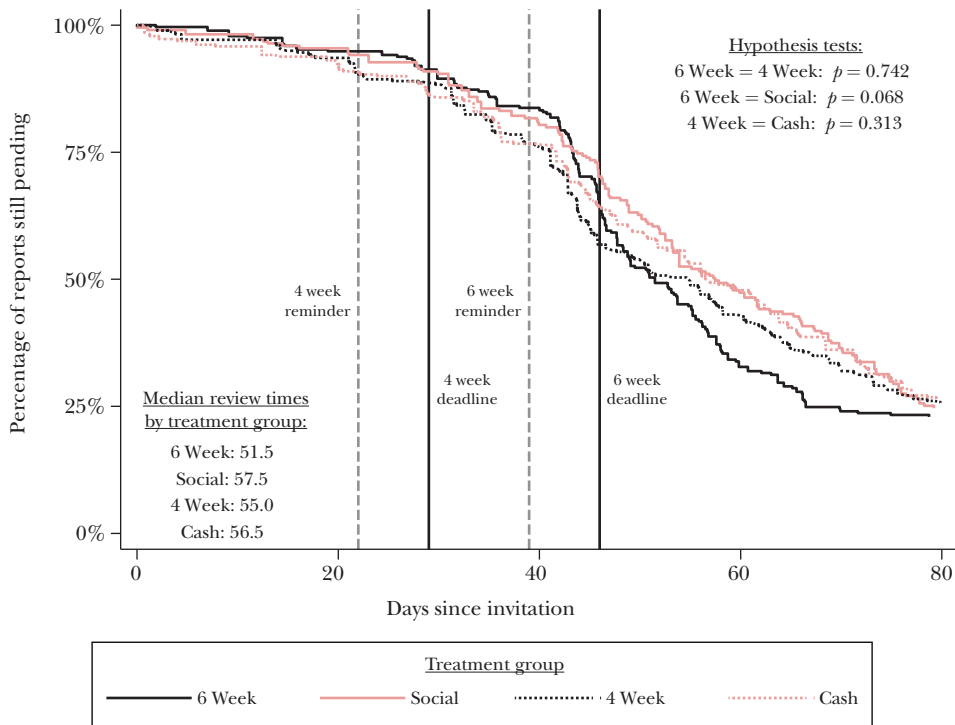
Second, we compare the *pre-experiment* review times of referees who agreed to review papers in each of the four experimental groups. For this analysis, we focus on the 67 percent of referees in our primary experimental sample who reviewed a manuscript for the journal before the experiment began (from 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. Nonparametric (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

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.

Figure 1

Pre-Experiment Review Times for Referees Who Accept Invitations during Experiment, by Treatment Group



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 versus the number of days elapsed since the referee received the invitation. Before the experiment, all referees faced the six-week deadline and reminders were not sent systematically. 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 during the experiment. 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 nonparametric 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.

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 online Appendix D. We report median survival times (the point at which 50 percent of reports have been submitted) and nonparametric (Wilcoxon) tests for the equality of the survival curves in each figure (see online 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 (Figure 2B titled “Reweighted Estimates”), 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.

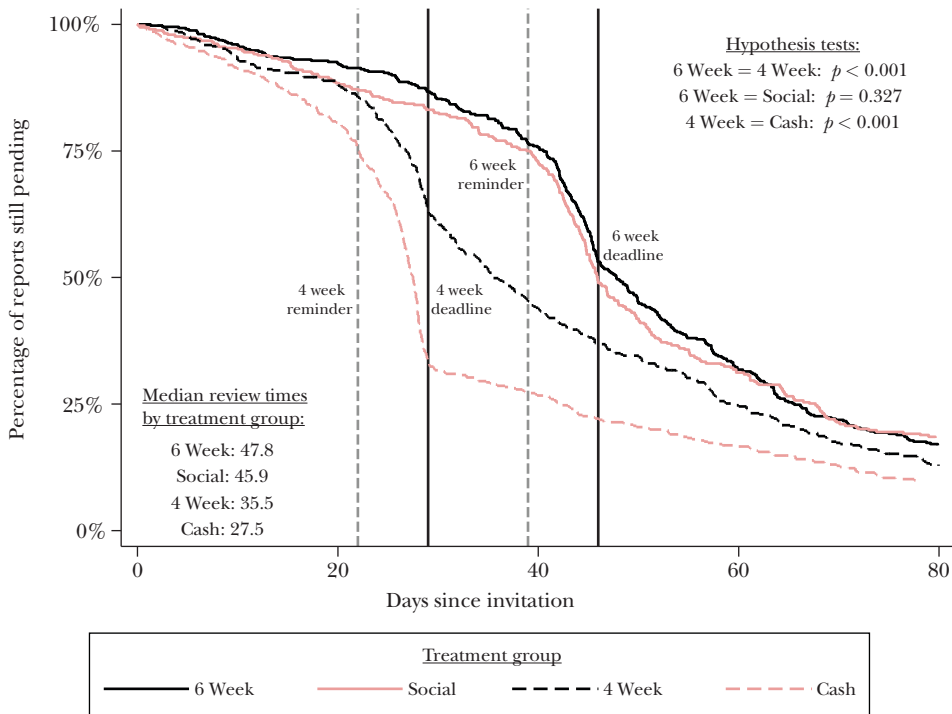
Shortening the deadline from six weeks (45 days) to four weeks (28 days) reduces median review times by 12.3 days, based on the baseline estimates Figure 2A. Hence, we estimate that shortening the deadline by one day reduces median review times by $12.3 / (45 - 28) = 0.72$ days. The effect is so large 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 nonlinear responses. Indeed, the response to the cash treatment is so large that one can show using a nonparametric bounding approach, as in Lee (2009), that selection effects account for very little of the impact. Recall from Table 2 that

Figure 2
Review Times by Treatment Group during Experiment

A: Baseline Estimates



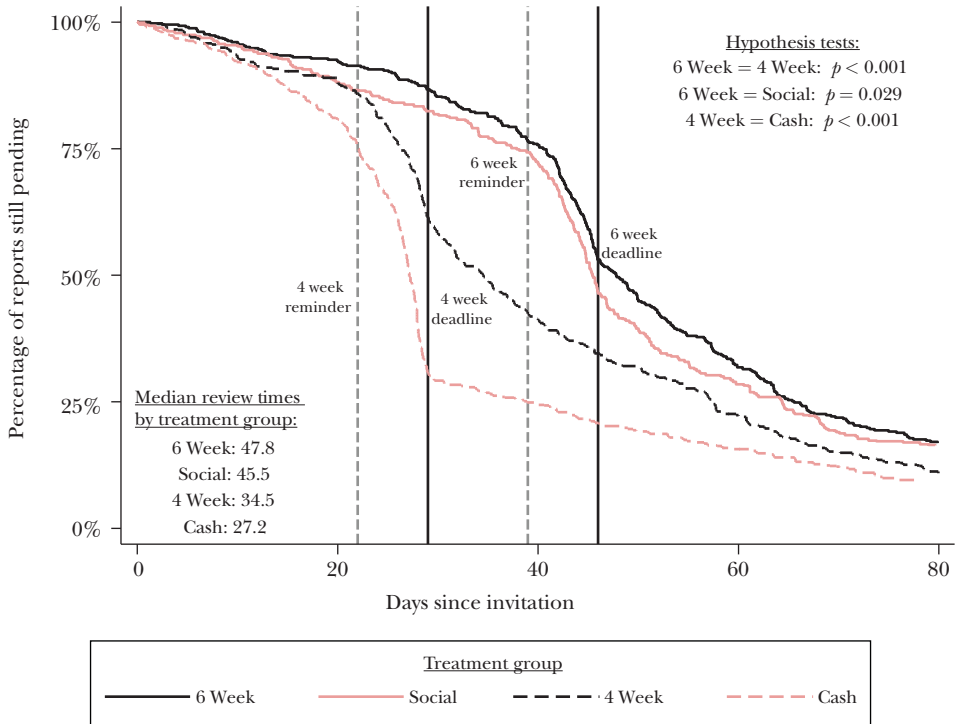
(continued on next page)

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,

Figure 2 (continued)

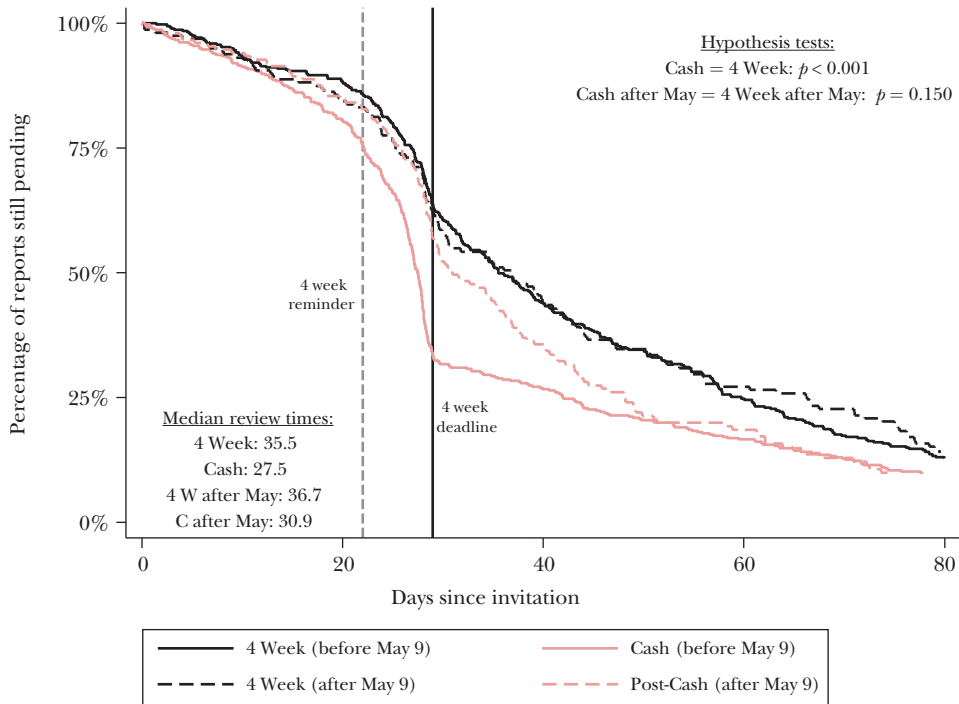
B: Reweighted Estimates



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 versus 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 online 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 nonparametric 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.

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, Koestner, and Ryan 1999; Frey and Jegen 2001; Kamenica 2012). Our experiment offers a new test of this hypothesis in the field, which complements earlier work on economic incentives and prosocial behavior

Figure 3

Review Times Before versus After the 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 versus 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 versus 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 nonparametric Wilcoxon tests for the hypothesis that review times are the same in the cash and four-week groups before and after May 9.

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 versus

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 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 (where turnaround times were to be publicly posted). We find a statistically 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

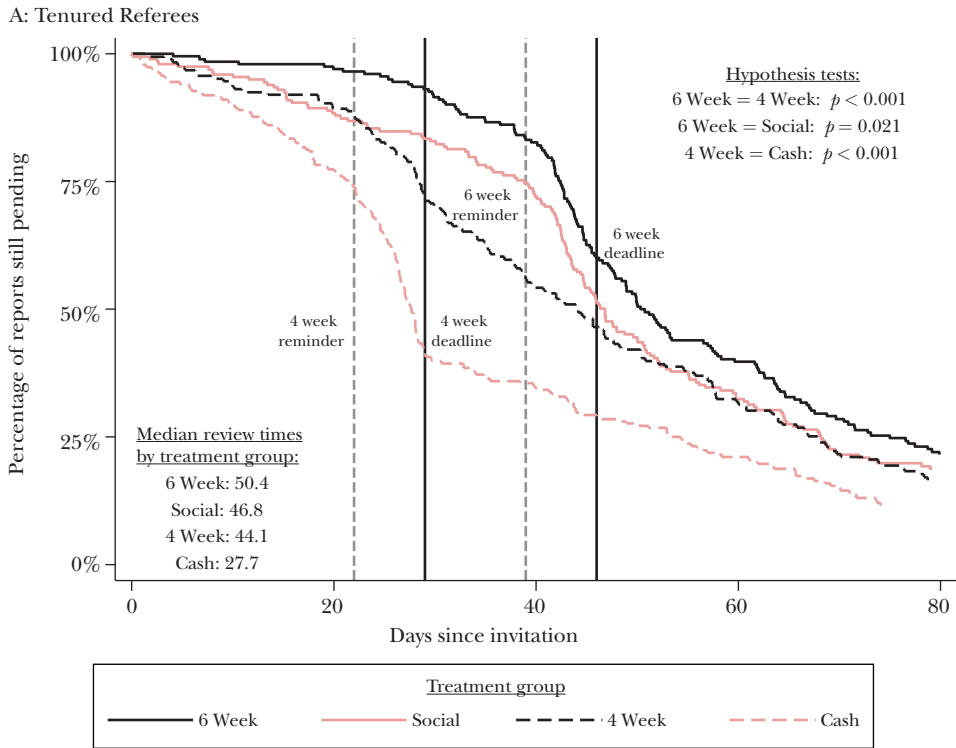
Finally, we explore the heterogeneity of the treatment effects by referee characteristics. We find no significant heterogeneity in treatment effects by several of

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-cash” survival curve in Figure 3 to drop steeply in the week before the four-week deadline. This does not occur: the “Post-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 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 semiparametric Cox hazard models in online Appendix E. Consistent with the graphical evidence in Figure 2, we find that the cash incentive and four-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 online Appendix Table 4, are robust to changes in the control vector and sample specifications. In online 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 as expected, we obtain more precise estimates when using all the data.

Figure 4
Heterogeneity in Treatment Effects by Tenure Status

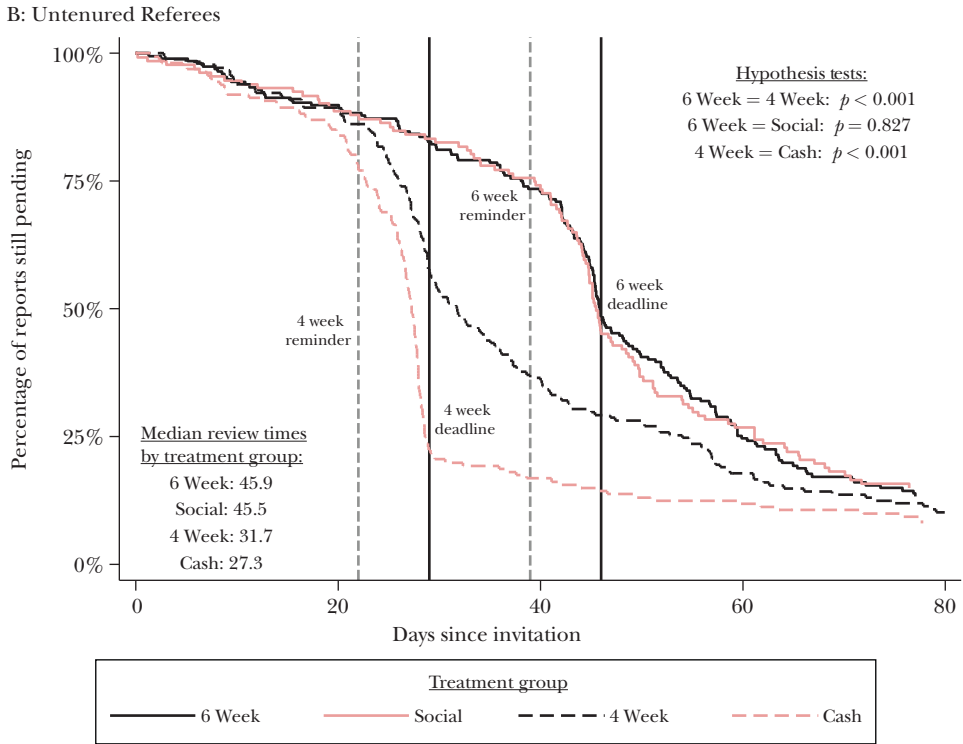


(continued on next page)

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 (Figure 4A) and those who were not tenured at that time (Figure 4B). 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 as 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 *larger* effects on

Figure 4 (continued)



Notes: Figure 4 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 online 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 versus the number of days elapsed since the referee received the invitation. See notes to Figure 2 for further details.

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 coeditors, 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

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 online Appendix Figure 3.

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 multitasking 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 because 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 shorter (11 percent shorter) in the social and cash groups relative to the six-week 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 online Appendix F for a complete list). We analyze referee invitations from other journals that are received 1) *after* referees have received an

Table 3
Measures of Review Quality by Treatment Group

	Group:			
	6 Week	Social	4 Week	Cash
Agreement between editor decision and referee's recommendation				
Editor follows referee's recommendation	77.9% (2.00)	76.2% (2.34)	77.5% (2.20)	76.2% (2.15)
<i>p</i> -value for equality with control		0.585	0.884	0.686
Observations	403	324	373	404
Length of referee report				
Median number of words in referee report	877 (29.1)	757 (32.5)	864 (30.3)	786 (29.2)
<i>p</i> -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 versus 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 four-week 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. 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.

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 percent in all four groups, with no statistically significant differences across the groups (see online Appendix Table 5). Median review times are approximately 56 days in all four groups, again with no statistically significant differences across the groups (see online 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

⁸The 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.

postponed. If referees postpone activities with pure private benefits such as leisure, social welfare may increase because 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 improve the efficiency of the peer review process substantially with relatively low-cost interventions, demonstrating the value of studying the peer review process empirically (as in Card and DellaVigna, this issue). Our results reject the view that the review process in

⁹ These findings contrast with the results of Squazzoni, Bravo, and Takács (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.

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.

■ *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 found at the JEP website, with the paper.*

References

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

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96(5): 1652–78.

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): 627–68.

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): 947–93.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* 425(6960): 785–91.

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 Christopher 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(Special Issue): 24–52.

Kamenica, Emir. 2012. "Behavioral Economics and Psychology of Incentives." *Annual Review of Economics* 4: 427–52.

Lacetera, Nicola, Mario Macis, and Robert Slonim. 2013. "Economic Rewards to Motivate Blood Donations." *Science* 340(6135): 927–28.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth, 111–94. Princeton University Press.

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." Unpublished paper.

Squazzoni, Flaminio, Giangiuseppe Bravo, Károly Takács. 2013. "Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study." *Research Policy* 42(1): 287–94.

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. Forthcoming. "Voluntary Giving to Public Goods: Moving Beyond the Linear VCM." In *Handbook of Experimental Economics*, Vol. 2, edited by John H. Kagel and Alvin E. Roth. Princeton University Press.