aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
1–5

DOI: 10.1177/1745691617693625
www.psychologicalscience.org/PPS
SAGE

Reflections on the Time-Pressure Cooperation Registered Replication Report

David G. Rand
Department of Psychology, Department of Economics, School of Management, Yale University

I am honored to have had Rand, Greene, and Nowak (2012) Study 7 chosen as the subject of a Registered Replication Report (RRR) and would like to thank everyone involved for the time and effort that they have invested in exploring the relationship between intuition and cooperation. I was glad to see that the RRR replicated the results reported in our original article. Beyond replicability, however, the lack of effect when including noncompliant participants in the RRR does raise important questions about the existence of a causal effect of time pressure on cooperation: The results are ambiguous on this point, as they are consistent with both a selection effect (no causal effect) and a meaningful positive causal effect.

Here, I present two analyses that, although not conclusive, provide support for the existence of a causal effect in the RRR data: One showing lack of selection bias based on observed individual difference covariates, and another based on the pattern of cross-lab variation in the impact of excluding noncompliant participants. Taking the RRR results (and the analyses I present here) together with the findings of a recent meta-analysis with over 15,000 observations and no indications of publication bias (Rand, 2016), I believe that the overall body of empirical evidence supports the general conclusion of the original 2012 article (and subsequent theorizing based on the social heuristics hypothesis, e.g., Bear & Rand, 2016; Rand et al., 2014) that deliberation undermines cooperation in one-shot anonymous interactions while also showing the limitations of time pressure as an approach for investigating this question. I hope the results of this RRR, and the questions that they raise, will inspire future work to more fully illuminate the impact of time pressure—and intuition more broadly—on human cooperative behavior.

Successful Direct Replication of Our Original Finding

The main analysis in our article excluded decisions by noncompliant participants (those who took longer than 10 s in the time-pressure condition or less than 10 s in the time-delay condition), which is in line with common practice for time-pressure experiments in both psychology and economics (e.g. Cappelletti, Güth, & Ploner, 2011; Kocher, Pahlke, & Trautmann, 2011; Shalvi, Eldar, & Bereby-Meyer, 2012; Sutter, Kocher, & Straub, 2003; Young, Goodie, Hall, & Wu, 2012; as well as the papers whose null results partly motivated this RRR, Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014).

When analyzed using this same exclusion criterion, the RRR data yielded a similar result to the original article: Both found substantial (and significant) positive effects of time pressure on cooperation—effects that did not significantly differ from each other ($z = .73$, $p = .46$). Thus, the RRR results successfully replicated our original published finding.

Intent-to-Treat Results Are Consistent With Either Selection Bias or a Causal Effect of Time Pressure

The RRR's primary planned analysis was not the one conducted in the original article. The alternative approach employed by the RRR was an "intent-to-treat" (ITT) analysis that compared all participants assigned to the two conditions, regardless of whether they complied with the timing manipulation. ITT analyses allow clear causal inference by avoiding selection biases. However, they also necessarily underestimate the actual effectiveness of treatments that do have a causal effect (because the ITT analysis includes untreated participants). This is especially true when a large fraction of participants were not successfully treated, as in the RRR, in which only 34.1% of participants in the time-pressure condition complied

Corresponding Author:
David G. Rand, Department of Psychology, Yale University, Sheffield-Sterling-Strathcona Hall, Room 204, 1 Prospect Street, New Haven CT 06511
E-mail: david.rand@yale.edu

with the manipulation. Although the RRR's ITT analysis found essentially zero effect of time pressure on cooperation, an ITT analysis combining data from the two time-pressure studies in the original article (but not reported in the original article) indicates a positive effect of 6.0 percentage points with a 95% CI of [1.1, 10.9].

Critically, although the RRR data do not provide evidence of a causal effect, they also do not rule out such an effect: given the low compliance rate, the 95% CI for the RRR's ITT estimate includes a true causal effect of up to 5.45 percentage points.[1] There may therefore be a true effect in the population, but the ITT effect in the RRR sample underestimated it (i.e. the ITT null result is not precisely estimated because of low compliance): The difference between the RRR result and the original might be due to sampling error coupled with a high rate of noncompliance.

Another way that the RRR data could be consistent with a positive causal effect of time pressure on cooperation was suggested in the RRR discussion: Noncompliant subjects in the RRR could have actually received a negative treatment (rather than simply being untreated). In that case, the negative effect for noncompliant participants would offset the positive effect of the treatment for compliant participants, resulting in no ITT effect overall despite a true positive treatment effect reflected in the compliant-only analysis.

A possible source of such negative treatment comes from a difference in design between the original studies and the RRR, namely the RRR's introduction (at my suggestion) of timers on the decision screen. Although I suggested the timers in the hope of increasing compliance relative to the original—a hope that was clearly not fulfilled given that the RRR compliance rate was roughly half that of the original study—in retrospect, the timers may have introduced a confound by making it clear to the noncompliant participants that they had failed to respond in time while still allowing them to subsequently make (or change) their choice.

This may have made those participants (a) think their choice did not matter since the time had expired, leading to random play;[2] (b) confused or surprised because the page did not auto-advance and still allowed them to change their choice, thus triggering the deliberation that time pressure was supposed to suppress; and/or (c) feel annoyed or frustrated with the experimenter for demanding that they respond more quickly than they were able. Any of these effects could have led noncompliant participants in the time-pressure condition not just to be untreated, but in fact to have received a cooperation-reducing treatment.[3] I recognize that these concerns regarding the timers are post-hoc, and more work is needed to compare the effect on cooperation of time pressure implemented with and without timers.

RRR Fits Well With the Meta-Analysis of Rand (2016)

The RRR's observations regarding time pressure are in line with a recent meta-analysis (Rand, 2016) of various cognitive processing manipulations (which, importantly, showed no signs of publication bias using Egger's test for small-study effects or *p* curve). This meta-analysis found an overall strong positive effect of promoting intuition on cooperation in social dilemmas in which defection is strictly payoff-maximizing (as in the RRR and Rand et al., 2012) even when conducting an ITT analysis—thus supporting the conclusions of the original article (Rand et al., 2012). Regarding time-pressure manipulations specifically, however, the meta-analysis (like the RRR) only found a positive effect when excluding noncompliant participants (see Fig. 1).[4] This was in contrast to the other manipulations, which had much higher compliance rates and continued to show a strong positive effect of intuition when including noncompliant participants in an ITT analysis.[5] (Note that this observation regarding the limitations of time pressure based on my meta-analysis was preregistered.)

What to conclude about the effect of time pressure based on the existing data, therefore, depends on the consequences of excluding noncompliant participants. Does this exclusion provide a better estimate of the true effect of time pressure by focusing the analysis only on participants who were effectively manipulated? Or does excluding noncompliant participants create the false appearance of a causal time-pressure effect via bias induced by selection?

I now present two analyses (designed and implemented after seeing the results of the RRR) that support the exclusion of noncompliant participants and speak against the existence of selection bias. Details of these analyses, as well as all code required to reproduce them, are presented in the Supplemental Material (SM): https://osf.io/xyfqn/. The relevant SM section is indicated alongside each analysis.

Evidence Against Selection Bias Based on Observed Individual Differences

One non-causal explanation for the treatment difference in the compliant-only analysis is selection bias based on individual differences: If people who complied differed from those who did not in attributes that influence cooperativeness, excluding noncompliant participants could artificially create a difference in cooperation across conditions. Based on this logic, an individual difference could potentially lead to selection bias if it is associated with both compliance and cooperation.

Accordingly, it is possible to test for selection bias based on the substantial number of individual difference

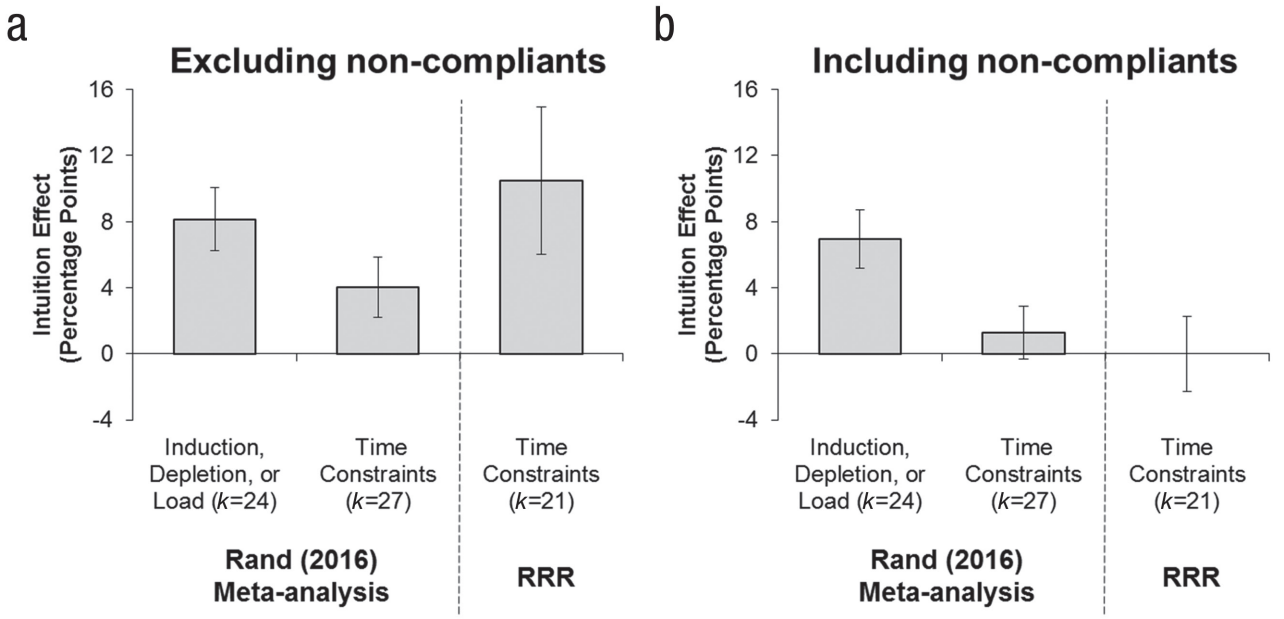

Fig. 1. Effect of increasing use of intuition (relative to deliberation) on social dilemma cooperation in experiments meta-analyzed in Rand (2016) and the RRR. Error bars indicate 95% confidence intervals.

measures that were collected in the RRR, many of which are theoretically linked to cooperation and thus are potential candidates for selection bias (e.g. trust, individualism vs. collectivism, prior experience with economic games, knowledge of the experimental hypothesis, comprehension of the payoff structure). Two of the collected measures did have robust (albeit small, $r \sim .1$) correlations with compliance, but neither was associated with increased cooperation: Compliant participants were more likely to be male, but in fact men cooperated somewhat less than women, and compliant participants spent less time reading the instructions, which was unrelated to cooperation. See SM1.1 for details.

More generally, cooperation was related to many of the individual differences collected in the RRR (as expected, given that these measures were chosen in large part because of their hypothesized relationship with cooperation): Regression models using all the individual difference measures to predict cooperation in the time-pressure condition for each lab (allowing arbitrary nonlinear functional forms for most measures) had an average $r^2 = .94$. If selection on any of these measures was responsible for the time-pressure effect observed in the compliant-only analysis, then including them as covariates should reduce that effect. Yet this is not the case: Predicting cooperation among compliant participants while including all individual difference measures as covariates yielded a meta-analytic time-pressure effect estimate of 11.8 percentage points, 95% CI [5.0, 18.6], which is actually slightly larger than the estimate from the analysis without covariates (10.3 percentage points, 95% CI [6.2, 14.4], as reported in the RRR). See SM1.2 for details.

Thus, there is no evidence that selection on any of the collected individual difference measures led to bias that inflated the time-pressure estimate. Of course, these analyses can only speak to selection bias based on the covariates that were collected. It could still be the case that the compliant-only time-pressure effect would be eliminated by covariates not measured in the RRR.

Evidence Against Selection Bias Based on an Underlying Decision Time Correlation

An alternative selection bias account argues that excluding noncompliant participants merely creates the illusion of a causal time-pressure effect because of an underlying negative correlation between decision time and cooperation (Tinghög et al., 2013). By this account, excluding noncompliant participants selectively excludes slow (and therefore selfish) participants from the time-pressure condition and fast (and therefore cooperative) participants from the time-delay condition, leading to higher cooperation in the post-exclusion time-pressure condition. In other words, decision time is a proxy for cooperativeness, and exclusions based on compliance (i.e. decision time) are really selection on cooperativeness.

Evidence against this selection bias account (and in support of a true causal effect) comes from considering the impact of exclusion—that is, how does exclusion

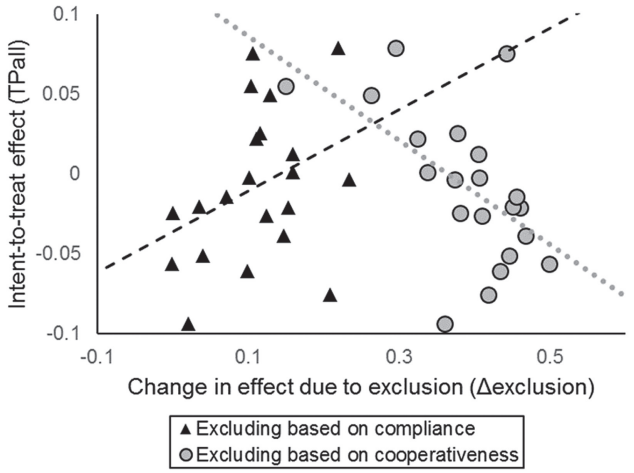

Fig. 2. Relationship between TP_{all} and $\Delta_{exclusion}$ in the RRR data when exclusions are made based on compliance with the time constraints (black) versus cooperativeness (gray). One point per lab. Least-squares fit lines added for visualization purposes.

change the magnitude of the time-pressure effect? Specifically, theory and computer simulations (see SM2 for details) indicate that if there is a true causal effect and no selection bias, a positive relationship should be observed across labs between (a) the amount that the time-pressure effect increases when excluding noncompliant participants ($\Delta_{exclusion}$),[6] and (b) the intent-to-treat time-pressure effect when including all participants (TP_{all}). In this "causal effect with no selection bias" scenario, excluding noncompliant participants gives an accurate measure of the true effect (by only considering people who were successfully treated), whereas the intent-to-treat effect TP_{all} is a diluted representation of the true effect (because of noncompliance). Therefore, there should be a positive correlation between the size of intent-to-treat effect (TP_{all}) and the difference between the compliant-only effect and the intent-to-treat effect ($\Delta_{exclusion}$); because the intent-to-treat effect is just a diluted version of the compliant-only effect, a bigger intent-to-treat effect should be associated with an even bigger compliant-only effect. Put differently, if excluding noncompliant participants helps to refine a true effect by eliminating the many participants who were unsuccessfully treated (as opposed to inducing selection bias), there should be a positive relationship between TP_{all} and $\Delta_{exclusion}$ because the exclusion amplifies the signal produced by the true causal effect (which is present but weaker in the intent-to-treat analysis of TP_{all}).

Indeed, in the RRR data there was a positive relationship between TP_{all} and $\Delta_{exclusion}$ across labs ($r = .36$). This is in stark contrast to what is observed when purposely introducing selection bias instead of excluding based on compliance: Selectively excluding noncooperative participants from the time-pressure condition and cooperative participants from the time delay condition leads to a strong negative correlation across labs between TP_{all} and $\Delta_{exclusion}$ ($r = -.56$). See Figure 2 and SM3 for details and statistical analysis.[7]

These analyses make two points. First, the positive correlation between TP_{all} and $\Delta_{exclusion}$ when excluding noncompliant participants is consistent with a true causal effect of time pressure with no selection bias. Second, the fact that excluding based directly on cooperation (instead of compliance) produces the opposite correlation shows that compliance is not just a proxy for cooperativeness and therefore that excluding based on compliance is not equivalent to introducing selection bias based on cooperativeness—contradicting the deflationary account based on selection bias described at the beginning of the this section.

Conclusion

In sum, the results of this RRR replicate our previously published findings but are ambiguous regarding the existence of a causal effect of time pressure on cooperation because of a very high rate of noncompliance coupled with a null intent-to-treat effect. Nonetheless, the RRR data provide new insight into the functioning of time-pressure experiments by allowing novel analyses that suggest excluding noncompliant participants may indeed help to reveal a true underlying time-pressure effect (rather than introducing selection bias). When considered together, the extant data regarding cognitive process manipulations support the conclusion that intuition promotes cooperation in social dilemmas, consistent with the predictions of the social heuristics hypothesis, while also emphasizing the limitations of time pressure (at least as implemented here) as a method for investigating this issue.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. If the 65.9% of untreated subjects produced a zero effect, then the observed ITT effect size would be only 34.1% of the true effect size. Taking the upper bound of the ITT 95% CI (1.86) and dividing by .341 gives the upper bound estimate of the true effect 95% CI as 5.45 percentage points.
2. Because the average cooperation rate was about 50% in most labs, increasing randomness (i.e. shifting cooperation towards 50%) would lead to a decrease in cooperativeness.
3. It is not possible to evaluate these hypotheses using the RRR data because, for many participants, Qualtrics only recorded time taken to submit the final decision and not the times associated with first and last click pre-submission click (which prevents one from checking if noncompliant participants changed their decision after seeing the time expire).
4. Evidence in support of the existence of a true time-pressure effect, at least in some populations using certain designs, comes from looking only at those time-pressure studies run

by my lab (15 studies, total unique $N = 5{,}831$) As reported in the personal meta-analysis of Rand et al. (2014), this analysis avoided file-drawer effects by including all studies (published or unpublished), and avoided p hacking by using the same simple comparison-of-means across all studies. The data showed a significant positive effect of time pressure both when excluding noncompliant participants and when performing an ITT analysis including noncompliant participants.

5. There continued to be no evidence of publication bias when considering only the non-time-pressure studies; see Rand (2016).

6. $\Delta_{exclusion} = TP_{exclusion} - TP_{all}$, where $TP_{exclusion}$ is the time-pressure effect when excluding noncompliant participants and TP_{all} is the intent-to-treat time-pressure effect when including all participants.

7. Bootstrapped p values: difference between correlations for compliance-based and cooperation-based exclusions, $p = .002$; positive correlation for compliance-based exclusion, $p = .057$; negative correlation for cooperation-based exclusion, $p = .007$. Excluding data from the Srinivasan lab (whose reaction time data likely were corrupted, as RTs were longer in time pressure than forced delay): difference between the two correlations, $p = .0003$; compliance-based exclusion, $r = .40$, $p = .046$; cooperation-based exclusion, $r = -.68$, $p = .002$. Also, note that excluding participants at random from the RRR data leads to no correlation between TP_{all} and $\Delta_{exclusion}$, average $r = .00$. See SM3 for details.

References

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, *113*, 936–941.

Cappelletti, D., Güth, W., & Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology*, *32*, 940–950.

Kocher, M. G., Pahlke, J., & Trautmann, S. T. (2011). *Tempus fugit: Time pressure in risky decisions* (Munich Discussion Paper No. 2011-8). Munich, Germany: University of Munich.

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, *27*, 1192–1206.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Green, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5*, Article 3677. doi:10.1038/ncomms4677

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, *23*, 1264–1270.

Sutter, M., Kocher, M., & Straub, S. (2003). Bargaining under time pressure in an experimental ultimatum game. *Economics Letters*, *81*, 341–347.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., . . . Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *497*, E1–E2.

Verkoeijen, P. P. J. L., & Bouwmeester, S. (2014). Does intuition cause cooperation? *PLoS ONE*, *9*(5), e96654. doi:10.1371/journal.pone.0096654

Young, D. L., Goodie, A. S., Hall, D. B., & Wu, E. (2012). Decision making under time pressure, modeled in a prospect theory framework. *Organizational Behavior and Human Decision Processes*, *118*, 179–188.

Supplemental Material

for

Reflections on the Time-Pressure Cooperation Registered Replication Report

David G. Rand

Contents

SM1. Testing for selection bias using observed individual difference measures

The "selection bias based on individual differences" account argues that the difference in cooperation between the time pressure and time delay conditions in the compliant-only analysis is driven by selection: because many non-compliant participants were excluded from the time pressure condition, there could be systematic differences in participant characteristics across conditions in the compliant-only analysis, and these differences could be responsible for the observed difference in cooperativeness.

For example, imagine that compliant participants were older than non-compliant participants, and that older people are more likely to cooperate. If this was the case, then excluding non-compliant participants from the time pressure condition would cause the average age to be higher in the time pressure condition compared to the forced delay condition, and correspondingly would lead to more cooperation in the time pressure condition than the forced delay condition.

Although it could always be that selection bias is occurring based on unobserved characteristics, it is at least possible to test for selection bias based on those characteristics that *were* observed for each participant in the RRR.

SM1.1 Correlations between individual differences and compliance versus cooperation

I begin by looking for evidence of selection – do compliant and non-compliant participants in the time pressure condition differ significantly on any of the observed characteristics? Table S1 shows the correlation between compliance and each individual difference measure recorded in the RRR. Column 1 indicates correlations partialing out the variance in compliance explained by the lab in which the study was conducted (since the effect sizes for the meta-analysis are calculated within each lab, partialing out lab variance is critical). For completeness, I also report simple pairwise correlations in column 2. When a correlation is significant in column 2 but not in column 1, that indicates that variance in that individual difference measure was merely proxying for variation in compliance across labs, rather than reflecting a true relationship with compliance. For further completeness, column 3 indicates partial correlations from the full model including all covariates as well as lab dummies (these partial correlations are difficult to interpret, as many of the individual difference measures were strongly correlated with each other). Note that in all analyses in this section, I follow the RRR analysis protocol and exclude the 21 participants who indicated "Other" as their gender.

From Table S1, it can be seen that the only variables on which there is robust evidence of significant selection are gender (compliant participants are more likely to be male) and time taken reading the instructions (compliant participants, who respond quickly on the decision screen, also pass more quickly through the preceding instructions screen). All other correlations are both not statistically significant when partialing out variance based on lab, and also small in magnitude (all $r < .10$).

Table S1. *Correlations with compliance, considering all participants in the time pressure condition.*

	Partial correlation including lab dummies	Simple pairwise correlation	Partial correlation from full model
Male	0.11*	0.12*	0.11*
Age	0.01	0.03	0.01
Trust in strangers	0.06	0.08*	0.05
Trust in daily life interaction partners	0.03	0.05	0.00
People can be trusted	0.01	0.03	-0.03
People are fair	0.02	0.03	0.01
People try to help	0.01	0.03	0.01
Can't count on strangers (Reverse coded)	0.03	0.04	-0.01
Prior economic game experience (0=no, 1=yes)	0.02	0.04	0.00
Prior subject pool study experience (0=no, 1=yes)	0.03	0.05	0.02
Prior paid study experience (0=no, 1=yes)	0.01	0.07	-0.02
Prior MTurk pool study experience (0=no, 1=yes)	-0.02	-0.05	-0.02
Prior experience with deception studies (0=no, 1=yes)	0.05	0.04	0.03
Horizontal Collectivism	0.01	0.00	0.02
Horizontal Individualism	0.01	0.03	0.01
Vertical Collectivism	-0.04	-0.05	-0.02
Vertical Individualism	-0.02	0.00	-0.02
Know other participants (0=no, 1=yes)	0.07	0.09*	0.07
log10(Time reading instructions [s])	-0.12*	-0.10*	-0.12*
Heard about study	0.02	0.03	0.02
Knew what was investigated	-0.02	0.01	-0.02
Unsure what researchers trying to demonstrate	-0.01	-0.03	-0.01
Had good idea about hypotheses	0.00	0.02	0.00
Unclear about study aims	0.02	0.00	0.02
Correct comprehension of payoffs (0=No, 1=Yes)	0.06	0.09*	0.06

**Significant at the 5% level (including Bonferonni correction for 25 comparisons)*

The real concern is not selection per se, however, but instead is selection *bias:* do selection-based differences across conditions explain the observed difference in cooperation between conditions in the compliant-only analysis? (For example, compliant participants could be more likely to have blue eyes, but this difference would not lead compliant participants to be more cooperative – and therefore would be irrelevant for explaining the observed difference in cooperation between conditions in the compliant-only analysis.)

As a first test for selection bias, I ask whether there are correlations between cooperation in the time pressure condition and the variables on which selection occurred (see Table S2). If the characteristics that are over-represented in compliant participants are associated with increased cooperation, that could induce selection. On the contrary, however, being male is if anything *negatively* related to cooperation, such that selection based on gender would *reduce* the apparent time pressure effect rather than inflating it; and there is no relationship between time taken reading in the instructions and cooperation.

Table S2. *Correlations with cooperation, considering all participants in the time pressure condition.*

	Partial correlation including lab dummies	Simple pairwise correlation	Partial correlation from full model
Male	-0.06	-0.09*	-0.05
Age	0.03	0.03	0.03
Trust in strangers	0.22*	0.21*	0.10*
Trust in daily life interaction partners	0.15*	0.15*	0.03
People can be trusted	0.13*	0.12*	0.00
People are fair	0.13*	0.14*	0.04
People try to help	.09*	0.09*	0.01
Can't count on strangers (Reverse coded)	.16*	0.17*	0.05
Prior economic game experience (0=no, 1=yes)	-0.01	-0.10*	-0.01
Prior subject pool study experience (0=no, 1=yes)	0.01	-0.05	0.02
Prior paid study experience (0=no, 1=yes)	-0.01	-0.12*	-0.02
Prior MTurk pool study experience (0=no, 1=yes)	0.02	0.03	0.01
Prior experience with deception studies (0=no, 1=yes)	0.01	-0.02	-0.01
Horizontal Collectivism	0.15*	0.14*	0.06
Horizontal Individualism	-0.04	-0.08*	0.00
Vertical Collectivism	0.03	0.03	0.02
Vertical Individualism	-0.15*	-0.16*	-0.08*
Know other participants (0=no, 1=yes)	0.02	0.04	0.02
log10(Time reading instructions [s])	0.00	-0.03	0.00
Heard about study	0.04	0.04	0.05
Knew what was investigated	0.01	0.01	-0.01
Unsure what researchers trying to demonstrate	0.01	0.01	-0.03
Had good idea about hypotheses	0.02	0.01	0.01
Unclear about study aims	0.06	0.06	0.05
Correct comprehension of payoffs (0=No, 1=Yes)	0.01	-0.01	0.00

**Significant at the 5% level (including Bonferonni correction for 25 comparisons)*

SM1.2 Controlling for individual differences in the compliant-only analysis

As a more conclusive test for selection bias, I perform the compliant-only analysis from the RRR (predicting cooperation using condition, restricting to participants who obeyed the time constraints) and include as covariates *all* of the observed individual difference measures. If there was selection on a particular characteristic (e.g. compliant participants were older) and this difference was contributing to the difference in cooperation across conditions (e.g. because of exclusion, average age was higher in the time pressure condition, and as a result cooperation was also higher), then including that characteristic as a covariate in the regression would reduce the coefficient on the condition dummy. Because differences in the value of the covariate was explaining differences in cooperation across conditions, the model's estimate of the influence of condition would be reduced.

SM1.2.1 Simulations validating the method

To further cement this logic, consider a toy example in which I generate a dataset involving selection bias. To do so, I generate 3500 observations, each of which has

- A randomly generated condition variable *C* (0=control, 1=treatment)
- A randomly generated individual difference variable *ID* (0=low type, 1=high type)
- A randomly outcome value *X* drawn from a uniform distribution over [0,1] if *ID*=0 and over [0, 1.5] if *ID*=1

Thus, in this example, condition is totally irrelevant for the outcome whereas the individual difference measure has a strong impact on the outcome.

I then introduce selection bias by excluding all low type participants from the treatment (i.e. dropping participants for whom *C*=1 and *ID*=0). This leads to an expected difference in outcome *X* between conditions of 0.125 (mean *X* = .625 in the control, and *X* = .75 in the treatment).

Over 10,000 replicates, I compare the regression coefficients generated by predicting *X* with *C* as the only independent variable compared to predicting *X* using by *C* and *ID*. When only predicting with *C*, there is an apparent effect of *C* of b=.125. Conversely, when also including *ID*, the coefficient on *C* becomes b=.000, while the coefficient on *ID* is b=.25.

This selection-only example can be contrasted with another example that involves both selection bias and a real causal effect. For this example, the code is identical except that *X* values are drawn from a uniform distribution over [0,1] if *ID*=0 & *C*=0, over [0, 1.25] if *ID*=1 or *C*=1, and over [0, 1.5] if both *ID*=1 & *C*=1. (That is, there are independent increases in *X*'s expected value of 0.125 from having the characteristic and from being treated.)

In this second example, when only predicting *X* using *C*, there is an apparent effect of *C* of b=.188; and when also including *ID*, the coefficient on *C* becomes b=.125 and the coefficient on *ID* is b=.125.

Thus, in these toy examples we see that including the individual difference on which selection was occurring eliminates the apparent treatment effect when there is no causal effect, and effectively separates the true effect from the effect induced by selection bias when both are present.

SM1.2.2 Analysis of the RRR data

I now turn to the actual RRR data and perform the compliant-only analysis including all collected individual difference measures as covariates. Specifically, I run a regression for each lab predicting cooperation using a time pressure dummy and all the covariates (and excluding non-compliant participants), and meta-analyze the resulting coefficients on the time pressure dummy.

To provide a maximally stringent test for selection bias, my main analysis allowed non-linear functional forms for the continuous covariates. For age and reading time on the instructions,

which had a very large range of possible values, I included quadratic terms. I treated all other continuous measures as categorical (i.e. included dummies for each level), allowing arbitrarily complex functional forms (for the individualism and collectivism scales, I rounded to the nearest integer value before creating dummies). Note that because this analysis is carried out separately for each lab (and the resulting time pressure dummy coefficients meta-analyzed), the coefficients on each level of each covariates can be different for each lab.

Before looking at the effect of these covariates on the time pressure effect, I establish that these covariates do in fact relate to cooperativeness in the time pressure condition (and therefore could potentially contribute to selection bias). To do so, I examine the r^2 values generated by running a regression separately for each lab that predicts cooperation in the time pressure condition (including non-compliant participants) using all of these covariates. This results in an average r^2 = .94, indicating that these covariates can account for nearly all of the variance in cooperation behavior in the time pressure condition (the results for the forced delay condition are similar, r^2 = .92).

There is almost certainly a great deal of over-fitting happening in these regression models – because I am treating most of the covariates as categorical (and thus including dummies for each level), the number of independent variables is very large. Overfitting is not a problem for my analysis, however, because the goal is not to assess the relationship between the covariates and cooperation per se. Instead, the goal is to ask whether the time pressure effect in the compliant-only analysis remains strong even including the covariates. Therefore, by using this formulation which allows the covariates to account for as much as variation in cooperation as possible, I provide a maximally conservative and stringent test for selection bias based on the observed covariates. That is, if the compliant-only time pressure effect survives the inclusion of covariates, the presence of covariate overfitting strengthens – rather than undermining – the conclusion that that effect is not due to selection bias on the observed covariates.

Contrary to the selection bias account, predicting cooperation across conditions using a time pressure dummy as well as all of the covariates, and excluding non-compliant participants, produced a meta-analytic estimate of the time pressure effect size of 11.8 percentage points, 95% CI [5.0, 18.6]. This estimate is slightly *larger* than the effect size without covariates of 10.3 percentage points. Thus, even though the covariates could account for the vast majority of variation in cooperation, controlling for them did not diminish the observed time pressure effect.

The results were similar when reducing the possibility of overfitting by treating all continuous variables as linear. The covariates still explained a good deal of variation in cooperation behavior, although naturally much less than in the non-linear formulation above (time pressure condition, r^2 = .36; forced delay condition, r^2 = .37; including non-compliant participants in both analyses). And, critically, I continued to observed no meaningful dilution of the compliant-only time pressure effect size when including all covariates: 10.1 percentage points, 95% CI [6.6, 13.7].[1]

[1] This analysis implicitly assumes that the slopes for each covariate are parallel across conditions within each lab. I cannot credibly verify this assumption, however, because there are 21 labs x 25 covariates.

The results were also qualitatively equivalent when excluding data from the Srinivasan lab (where decision times were highly irregular compared to the other labs; see SM3 for details). Without covariates, the compliance-only time pressure effect estimate was 10.0 percentage points, 95% CI [5.7, 14.2]. When including covariates using non-linear functional forms, the compliance-only time pressure effect estimate was 10.6 percentage points, 95% CI [4.8, 16.5]. When including covariates using linear functional forms, the compliance-only time pressure effect estimate was 9.7 percentage points, 95% CI [6.1, 13.3].

Thus, I find no evidence that excluding non-compliant participants induces selection bias based on any of the observed individual differences. It is of course possible that there are unobserved covariates which *are* leading to selection.

Finally, for completeness I report in Table S3 the meta-analytic estimate of the difference in means between conditions (excluding non-compliant participants) for each covariate, which provides a sense of how much grist there is for selection to act on each covariate. I report both the estimate of the absolute value of the difference (since variation in either direction could lead to selection) and the raw value of the difference.

Table S3. *Meta-analytic estimate of the difference in means between conditions (excluding non-compliant participants) for each covariate.*

	Abs(Difference across conditions) excluding non-compliant participants	Difference across conditions excluding non-compliant participants
Male	0.077	0.049
Age	0.391	0.133
Trust in strangers	0.350	0.219
Trust in daily life interaction partners	0.309	0.212
People can be trusted	0.144	0.026
People are fair	0.113	0.002
People try to help	0.092	0.024
Can't count on strangers (Reverse coded)	0.106	0.011
Prior economic game experience (0=no, 1=yes)	0.068	0.030
Prior subject pool study experience (0=no, 1=yes)	0.073	0.052
Prior paid study experience (0=no, 1=yes)	0.058	0.031
Prior MTurk pool study experience (0=no, 1=yes)	0.027	0.004
Prior experience with deception studies (0=no, 1=yes)	0.085	0.067
Horizontal Collectivism	0.174	0.036
Horizontal Individualism	0.187	0.048
Vertical Collectivism	0.233	0.009
Vertical Individualism	0.209	-0.030
Know other participants (0=no, 1=yes)	0.082	0.054
log10(Time reading instructions [s])	0.043	-0.036
Heard about study	0.019	0.012
Knew what was investigated	0.398	-0.009
Unsure what researchers trying to demonstrate	0.356	-0.115
Had good idea about hypotheses	0.386	-0.012
Unclear about study aims	0.322	-0.028
Correct comprehension of payoffs (0=No, 1=Yes)	0.075	0.031

SM2. Expected correlation pattern in the presence of a true effect and no selection bias

Here I show that if there was a true causal effect of time pressure for participants who were successfully treated, and exclusion introduced no selection bias, one would expect a positive correlation across labs between (i) the time pressure effect when including all subjects (TP_{all}), and (ii) the amount that the time pressure effect changes after the exclusion is implemented ($\Delta_{exclusion} = TP_{exclusion} - TP_{all}$). To do so, I explore the predictions derived from a situation in which successful treatment increases cooperation, and the only consequence of excluding non-compliant participants is to make the analysis focused only on treated participants.

Before presenting a formal derivation and computer simulations, I will lay out the basic logic and assumptions.
a) Assume there is a true causal effect of time pressure among those who are compliant: participants who comply with instructions in the pressure condition (i.e. decide in less than 10 seconds) actually experience the treatment and are more cooperative. That is, participants who complied in the time pressure condition cooperate more than those in the forced delay condition.
b) Assume that those who don't comply with the instructions do not experience any treatment, so they cooperate at default levels. That is, non-compliant participants show no time pressure effect.
c) For simplicity, assume that the forced delay instruction also has no effect so that people in that condition contribute at default levels.
d) With these assumptions, we have two groups of participants: Those who cooperate more because they experienced the time pressure treatment and those who respond at some default level of cooperation.
e) In such a situation, the time pressure effect (difference in cooperation between time pressure and time delay conditions) will be larger in the compliant-only analysis than in the intent-to-treat (ITT) analysis: the ITT effect will be diluted by including non-compliant participants (who cooperated at the same default level as those in the time delay condition).
f) The average ITT effect should still be positive, as long as some participants complied (i.e. cooperated more), but that effect might be small compared to sampling noise (due to, for example, the widely documented individual differences in cooperation), especially when the fraction of compliant participants is low.
g) Nonetheless, a larger ITT effect will be associated with a larger compliant-only effect, because the ITT effect is positive only to the extent the treatment increased cooperation for those who complied.
h) When there is a substantial rate of non-compliance, the ITT effect is diluted, but it increases as the effectiveness of the treatment increases.
i) Critically, as the effectiveness of the treatment increases, the compliant-only effect will increase at a faster rate than the ITT, because the compliant-only effect is undiluted.
j) Consequently, the size of the ITT effect will be positively correlated with the difference between the compliant-only effect and the ITT effect.
k) Conversely, if there is no causal effect of time pressure, both the ITT effect and the compliant-only effect would be driven entirely by noise, and there would be no correlation between the size of the ITT effect and the difference between the compliant-only effect and the ITT effect.

SM2.1 Formal analysis

Because the distribution of contributions was strongly bimodal at 0 and 1, for simplicity I assume that people are either cooperators (contribute everything) or defectors (contribute nothing).

Let X_i be the actual time pressure effect in lab *i*'s subject pool: successfully decreasing the decision time of a participant to under 10s in subject pool *i* increases the probability of cooperating by *i* percentage points (relative to what that person would have done if in the forced delay condition).

However, the decision time of all participants in the time pressure condition of each lab is not successfully manipulated. Let c_i be the fraction of participants in the time pressure condition from lab *i* that *are* successfully manipulated (i.e. who are compliant with the manipulation check), and thus receive the treatment X_i. The (1-c_i) fraction of participants in the time pressure condition who are *not* successfully treated, conversely, receive no treatment and thus have an average effect size of 0.

The intent-to-treat time pressure effect over all participants TP_{all} in the sample from lab *i* is therefore given by

$$\begin{aligned} TP_{all,i} &= c_i X_i + (1\text{-}c_i)\ 0 + e_{i1} \\ &= c_i X_i + e_{i1.} \end{aligned}$$

where e_{i1} is a random variable representing sampling noise (i.e. random variation in the cooperativeness of participants in each condition). Note that if c_i is sufficiently small and e_{i1} is sufficiently negative, this ITT effect can be zero or negative even if X_i is greater than 0 (i.e. even if there is a true causal effect).

The compliant-only analysis excludes the (1-c_i) fraction of non-compliant participants and only examines the c_i fraction of compliant participants. Therefore, the resulting time pressure effect $TP_{exclude,i}$ is given by

$$TP_{exclude,i} = X_i + e_{i1} + e_{i2}$$

where e_{i2} is a random variable representing sampling noise based on endogenous reaction times (e.g. some untreated people wind up deciding more quickly than 10s of their own volition).

Therefore, the difference between the intent-to-treat and compliant-only analyses, $\Delta_{exclude,I}$, is given by:

$$\begin{aligned} \Delta_{exclude,i} &= TP_{exclude,I} - TP_{all,i} \\ &= X_i\ (1\text{-}c_i) + e_{i2} \end{aligned}$$

Note that this expression does not share a noise term with $TP_{all,}$ and therefore TP_{all} and $\Delta_{exclude}$ will not be correlated across labs due solely to correlated noise.

Instead, because both TP_{all} and $\Delta_{exclude}$ are increasing in X_i, they *will* be positively correlated across labs (holding c constant) if X_i varies across labs, such that there is a causal effect of time pressure that is variable in size (as predicted by the Social Heuristics Hypothesis). Conversely, they will be uncorrelated if X_i does not vary (e.g. because there is no causal effect of time pressure such that $X_i = 0$ for all i).

Thus, a positive correlation across labs TP_{all} and $\Delta_{exclude}$ is consistent with the existence of a true causal effect of time pressure which is exposed by exclusion, and a lack of selection bias.

SM2.2 Computer simulations

To further support this argument, I complement this formal analysis with computer simulations. In each simulation run I generate decisions for 75 players in the time pressure condition and 75 players in the forced delay condition for each of 21 labs (0 = defection, 1 = cooperation).

Each lab i has:

- A causal effect of time pressure X_i which specifies the increase in probability of cooperating for players in the time pressure condition who are successfully treated (randomly sampled from a uniform distribution over the interval [0, X_{max}], where X_{max} is varied across simulations).
- A baseline cooperation rate b_i which specifies the probability of cooperating in the absence of the time pressure treatment (randomly sampled from a uniform distribution over the interval [.5, .8], based on the range of RRR labs' average cooperation rate in the forced delay condition).
- A compliance rate c_i which specifies the probability of players in the time pressure condition being successfully treated (randomly sampled from a uniform distribution over the interval [.2, .5], based on the range of RRR labs' compliance rates in the time pressure condition).

Then, to generate each player's decision:

- Players in the forced delay condition cooperate with probability b_i.
- Players in the time pressure condition who are successfully treated cooperate with probability $b_i + X_i$.
- Players in the time pressure condition who are *not* successfully treated cooperate with probability b_i.

After generating the decisions for each player in each lab, I compute TP_{all} (sum of decisions of all players in the time pressure condition minus sum of decisions of all players in the forced delay condition), $TP_{exclude}$ (sum of decisions of compliant players in the time pressure condition minus sum of decisions of all players in the forced delay condition), and $\Delta_{exclude} = TP_{all} - TP_{exclude}$. Shown in Figure S1 is the average correlation between TP_{all} and $\Delta_{exclude}$, as a function of X_{max}, averaged over 100,000 simulation runs per X_{max} value. The specific magnitudes of the correlation should not be over-interpreted, as additional simulations show that these magnitudes can vary widely depending on the specific assumptions made about the distribution of X_{max} values. Instead of magnitudes, the key point here is simply that positive correlations occur.

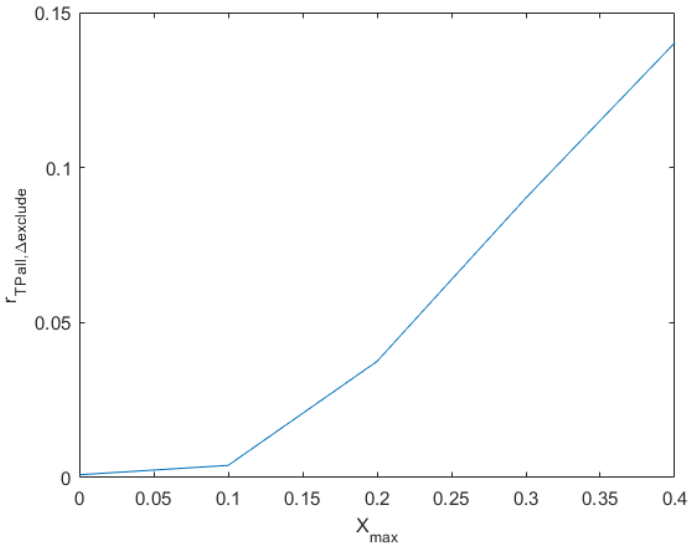

Figure S1. *Average cross-lab correlation over 100,000 simulation runs between* TP_{all} *and* $\Delta_{exclude}$ *as a function of* X_{max}.

SM3. Correlation pattern across labs when excluding based on cooperativeness

The "selection bias based on underlying correlation" account argues that excluding non-compliant participants selectively excludes slow (and therefore selfish) participants from the time pressure condition and fast (and therefore cooperative) participants from the time delay condition, leading to higher cooperation in the post-exclusion time pressure condition. In other words, decision time is a proxy for cooperativeness, and so exclusions based on compliance (i.e. decision time) are really just selection on cooperativeness. Here I use the RRR data to explore the implications of this proposal for the relationship between TP_{all} and $\Delta_{exclusion}$.

The values of TP_{all} (the intent-to-treat effect size) for each lab can be directly observed from the RRR data. The values of $TP_{exclusion}$, and therefore $\Delta_{exclusion}$, for each lab resulting from actually excluding non-compliant participants are also directly observable from the RRR data. As reported in my comment, there is a positive correlation between TP_{all} and the actual $\Delta_{exclusion}$ of r = .36, as predicted by a true causal effect and no selection bias (see in Section SM2 above).

Instead of using the actual $\Delta_{exclusion}$ values, what I do now is ask what values of $\Delta_{exclusion}$ result from choosing participants to exclude based directly on their cooperativeness, rather than based on the compliance of their decision times. For each lab, I determine the number of participants actually excluded because of their decision times (i.e. the number of actually non-compliant participants) – for lab i, let c_{Ti} be the number of compliant (non-excluded) participants in the time pressure condition and c_{Di} be the number of compliant (non-excluded) participants in the forced delay condition. Then, instead of basing inclusion/exclusion on decision times, I select a corresponding number of participants based on cooperativeness to include (selectively excluding non-cooperative participants from the time pressure condition and cooperative participants from the forced delay condition – who may or may not be the actually non-compliant participants).

Specifically, I calculate $\Delta_{exclusion}$ by including only the c_{Ti} most cooperative participants in the time pressure condition for each lab, and the c_{Di} least cooperative participants in the forced delay condition. Doing so produces large positive values of $\Delta_{exclusion}$ (as one would expect), but a *negative* correlation between TP_{all} and this cooperation-based $\Delta_{exclusion}$ of r = -.56.[2] The stark difference between this finding for cooperation-based exclusion and the actual pattern resulting from compliance-based exclusion can be seen in Figure 2 in my comment, reproduced here as Figure S2 for expositional purposes. (It is also important to note that excluding participants at random from the RRR data, rather than based on compliance or cooperativeness, does not lead any correlation between TP_{all} and $\Delta_{exclusion}$, average r = .00 over 10,000 simulation runs.)

[2] Similar results are obtained when the cooperation-based $\Delta_{exclusion}$ is calculated by probabilistically selecting c_{Ti} participants proportional to cooperativeness to include in the time pressure condition for each lab, and c_{Di} participants proportional to non-cooperativeness (1-cooperativeness) in the forced delay condition. Averaging over 10,000 such selections yields an average correlation between TP_{all} and this proportional version of the cooperation-based $\Delta_{exclusion}$ of r = -.47.

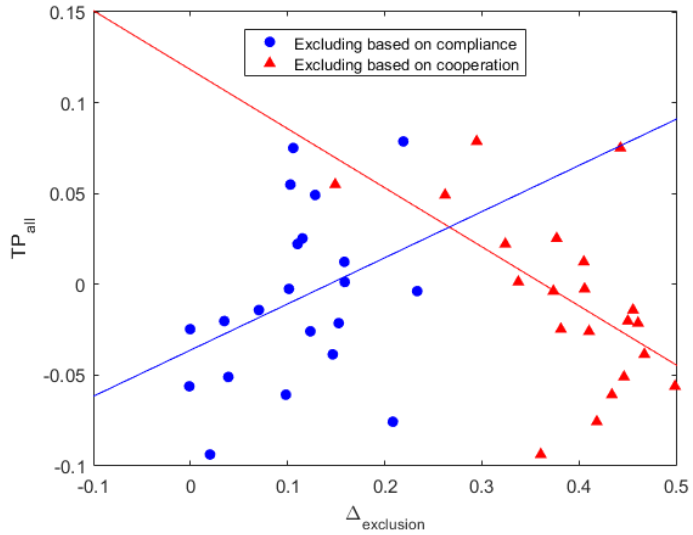

Figure S2. *Relationship between TP_{all} and $\Delta_{exclusion}$ in the RRR data when exclusions are made based on compliance with the time constraints (blue) versus cooperativeness (red). One dot per RRR lab. Least-squares fit lines added for visualization purposes.*

Because these analyses involve only 21 data points (one per lab), and correlations are potentially unstable with small sample sizes, I assess both correlations' robustness (and the robustness of the difference between the two correlations) using bootstrapping. To do so, I generate 500,000 bootstrap replicates where data for 21 labs are sampled with replacement from the actual 21 labs' data, and calculate the correlation between TP_{all} and both the compliance-based $\Delta_{exclusion}$ and the cooperation-based $\Delta_{exclusion}$. The results are shown in Figure S3.

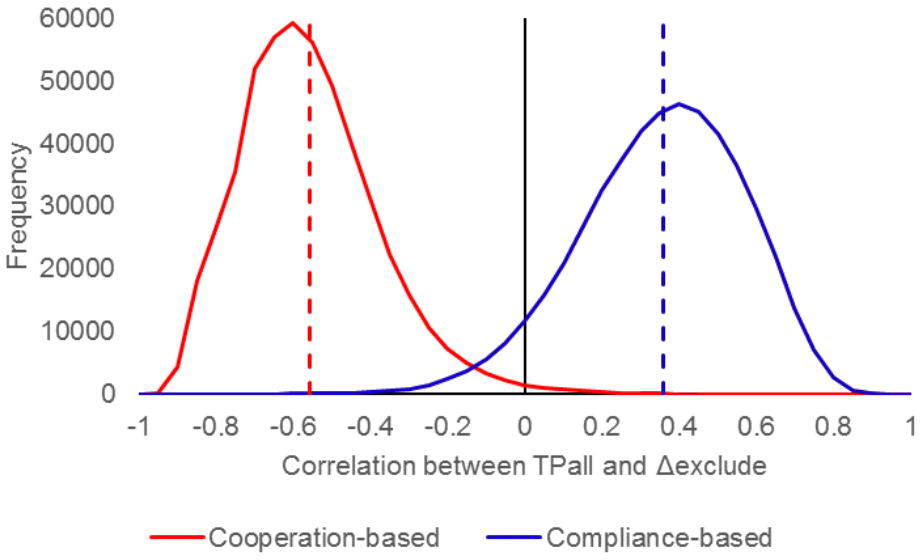

Figure S3. *Correlation between TP_{all} and $\Delta_{exclusion}$ in the RRR data when exclusions are made based on compliance with the time constraints (blue) versus cooperativeness (red), over 500,000 bootstrapped samples. Values observed in the actual data are indicated by vertical dotted lines.*

As can be seen, the correlations observed in the actual data sit nicely at the respective modes of the bootstrapping results, showing that the observed correlations are representative. Furthermore, the compliance-based exclusion correlation was larger than the cooperation-based exclusion in 498,848 of the bootstrap samples (implying $p = .002$ for the difference between the correlations based on the two different forms of exclusion); the compliance-based exclusion correlation was positive in 471,344 of the bootstrap samples (implying $p = .057$ for the existence of a positive compliance-based correlation); and the cooperation-based exclusion correlation was negative in 496,585 of the bootstrap samples (implying $p = .007$ for the existence of a negative cooperation-based correlation).

These results are even more stark when excluding data from the Srinivasan lab – as described in the RRR main text, the decision times from that lab were highly irregular, with the mean decision time in the time pressure condition being much larger than in any of the other labs (41s versus 11.5s to 15.2s in the other labs) and in fact being larger even than that lab's time delay condition. Thus, the compliance assignments in that lab's data are likely to be meaningless.

Accordingly, the correlation between TP_{all} and both types of exclusion are larger in magnitude when excluding this lab's data: compliance-based exclusion, $r = .40$, bootstrapped $p = .046$; cooperation-based exclusion, $r = -.68$, bootstrapped $p = .002$; difference between the two, bootstrapped $p = .0003$.

These analyses of the RRR data indicate that the observed positive relationship between TP_{all} and $\Delta_{exclusion}$ is not consistent with the argument that exclusion is inducing selection bias by selectively excluding non-cooperative participants from the time pressure condition and cooperative participants from the forced delay condition.