ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2017, Vol. 12(3) 527–542

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691617693624
www.psychologicalscience.org/PPS

Registered Replication Report: Rand, Greene, and Nowak (2012)

S. Bouwmeester*, P. P. J. L. Verkoeijen*, B. Aczel, F. Barbosa, L. Bègue, P. Brañas-Garza, T. G. H. Chmura, G. Cornelissen, F. S. Døssing, A. M. Espín, A. M. Evans, F. Ferreira-Santos, S. Fiedler, J. Flegr, M. Ghaffari, A. Glöckner, T. Goeschl, L. Guo, O. P. Hauser, R. Hernan-Gonzalez, A. Herrero, Z. Horne, P. Houdek, M. Johannesson, L. Koppel, P. Kujal, T. Laine, J. Lohse, E. C. Martins, C. Mauro, D. Mischkowski, S. Mukherjee, K. O. R. Myrseth, D. Navarro-Martínez, T. M. S. Neal, J. Novakova, R. Pagà, T. O. Paiva, B. Palfi, M. Piovesan, R.-M. Rahal, E. Salomon, N. Srinivasan, A. Srivastava, B. Szaszi, A. Szollosi, K. Ø. Thor, G. Tinghög, J. S. Trueblood, J. J. Van Bavel, A. E. van 't Veer, D. Västfjäll, M. Warner, E. Wengström, J. Wills, and C. E. Wollbrant
*Proposing authors

Multilab direct replication of: Study 7 from Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.

Protocol vetted by: David Rand

Protocol and manuscript edited by: Daniel Simons

Abstract

In an anonymous 4-person economic game, participants contributed more money to a common project (i.e., cooperated) when required to decide quickly than when forced to delay their decision (Rand, Greene & Nowak, 2012), a pattern consistent with the social heuristics hypothesis proposed by Rand and colleagues. The results of studies using time pressure have been mixed, with some replication attempts observing similar patterns (e.g., Rand et al., 2014) and others observing null effects (e.g., Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014). This Registered Replication Report (RRR) assessed the size and variability of the effect of time pressure on cooperative decisions by combining 21 separate, preregistered replications of the critical conditions from Study 7 of the original article (Rand et al., 2012). The primary planned analysis used data from all participants who were randomly assigned to conditions and who met the protocol inclusion criteria (an intent-to-treat approach that included the 65.9% of participants in the time-pressure condition and 7.5% in the forced-delay condition who did not adhere to the time constraints), and we observed a difference in contributions of −0.37 percentage points compared with an 8.6 percentage point difference calculated from the original data. Analyzing the data as the original article did, including data only for participants who complied with the time constraints, the RRR observed a 10.37 percentage point difference in contributions compared with a 15.31 percentage point difference in the original study. In combination, the results of the intent-to-treat analysis and the compliant-only analysis are consistent with the presence of selection biases and the absence of a causal effect of time pressure on cooperation.

Keywords

cooperation, social heuristic hypothesis, decision making, economic games, social psychology, replication

Corresponding Author:
Peter P. J. L. Verkoeijen, Department of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands/Learning and Innovation Center, Avans University of Applied Sciences, Breda, The Netherlands.
E-mail: verkoeijen@fsw.eur.nl

Rand, Greene, and Nowak (2012) argued that our social intuitions are shaped by our daily experiences and that those intuitions can determine whether our default response is selfish or cooperative. According to this *social heuristic hypothesis* (formalized mathematically by Bear & Rand, 2016; but see Myrseth & Wollbrant, 2016), people who regularly experience and benefit from cooperation in their daily lives will tend to develop cooperative intuitions as a default response, and those who are rewarded for non-cooperation will tend toward selfish intuitive responses. Although intuitions vary across people, deliberation is theorized to always favor self-interested behavior. For example, in one-shot, anonymous economic games in which selfish actions maximize one's payoff deliberation will favor non-cooperation, overriding any potential intuitive bias toward cooperation. The social heuristic hypothesis predicts more cooperation for judgments made intuitively (because some participants will default to cooperative responses) than when judgments are made with more deliberation (because deliberation will favor selfishness for all participants).

Rand and colleagues (2012) conducted a series of studies to assess the social heuristics hypothesis, using correlational quasi-experimental and experimental designs. Two of these experiments manipulated time pressure in a one-shot public goods game, with one testing participants from Amazon Mechanical Turk (Study 6) and another testing college students in a lab (Study 7). In these experiments, participants were either required to decide how much to contribute to the group within 10 s (time-pressure condition/intuitive decision making), or they were asked to wait at least 10 s before deciding on their contribution (forced-delay condition/reflection). In both experiments, when noncompliant participants were removed prior to analysis, the mean contribution was greater in the intuition/time-pressure condition than in the reflection/forced-delay condition. However, when including all participants in an intent-to-treat analysis—thereby preserving random assignment to conditions and avoiding selection biases—the contributions did not differ significantly between conditions in either experiment (Tinghög et al., 2013; they did differ significantly when combining across the two studies: Rand, Greene, & Nowak, 2013).

Since its publication, Rand et al.'s (2012) article has been highly influential. Yet, some studies have not found a difference in cooperation between participants placed under time pressure and those forced to delay their decision (Lohse, 2016; Tinghög et al., 2013; Verkoeijen & Bouwmeester, 2014). A recent meta-analysis of 51 published and unpublished studies (Rand, 2016; $N = 15{,}850$; the included studies showed no evidence of publication bias using p curve or Egger's test) testing different ways to induce intuition/deliberation in one-shot economic cooperation games reported a positive link between intuition and cooperation for the subset of studies in which defection always maximizes payoff (as in Rand et al., 2012 and in the unsuccessful replications cited earlier). However, the meta-analysis also showed a great deal of heterogeneity across studies, and many studies did not find a significant effect when considered individually.

One potentially critical issue involves how the analyses account for participants who did not comply with the time-pressure instructions (Tinghög et al., 2013). In most studies using the time-pressure and forced-delay procedure, many participants in the time-pressure condition respond too slowly (a substantially smaller proportion fail to respond slowly enough in the forced-delay condition). Rand et al. (2012) restricted their analysis to participants who had adhered to the task instructions (a compliant-only analysis; an approach adopted by others, e.g., Verkoeijen & Bouwmeester, 2014) rather than analyzing all participants assigned to each condition (an intent-to-treat analysis). Given the high exclusion rates and that compliance means different things in the two conditions (i.e., too fast in one condition and too slow in the other), a compliant-only analysis can introduce systematic differences between the participants in each condition. For example, a compliant-only analysis might selectively eliminate slow-responding participants from the time-pressure condition and not from the forced-delay condition, thereby disrupting random assignment to conditions. Such selective exclusion could produce a spurious difference between conditions that is driven by selection bias (i.e., differences between the participants) rather than by the experimental intervention.

Consistent with this possibility, correlational studies have found a negative correlation between response time and cooperation (Rand et al., 2012). Consequently, excluding slow respondents would be expected to introduce a bias favoring greater cooperation among the remaining participants in the time-pressure condition; of those randomly assigned to the time-pressure condition, the compliant-only participants are more likely to be fast responding and more cooperative. Given that about 50% of the participants in Rand et al. (2012) failed to respond on time, selection biases could partly explain the difference between conditions in the compliant-only analysis.

In contrast, an intent-to-treat analysis preserves random assignment because all people assigned to each condition are included in the analysis regardless of whether they adhered to the instructions. Consequently, it permits a valid causal inference about the generality of any observed difference between the conditions. However, an intent-to-treat analysis does not always allow a clear inference about the effectiveness of the treatment itself. It gives an unbiased estimate of the direction of the effect, but it can underestimate the potency of a treatment if some participants fail to adhere to the instructions. For example, if no participants adhered to the instructions, an intent-to-treat analysis would show no effect (the participants

in the two conditions essentially did the same thing because none of them followed instructions). But, it would not test whether time pressure would have been effective had participants in each condition actually followed the instructions. An intent-to-treat analysis tests whether there is a difference between people who were instructed to respond quickly and those who were instructed to respond after a delay, regardless of whether they actually adhered to those instructions.

For this Registered Replication Report (RRR), we use an intent-to-treat analysis as our primary analysis because it does not undermine random assignment to conditions and thereby allows for an unbiased causal inference. Given that the original study analyzed the data by excluding noncompliant participants, our protocol specified that we too would do this analysis if more than 10% of participants failed to comply with the time constraints.

In addition to these analyses, we conducted exploratory analyses that excluded participants with prior experience in tasks of this sort or who failed to comprehend the task. Participants who had previous experience with economic games of this sort may have learned that their intuitions can lead them astray in one-shot games and thus have been less likely to show the intuitive cooperation effect (Rand et al., 2012, 2014; Rand & Kraft-Todd, 2014). Also, participants who mistakenly believed that cooperation maximized payoff (i.e., those who did not comprehend the nature of the task) would have been more likely to cooperate even with deliberation (Strømland, Tjøtta, & Torsvik, 2016). In the analyses, we report the difference between the time-pressure and forced-delay conditions with and without these participants.

In summary, the goal of this RRR is to shed further light on the link between intuition and cooperation by assessing the size and variability of the difference in cooperation between participants responding under time pressure and those responding after a delay. More specifically, the RRR will replicate the between-subjects comparison (time pressure vs. forced delay) from Study 7 of Rand et al.'s (2012) study in a laboratory setting with college students as participants. The primary planned analysis includes all participants who met the protocol requirements and completed the task. The secondary analyses examine how the difference between conditions varies when excluding participants who have had prior experience with tasks like this one, who fail comprehension checks, and who do not comply with the task requirements (other exploratory moderator analyses, including individualism vs. collectivism, are reported at https://osf.io/scu2f/).

Protocol and Participating Laboratories

The protocol for a replication of Study 7 from Rand et al. (2012) was developed by Samantha Bouwmeester and Peter Verkoeijen. The original study's first author, David Rand, provided extensive input and guidance throughout the process, including the original materials and scripts. *Perspectives on Psychological Science* publicly announced a call for laboratories interested in participating in the RRR project on June 8, 2015, and after the July 6, 2015, deadline, 23 laboratories were accepted to join the project. Twenty-one laboratories completed the study, collecting enough data to meet the inclusion criteria. The final set of replications included studies from a range of institutions across 12 countries, with many participating laboratories headed by experts on decision making, public good games, and/or social psychology. Each laboratory preregistered their plans for implementing the approved protocol, and these plans were preapproved by the editor, who verified that they met all of the requirements for the study. The results from all completed studies are included in this report, regardless of their outcome.

The protocol specified minimum sample sizes, exclusion rules, and testing conditions, and each laboratory's preregistered implementation of the protocol specified their target and minimum sample size, testing setting, recruiting procedures, and other aspects of their implementation. All labs used the same experimental script for data collection, modifying it only when it was necessary to translate materials to languages other than English. The full protocol is available on the Open Science Framework (https://osf.io/scu2f/) and that project page includes links to each participating laboratory's implementation of the study.

Method

Subjects

The protocol required testing of at least 75 participants in each of the two conditions, and labs were strongly encouraged to test as large a sample as possible. With the minimum sample size, individual studies would be underpowered to reject the null hypothesis for the original sample size, but the goal of the RRR is not to determine whether each individual study obtains a statistically significant result. Rather, the goal is to estimate the effect size meta-analytically across studies. Consequently, these projects trade off power in individual studies against the desire to increase the number of participating laboratories.

Participants were between the ages of 18 and 34 and were recruited from undergraduate subject pools or the equivalent; each sample ranged between 20% and 80% women. Participants received a show-up payment or course credit for participating, and they also had an opportunity to earn more money as a result of a public goods game (see below). The show-up fees varied somewhat across laboratories depending on their typical payments for studies of this sort. For recruiting purposes, the

study was described as a "study of decision making," and other than the duration, location, and compensation, no other information was provided about the content of the study. In order to increase the likelihood that participants would be unfamiliar with studies of this sort, laboratories were encouraged to collect their data at the start of the semester and to recruit from student samples with less experience in psychology studies. The protocol also asked laboratories to collect data on prior study experience for each participant.

Given the design of the study, participants were tested in groups that were multiples of 4, with a minimum testing session size of 12 participants. In rare cases, when fewer than 12 participants attended a scheduled session, data from groups with 8 participants were permitted. Whenever the total number of participants to attend a session was not a multiple of 4, the extra participants were paid a "show up" fee and were not tested (or were asked to return for a later session). The minimum group size ensured that participants believed the explanation that the payoff depended on other people and that they could not determine which of the other people in the room were in their group.

Materials & procedures

The original study materials, including the instructions, scripts, and post-experiment questionnaires were converted into a Qualtrics script (http://www.qualtrics.com) that handled all data collection. The script is available at https://osf.io/scu2f/. Labs conducting testing in countries other than the United States and Canada translated the contents of the script and adapted the currency amounts to match a similar level of local purchasing power (see http://www.numbeo.com/cost-of-living/compare_countries.jsp). These translated Qualtrics scripts are available on each lab's project page.

The Qualtrics script randomly assigned participants to a time-pressure condition and a forced-delay condition, with the constraint that approximately equal numbers of participants would be assigned to each condition. The experimenter and other participants were blind to condition assignment, and participants were unaware of the existence of a condition different from their own. The Qualtrics script showed the following instructions to all participants:

> You have been randomly assigned to interact with 3 of the other people in the room. All of you receive this same set of instructions. You cannot participate in this study more than once. Each person in your group is given $4 for this interaction. You each decide how much of your $4 to keep for yourself, and how much (if any) to contribute to the group's common project (from 0 to 400 cents). All money contributed to the common project is doubled, and then split evenly among the 4 group members. Thus, for every 2 cents contributed to the common project, each group member receives 1 cent. If everyone contributes all of their $4, everyone's money will double: each of you will earn $8. But if everyone else contributes their $4, while you keep your $4, you will earn $10, while the others will earn only $6. That is because for every 2 cents you contribute, you get only 1 cent back. Thus you personally lose money on contributing. The other people really will make this decision too—there is no deception in this study. Once you and the other people have chosen how much to contribute, the interaction is over. None of you can affect each other's payoffs other than through the single decision in this interaction.

On the next screen, participants were asked to decide how much to contribute by using a slider, with a pointer that started in the center of the range and with several values marked (the starting position was not marked with a value). When participants moved the slider, it indicated the exact contribution for that slider position. Although the slider did not require a response, in order to select an exact contribution, participants needed to move the slider. If they pressed continue without moving the slider, their contribution was recorded as missing and their data were excluded from the analyses.

Participants in the time-pressure condition were told: "Please make your decision as quickly as possible. You must make your decision in less than 10 seconds!" The screen showed a timer that counted down from 10, stopping at zero. Participants in the forced-delay condition were told: "Please carefully consider you [sic] decision. You must wait and think for at least 10 seconds before making your decision!"[1] The screen showed a timer that counted up from 0 and continued counting until the participant responded. The script recorded each participant's contribution and the time when they submitted their decision. Note that the original study did not use timers; during protocol development, David Rand suggested adding them based on his experience from subsequent studies.

After their decision, participants answered questions and surveys to measure (a) comprehension of the task; (b) their justification for their contribution; (c) individualism or collectivism (Singelis, Triandis, Bhawuk, & Gelfand, 1995); (d) experience with tasks of this sort; (e) experience with research participation more generally; (f) self-reported perceptions of trust in others (a factor suggested by Rand as a possible moderator of the time-pressure effect); (g) awareness of the research hypothesis (PARH:

cf. Rubin, Paolini, & Crisp, 2010); (h) sex, birth year, and country; and (i) how many of the participants in the room they knew.

As in the original study, participants were paid by randomly grouping them with 3 other participants (without replacement) to determine the collective group contribution and payout amounts.

Data exclusions

Data were excluded for participants who were younger than 18 or older than 34 (determined by subtracting their self-reported birth year from 2015; participants who did not report their birth year were excluded), who did not complete all tasks, or who did not move the slider to select a specific contribution amount; data was also excluded when the experimenter/computer incorrectly administered the task or instructions. Exclusion decisions that depended on a judgment of an experimenter were made by someone blind to condition assignment and before examining that participant's contribution in the public goods task. All data, including those from excluded participants, are provided on each laboratory's Open Science Framework page and on the main page for the RRR. Secondary meta-analyses report the results when excluding participants who had experience with studies like this one (experience), who did not adhere to the time constraints (noncompliant), or who did not correctly answer the comprehension check questions (noncomprehending). The experience analysis included only those participants who responded with a "1—nothing like this scenario" to the question: "To what extent have you participated in studies like this one before? (i.e. where you choose how much to keep for yourself versus contributing to benefit others)."

Results

A total of 21 laboratories contributed data from a grand total of 3,596 participants. Table 1 presents sample demographics for each participating laboratory. The Appendix provides a brief description of each laboratory's study, including documentation of any departures from the official protocol or from their own preregistered plans.

Primary analysis

Given that labs varied in the currency used to pay participants, we calculated each person's contribution as a percentage of the maximum possible contribution. For each lab, we then computed the mean percentage contribution in the time-pressure and forced-delay conditions and the difference in means between them (pressure – delay).

The primary analysis includes all participants who met the protocol requirements and recorded a contribution (an intent-to-treat analysis). Figure 1 shows a forest plot and the results of a random-effects meta-analysis across all laboratories. Below that meta-analysis, it also provides the meta-analytic result when excluding experienced (n = 2,000 in the analysis), noncompliant (n = 2,276), or noncomprehending (n = 2,304) participants, and when excluding all three (n = 792). Table 2 summarizes the results for each laboratory separately for all participants and after applying the exclusion criteria.[2]

An intent-to-treat analysis of data from the original study showed an 8.6 [95% CI: −1.84, 19.00] percentage point difference in the amount contributed between the time-pressure (M = 49.4%) and forced-delay (M = 40.8%) conditions.[3] Across all participants, the meta-analytic effect size in the RRR was −0.37 percentage points [95% CI: −2.60, 1.86], a value smaller than observed in the original data and close to zero. The observed effects ranged from −9.36 to 7.87, and the variability across laboratories was consistent with what would be expected by chance, $Q(20)$ =16.84, p = .66, I^2 = 2.72%.

Analyses with data exclusions

The preregistered protocol specified that if more than 10% of participants failed to adhere to the time constraints, we would conduct a secondary analysis including only on those participants who complied with the time constraints (a compliant-only analysis; 34.1% and 92.5% compliance in the time-pressure and forced-delay conditions respectively). In this analysis, the meta-analytic difference between conditions was 10.37 percentage points compared with a 15.31 percentage point difference for the same analysis in the original study. The variability across laboratories was somewhat larger but not significantly different from what would be expected by chance, $Q(20)$ = 29.22, p = .084, I^2 = 33.04%.

The results of the meta-analyses excluding participants based on their experience or comprehension were similar to those of the primary meta-analysis that included all participants, with meta-analytical differences of −2.19 percentage points and −0.64 percentage points, respectively. The variability across labs again was consistent with what would be expected by chance: experienced $Q(20)$ = 25.31, p = .19, I^2 = 19.08%; noncomprehending $Q(20)$ = 14.93, p = .78, I^2 = 8.06%. Furthermore, when applying all three of these exclusion criteria (experience, noncompliance, and nonunderstanding), the meta-analytic difference between conditions was 12.34 percentage points, with cross-lab variability consistent with what would be expected by chance alone, $Q(18)$ = 15.20, p = .65, I^2 = 0.94%.

Table 1. Demographic Information for Each Contributing Lab

Lab	Country	Testing language	Condition	*N*	Women *n*	Age *M* (*SD*)	Understanding personal benefit *n*	Understanding group benefit *n*	Naive *n*
Aczel	Hungary	Hungarian	Time pressure	101	80	21.1 (1.8)	62	83	82
			Forced delay	102	77	21.4 (2.0)	66	84	87
Bègue	France	French	Time pressure	107	71	20.8 (2.0)	71	91	86
			Forced delay	114	84	20.7 (2.2)	75	97	97
Bouwmeester	The Netherlands	Dutch/English	Time pressure	87	70	20.2 (2.1)	71	80	65
			Forced delay	82	69	20.3 (2.3)	61	72	63
Espin	United Kingdom	English	Time pressure	77	39	22.6 (3.7)	36	50	39
			Forced delay	80	45	21.9 (3.7)	34	56	43
Evans	The Netherlands	Dutch	Time pressure	72	58	19.8 (1.7)	61	69	19
			Forced delay	72	58	19.5 (2.0)	59	69	22
Ferreira-Santos	Portugal	Portugese	Time pressure	81	62	20.5 (3.1)	54	57	57
			Forced delay	82	50	21.2 (3.4)	42	60	61
Fiedler	Germany	German	Time pressure	79	52	19.8 (2.3)	59	75	54
			Forced delay	75	49	19.9 (2.2)	56	68	48
Hauser	Unites States	English	Time pressure	84	56	21.5 (3.0)	75	81	26
			Forced delay	82	47	22.0 (3.3)	71	78	35
Hernan	United Kingdom	English	Time pressure	90	56	20.9 (2.1)	63	83	9
			Forced delay	92	53	20.9 (2.2)	68	84	6
Lohse	Germany	German	Time pressure	76	37	21.5 (2.4)	50	71	20
			Forced delay	80	31	21.8 (2.4)	58	71	26
Mischkowski	Germany	German	Time pressure	97	64	23.6 (2.7)	65	88	29
			Forced delay	91	57	24.5 (2.7)	62	86	19
Neal	United States	English	Time pressure	81	61	22.5 (4.1)	53	67	72
			Forced delay	75	53	21.6 (2.8)	48	63	64
Novakova	Czech Republic	Czech	Time pressure	101	70	22.2 (2.6)	78	89	56
			Forced delay	101	60	22.5 (2.6)	78	94	55
Pagà	Spain	Spanish	Time pressure	79	47	21.2 (3.0)	63	72	21
			Forced delay	78	45	21.9 (3.3)	62	70	27
Piovesan	Denmark	Danish	Time pressure	107	30	20.6 (1.6)	74	85	91
			Forced delay	109	30	20.6 (1.5)	61	97	83
Salomon	United States	English	Time pressure	98	70	19.9 (1.6)	81	82	74
			Forced delay	100	63	19.7 (1.4)	79	85	69
Srinivasan	India	Hindi	Time pressure	88	39	21.4 (2.5)	31	36	26
			Forced delay	89	45	22.0 (3.1)	35	41	43
Tinghög	Sweden	Swedish	Time pressure	83	34	22.4 (2.4)	55	71	58
			Forced delay	81	38	21.6 (2.2)	57	73	72
Trueblood	United States	English	Time pressure	75	56	21.4 (3.5)	58	73	57
			Forced delay	68	43	21.4 (3.3)	56	63	49
Wills	United States	English	Time pressure	75	31	22.3 (2.5)	60	69	18
			Forced delay	72	27	22.0 (2.1)	50	64	15
Wollbrant	Sweden	Swedish	Time pressure	66	25	25.1 (3.5)	44	57	23
			Forced delay	67	31	24.9 (3.1)	54	60	34

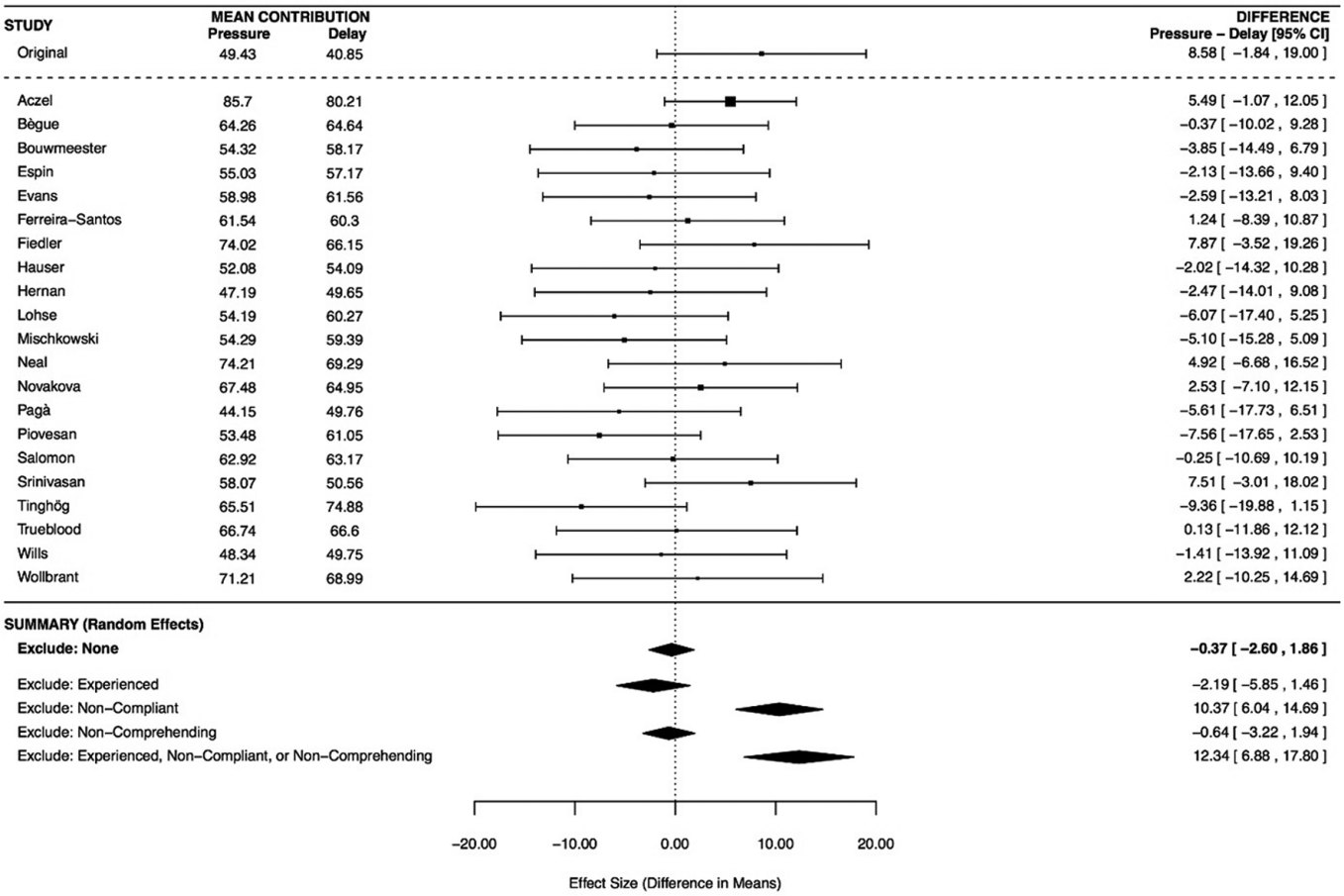

Fig. 1. Forest plot and meta-analytic result for the difference in contributions between the time-pressure and forced-delay conditions. Studies in the forest plot are listed alphabetically by the last name of the first author for that lab's study with the original result presented at the top. The mean difference for each lab is indicated by a square with the size corresponding to the inverse of the standard error of the difference score for that lab. The error bars indicate 95% confidence intervals around that laboratory's mean difference. The diamonds in the Summary section represent the results of random-effects meta-analyses of the RRR studies with the width representing a 95% confidence interval around the meta-analytic difference. None of these meta-analyses includes the original Rand, Greene, and Nowak (2012) result. The first diamond corresponds to the data in the forest plot and represents the primary planned meta-analysis with all participants. The next three diamonds show the meta-analytic difference after excluding experienced, noncompliant, or noncomprehending participants. The final diamond provides the meta-analytic difference when excluding participants who failed to meet any one of these criteria. A forest plot for the data excluding noncompliant participants is provided in the General Discussion section. Forest plots for the other meta-analyses are available at https://osf.io/scu2f/.

Additional exploratory analyses examined the role of a number of other moderators, including trust in others, gender, age, individualism or collectivism, whether or not participants knew other participants, total studies participated in previously, and participation in deceptive studies. The results of these meta-analyses are presented in Table 3, and the associated forest plots are available at https://osf.io/scu2f/.

General Discussion

This RRR featured data from 21 laboratories and a total of 3,596 participants. The studies were conducted according to a vetted design, and the analysis scripts were created while blind to the actual outcomes of the studies (although they were updated to address formatting issues, to provide more complete output, and to correct errors). The primary planned analysis in the RRR—an intent-to-treat approach including all participants—revealed a difference in contributions of −0.37 percentage points between the time-pressure condition and the forced-delay condition. This meta-analytic result is close to zero and smaller than the 8.6 percentage point difference computed from the original data. However, analyzing the data in the same way that the original article did—a compliant-only analysis that excludes participants who did not adhere to the time constraints—revealed a difference in contributions between conditions of 10.37 percentage points compared with a difference of 15.31 percentage points in the original study (see Fig. 2). The

Table 2. Decision Times, Contributions, and Sample Sizes in Each Lab, Shown With and Without Exclusions

		Time pressure			Forced delay		
Lab	Condition	*N*	Decision time *M* (*SD*)	Contribution *M* (*SD*)	*N*	Decision time *M* (*SD*)	Contribution *M* (*SD*)
Aczel	All participants	101	13.9 (5.4)	85.7 (20.8)	102	26.5 (11)	80.2 (26.5)
	Excluding experienced	82	14.2 (5.4)	87.4 (19)	87	26.3 (10.9)	81.2 (26.9)
	Excluding noncompliant	20	8 (0.9)	96.2 (10.1)	99	27.1 (10.7)	80.6 (25.5)
	Excluding noncomprehending	58	12.8 (4.3)	92.1 (15.3)	58	26.9 (10)	84.7 (22.5)
	Any exclusion	7	7.8 (0.6)	97.4 (6.9)	49	27 (10.5)	86.8 (21.3)
Bègue	All participants	107	14.7 (6.7)	64.3 (38.1)	114	33.5 (24.9)	64.6 (35.1)
	Excluding experienced	86	14.3 (6.7)	65.1 (38.9)	97	33.2 (21.7)	63.2 (37)
	Excluding noncompliant	25	8.3 (1.4)	86.5 (29.6)	109	34.7 (24.8)	63.6 (35.1)
	Excluding noncomprehending	65	14 (5.5)	66.7 (38.1)	71	35.9 (27.2)	62.7 (38.2)
	Any exclusion	16	8 (1.6)	89.6 (28.5)	59	35.7 (21.3)	60.5 (40.2)
Bouwmeester	All participants	87	15 (7.4)	54.3 (36.7)	82	32.6 (24)	58.2 (33.7)
	Excluding experienced	65	15.7 (8)	54.1 (36.5)	63	34.8 (26.5)	58.6 (34.8)
	Excluding noncompliant	15	7.4 (2)	69.2 (45.5)	80	33.2 (24)	58.4 (33.1)
	Excluding noncomprehending	67	15.4 (7.9)	56.1 (37.7)	58	34.4 (26.8)	58.7 (32.7)
	Any exclusion	6	7 (2.4)	83.3 (40.8)	43	37.7 (29.9)	58.2 (34.4)
Espín	All participants	77	12.7 (7.3)	55 (38.1)	80	24.2 (12.7)	57.2 (35.6)
	Excluding experienced	39	14.4 (9.5)	53.4 (42.8)	43	26.5 (13.6)	61.1 (33.9)
	Excluding noncompliant	31	7.9 (1.7)	68.8 (38)	71	26.3 (11.9)	55.6 (34.6)
	Excluding noncomprehending	29	12.3 (7.2)	52.6 (40.7)	32	23.3 (14.3)	57.8 (38.4)
	Any exclusion	8	7.8 (1)	64.8 (48.9)	15	32.7 (15.1)	51.1 (33.2)
Evans	All participants	72	15.3 (8.7)	59 (35.2)	72	32.4 (16.7)	61.6 (29.6)
	Excluding experienced	19	13.2 (4.8)	54.2 (39)	22	32.1 (19.7)	54 (30.1)
	Excluding noncompliant	23	8 (1.8)	71.3 (38.6)	72	32.4 (16.7)	61.6 (29.6)
	Excluding noncomprehending	60	16 (8.9)	56.8 (35)	58	32.5 (17.3)	60.5 (29.1)
	Any exclusion	5	8.3 (1)	60.1 (42.4)	18	33.7 (21.4)	52.6 (32.5)
Ferreira-Santos	All participants	81	13.8 (10)	61.5 (28.5)	82	24.7 (15.1)	60.3 (34)
	Excluding experienced	57	12.9 (5.8)	61 (29.7)	61	23.9 (10.8)	58.7 (33.8)
	Excluding noncompliant	22	8.3 (1.7)	76.8 (27.1)	78	25.6 (15)	59.7 (33.7)
	Excluding noncomprehending	43	13.4 (6.6)	67.9 (25.3)	36	25.2 (18.2)	62.1 (36.5)
	Any exclusion	8	8.1 (2)	79.4 (26.5)	28	23.9 (6.9)	60.4 (36.8)
Fiedler	All participants	79	12.2 (4.9)	74.0 (32.7)	75	27.6 (15.9)	66.2 (39.3)
	Excluding experienced	54	11.7 (4.5)	75.6 (31.5)	48	29.4 (16.4)	67.4 (38.4)
	Excluding noncompliant	31	7.5 (1.7)	92.4 (20.8)	68	29.8 (15.1)	62.7 (39.6)
	Excluding noncomprehending	58	12.7 (5.2)	74.2 (34.6)	53	28.0 (16.9)	69.5 (37.7)
	Any exclusion	15	7.4 (1.6)	89.8 (27.2)	33	30.4 (17.1)	70.1 (36.9)
Hauser	All participants	84	11.6 (4.5)	52.1 (39.8)	82	24.4 (14.8)	54.1 (41.1)
	Excluding experienced	26	11.8 (4.2)	65.8 (37.3)	35	26.8 (14.5)	56.7 (39.3)
	Excluding noncompliant	38	8 (1.4)	53.3 (42.2)	72	27 (14)	51.9 (39.9)
	Excluding noncomprehending	73	11.7 (4.5)	50 (39.7)	68	23.9 (15.1)	54.7 (41.8)
	Any exclusion	8	7.8 (1.3)	78.8 (36.4)	25	28.8 (14.7)	55.7 (40)
Hernan	All participants	90	11.5 (5.1)	47.2 (39.3)	92	27.2 (20)	49.7 (40.2)
	Excluding experienced	9	16.3 (8.3)	51.9 (33.3)	6	35.7 (34.7)	83.3 (27)
	Excluding noncompliant	43	7.5 (1.6)	48.3 (46.7)	88	28 (20)	50.8 (39.7)
	Excluding noncomprehending	60	11.1 (5.6)	45.9 (40.4)	64	28.2 (22.2)	50.2 (38.8)
	Any exclusion	1	7.5 (NA)	100 (NA)	4	36.2 (41)	75 (30.6)
Lohse	All participants	76	13.9 (7.8)	54.2 (38.6)	80	27.9 (15.3)	60.3 (33.5)
	Excluding experienced	20	12.7 (6.1)	57.3 (38.1)	26	29.3 (13.9)	55.3 (33.6)
	Excluding noncompliant	25	8.2 (1.4)	62.8 (43.2)	77	28.8 (15)	59.1 (33.5)
	Excluding noncomprehending	49	12.8 (6.7)	55.6 (41.3)	56	30.1 (16.7)	62.2 (35.9)
	Any exclusion	6	8.6 (1.8)	59.1 (48)	15	32.4 (14.3)	57.6 (38.4)
Mischkowski	All participants	97	12.2 (4.4)	54.3 (35)	91	22 (10.8)	59.4 (36.3)
	Excluding experienced	29	13.1 (5.2)	57.6 (34)	19	26 (14.1)	73.9 (30.7)
	Excluding noncompliant	29	7.9 (1.8)	60.5 (41.6)	84	23.2 (10.2)	60.2 (35.2)

(Continued)

Table 2. (Continued)

		Time pressure			Forced delay		
Lab	Condition	*N*	Decision time *M* (*SD*)	Contribution *M* (*SD*)	*N*	Decision time *M* (*SD*)	Contribution *M* (*SD*)
	Excluding noncomprehending	62	12.5 (4.9)	49.3 (37.1)	60	22.3 (10.1)	60.1 (38)
	Any exclusion	3	7.9 (1.5)	72.4 (47.8)	10	25 (9.4)	83.6 (23)
Neal	All participants	81	13 (5.9)	74.2 (34.4)	75	30.3 (18.8)	69.3 (39.5)
	Excluding experienced	72	13.1 (5.5)	73.7 (34.7)	64	30.3 (18.9)	71 (38.9)
	Excluding noncompliant	27	7.5 (1.8)	86.6 (30.9)	74	30.6 (18.7)	68.9 (39.6)
	Excluding noncomprehending	47	12.6 (6.4)	74 (36.2)	39	33.1 (21.7)	73.9 (36.5)
	Any exclusion	14	7.4 (1.8)	94.2 (21.6)	33	33.3 (21.1)	72.6 (36.3)
Novakova	All participants	101	13.3 (6.3)	67.5 (33)	101	25.8 (18.7)	65 (36.7)
	Excluding experienced	56	12.5 (6.1)	68.9 (30.9)	55	29.4 (20.3)	74 (27.6)
	Excluding noncompliant	36	7.7 (1.6)	77.5 (32.6)	87	28.8 (18.4)	63.5 (35.8)
	Excluding noncomprehending	69	13.2 (6.8)	65.1 (34.6)	75	25.7 (20.1)	63.4 (37.8)
	Any exclusion	18	7.7 (1.6)	69 (38.6)	33	31.7 (22.6)	73.5 (28.2)
Pagà	All participants	79	12.2 (5.9)	44.2 (38.8)	78	22.7 (17.2)	49.8 (38.7)
	Excluding experienced	21	12.1 (4.4)	39.4 (29)	27	17.4 (12.5)	57.4 (39.4)
	Excluding noncompliant	37	8.1 (1.5)	44.6 (44)	62	26.9 (17)	50.3 (36.5)
	Excluding noncomprehending	58	12.2 (6)	37.9 (38)	58	25.4 (18.3)	44.1 (37)
	Any exclusion	7	8.2 (1.5)	33.2 (37.8)	11	23.3 (12)	42.4 (35.4)
Piovesan	All participants	107	13.1 (6.9)	53.5 (36.2)	109	26.9 (18.8)	61 (39.4)
	Excluding experienced	91	12.9 (6.4)	52.1 (36.2)	83	28.2 (20.1)	61.9 (39)
	Excluding noncompliant	42	7.6 (1.7)	73.3 (35.9)	100	28.7 (18.6)	60 (39)
	Excluding noncomprehending	65	11.7 (5.4)	53.9 (37.8)	57	23.8 (15.1)	61.9 (42.5)
	Any exclusion	26	7.5 (1.8)	71.7 (35.6)	40	26.2 (14.6)	61.2 (42.5)
Salomon	All participants	98	12 (5.8)	62.9 (38.3)	100	33.7 (34.2)	63.2 (36.7)
	Excluding experienced	74	12.3 (5.9)	61.8 (38.1)	69	30.3 (24.9)	66.6 (35.6)
	Excluding noncompliant	40	7.6 (1.8)	69.4 (42.6)	91	36.4 (34.6)	59.5 (36.5)
	Excluding noncomprehending	71	11.7 (5.2)	66.7 (37.7)	72	33.9 (27.1)	65.4 (36.8)
	Any exclusion	19	7.2 (1.8)	78 (39.9)	46	34.8 (26.4)	64.4 (36.8)
Srinivasan	All participants	88	41.1 (32.6)	58.1 (35.4)	89	42.5 (29.5)	50.6 (36)
	Excluding experienced	26	38.4 (42.9)	53.6 (40.1)	43	35.9 (25)	52.6 (37.4)
	Excluding noncompliant	8	8.1 (2.1)	68 (38)	85	44.2 (29.1)	49.9 (35.7)
	Excluding noncomprehending	20	42.1 (34.7)	65.9 (39.6)	19	32.1 (17.5)	57.5 (42.1)
	Any exclusion	0	NA (NA)	NA (NA)	7	31.2 (11.8)	71.1 (36.7)
Tinghög	All participants	83	12 (5.3)	65.5 (35.6)	81	30 (24.7)	74.9 (33)
	Excluding experienced	58	11.9 (5.5)	68.8 (33.3)	72	29 (21.5)	75.7 (32.5)
	Excluding noncompliant	35	7.9 (1.9)	66.6 (41.5)	72	32.9 (24.7)	74 (32.9)
	Excluding noncomprehending	54	12.3 (5.5)	68.3 (34.9)	55	34.3 (28.1)	75.7 (33.2)
	Any exclusion	15	8.1 (1.7)	75.3 (36.8)	46	34.5 (23.8)	76 (32.4)
Trueblood	All participants	75	12.1 (5.6)	66.7 (33.2)	68	27.7 (13.1)	66.6 (39.8)
	Excluding experienced	57	11.8 (5.1)	70.1 (31.6)	49	28.5 (13.3)	68.6 (39.7)
	Excluding noncompliant	33	7.5 (1.9)	83.1 (31.7)	63	29.3 (12.1)	67.1 (39)
	Excluding noncomprehending	57	11.7 (5.1)	67.1 (36.5)	52	28 (14)	63.9 (42.1)
	Any exclusion	20	7.5 (1.9)	92 (25.5)	36	30.7 (13)	67.6 (40.3)
Wills	All participants	75	11.8 (4.5)	48.3 (39.3)	72	25.5 (23.4)	49.8 (38)
	Excluding experienced	18	11.8 (3.9)	39.5 (41.3)	15	32.3 (29.8)	63.9 (34)
	Excluding noncompliant	31	8 (1.3)	52.6 (47.4)	66	27.2 (23.8)	47 (36.9)
	Excluding noncomprehending	57	11.4 (3.8)	49.5 (39.5)	48	27.1 (26.7)	49.4 (39.4)
	Any exclusion	5	7.7 (1.5)	60 (54.8)	5	43.9 (46.7)	61 (38.5)
Wollbrant	All participants	66	11.6 (4.7)	71.2 (35.3)	67	26.8 (23.8)	69 (38)
	Excluding experienced	23	12.5 (6.3)	66.7 (38.4)	34	30.3 (29.8)	74.9 (35.6)
	Excluding noncompliant	27	7.9 (1.6)	80.3 (36.2)	60	29 (24.2)	67 (37.9)
	Excluding noncomprehending	41	11.1 (3.6)	65.5 (38.6)	52	26 (24.6)	70.2 (38.1)
	Any exclusion	6	8 (1.7)	83.3 (40.8)	23	34.3 (34.1)	73.8 (37.4)

Table 3. Results of the Moderator Meta-Analyses

Moderator	Type of analysis	Exclusions	Meta-analytic result (# of labs, value [95% CI], Q, I^2)
Trust (lab level)	Meta-regression	None	21, –3.66 [–8.46, 1.14], 14.89, 2.95%
		All	19, –11.66 [–28.64, 5.33], 13.54, 0.59%
Trust (individual)	Slope difference	None	21, –0.14 [–1.38, 1.09], 14.55, 0.00%
		All	19, –3.30 [–7.52, 0.92], 19.18, 13.81%
Age	Slope difference	None	21, –0.10 [–0.92, 0.72], 14.57, 0.00%
		All	19, –1.02 [–3.27, 1.22], 13.62, 0.00%
Horizontal individualism	Slope difference	None	21, 0.63 [–1.84, 3.09], 17.12, 0.00%
		All	19, –1.76 [–8.43, 4.91], 18.25, 2.37%
Vertical individualism	Slope difference	None	21, 0.43 [–1.11, 1.97], 14.83, 0.00%
		All	19, –2.09 [–7.65, 3.47], 26.60, 29.87%
Horizontal collectivism	Slope difference	None	21, –0.38 [–3.24, 2.49], 27.74, 22.30%
		All	19, –0.15 [–10.83, 10.53], 50.35, 61.09%
Vertical collectivism	Slope difference	None	21, –0.04 [–1.95, 1.87], 14.98, 0.00%
		All	19, 1.57 [–6.38, 9.51], 30.55, 21.40%
Gender (0 = male, 1 = female)	Effect size difference	None	21, –3.81 [–10.58, 2.96], 32.39, 38.40%
		All	18, –9.60 [–19.01, –0.18], 8.36, 0.00%
Subject pool study experience (0 = no, 1 = yes)	Effect size difference	None	21, 4.46 [–1.86, 10.78], 25.06, 16.66%
		All	18, 7.17 [–4.29, 18.63], 12.00, 0.00%
Paid study experience (0 = no, 1 = yes)	Effect size difference	None	20, –2.12 [–9.55, 5.31], 24.13, 16.19%
		All	12, –2.48 [–32.03, 27.06], 22.28, 53.20%
MTurk pool study experience (0 = no, 1 = yes)	Effect size difference	None	19, –0.61 [–8.98, 7.77], 15.35, 0.00%
		All	8, –8.07 [–34.78, 18.65], 6.25, 0.00%
Know other participants (0 = no, 1 = yes)	Effect size difference	None	20, –5.46 [–9.41, –1.52], 10.01, 0.00%
		All	17, –6.51 [–20.36, 7.34], 15.92, 0.00%
Deception (0 = no, 1 = yes)	Effect size difference	None	20, –0.99 [–6.40, 4.42], 14.73, 0.00%
		All	14, –9.25 [–25.98, 7.47], 10.07, 0.00%

Note: The Trust (lab level) moderator analysis compares the difference between conditions as a function of the mean level of trust for that lab using a meta-regression approach. Slope differences were used for continuous moderators (e.g., age), and they reflect the difference in slopes in the time-pressure and forced-delay conditions as a function of that moderator. It is conducted at the individual level for each lab, and the result is the meta-analytic difference in slopes across labs. Positive values for the meta-analytic result mean that the difference between the time-pressure and forced-delay slopes was larger for larger values of the moderator. Effect size differences were used for dichotomous moderators (i.e., gender), and they reflect the difference in the effect size for each level of the moderator. The meta-analytic result is the average difference in effect sizes across all labs for that analysis. Note, however, that some labs were not included in some analysis if, after exclusions, they had no data for one condition. The number of included labs is indicated in the rightmost column. The Exclusions column indicates the results with no exclusions and when participants were excluded based on experience, compliance or comprehension.

larger difference for a compliant-only analysis than for an intent-to-treat analysis is consistent with data from the original study and with a recent meta-analysis (Rand, 2016) that reported a 1.3 percentage point difference for an intent-to-treat analysis and a 4.3 percentage point difference for a compliant-only analysis in the subset of studies using time pressure to induce intuitive decision making.

The lack of a difference between the time-pressure and forced-delay conditions in the intent-to-treat analysis shows that instructing people to respond quickly or slowly had no effect on the amount of their contribution. The compliant-only analysis revealed a positive relationship between time pressure and contribution. However, this analysis does not allow for a causal inference because excluding participants based on their performance in the task can introduce a bias among the subjects assigned to each group that yields a spurious difference; any bias that undermines random assignment precludes a causal inference about the effect of time pressure on contributions. Excluding noncompliant participants could introduce many different forms of bias, and those biases could even vary depending on other factors (e.g., whether or not cooperation is appealing in that task; Evans, Dillon, & Rand, 2015; Krajbich, Bartling, Hare, & Fehr, 2015).

When an intent-to-treat analysis shows no difference, the only way that a difference between conditions in a compliant-only analysis could be consistent with the effectiveness of the treatment would be if those participants who did not comply actually experienced a causal

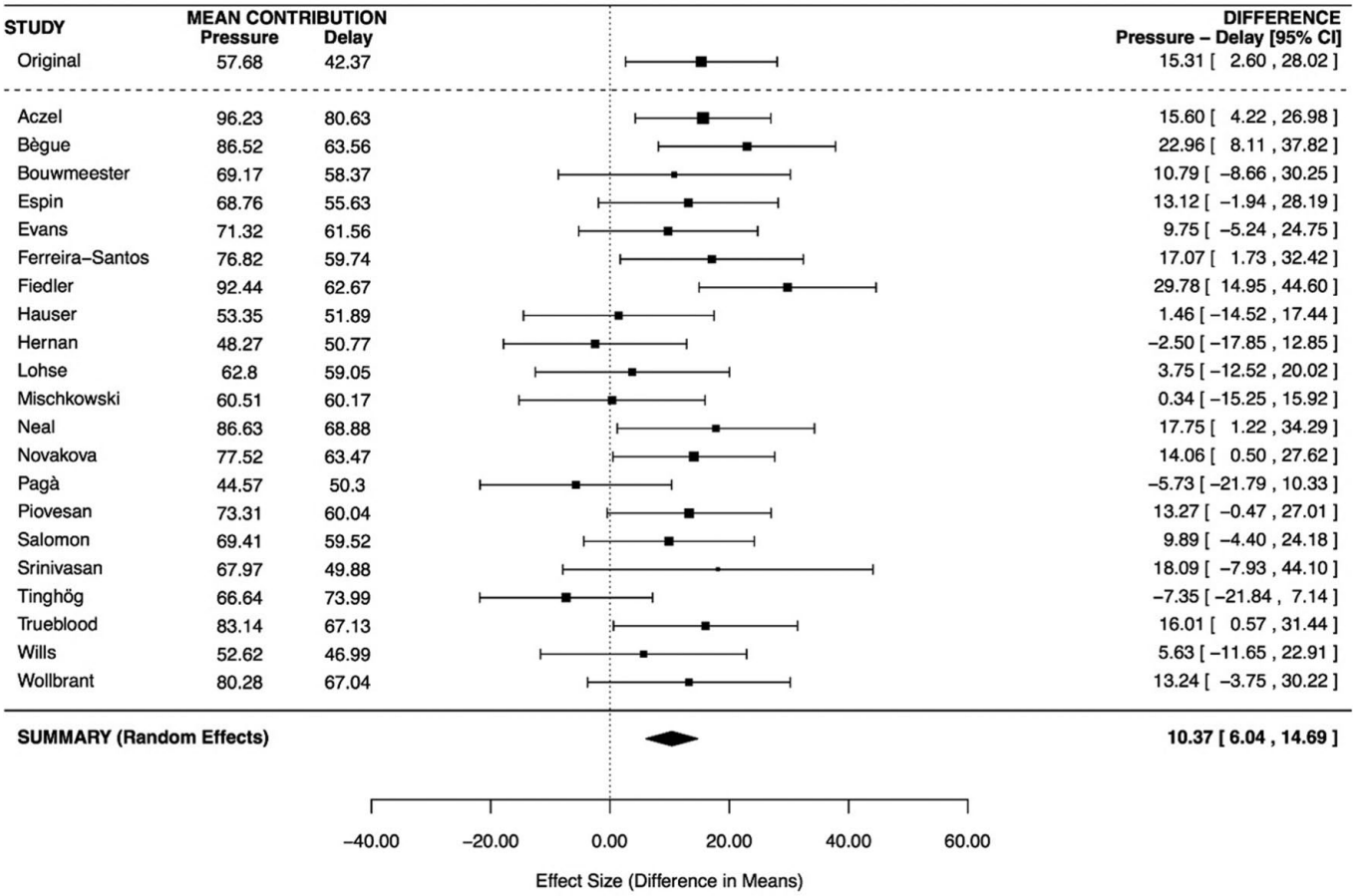

Fig. 2. Forest plot for the difference in contributions between the time-pressure and forced-delay conditions after excluding participants who did not comply with the time constraints. Studies in the forest plot are listed alphabetically by the last name of the first author for that lab's study, with the original result presented at the top. The mean difference for each lab is indicated by a square with the size corresponding to the inverse of the standard error of the difference score for that lab. The error bars indicate 95% confidence intervals around that laboratory's mean difference. The diamond represents the results of random-effects meta-analyses of the RRR studies with the width representing a 95% confidence interval around the meta-analytic effect size. The meta-analytic effect does not include the original Rand, Greene, and Nowak (2012) result.

induction that resulted in a more extreme effect in the opposite direction. In this case those who did not comply with the time-pressure instructions would have to have experienced a different treatment effect, one that made them even more deliberative than those who complied with the forced-delay instructions. Although such a pattern is possible in principle, it would require additional empirical evidence to demonstrate that causal relationship in the absence of selection biases. It might also require adjustments to the social heuristic hypothesis to explain why those who did not comply would be more likely to deliberate than would those who were in the condition designed to induce deliberation. Without such evidence, the most straightforward interpretation of the pattern of results is that the difference in the compliant-only analysis resulted from selection biases and that the RRR does not provide evidence for an effect of speeded versus delayed responses on cooperation.

Conclusion

Overall, the results of the primary analysis in this RRR showed essentially no difference in contributions between the time-pressure and forced-delay conditions: the point estimate was opposite that predicted by the hypothesis and was close to zero. A secondary, compliant-only analysis did show an approximately 10.37 percentage point difference between conditions (somewhat smaller than the original 15.31 percentage point difference). However, the compliant-only analysis does not allow for causal claims about the intervention due to potential selection biases.

Given the challenges of interpreting compliant-only analyses in the face of substantial exclusions, future studies of the effect of intuition on cooperation should strive to avoid high rates of non-compliance. One possibility would be to redesign the time-pressure procedures in

such a way that participants can consistently meet the time constraints. Given the challenges in doing so and the need to define an arbitrary timing threshold between intuitive and deliberative judgments (i.e., a 10-s cutoff), it might be more productive to test the social heuristic hypothesis using other ways of inducing intuitive or deliberative processing.

Appendix: Contributing Laboratories

Lead Lab

Samantha Bouwmeester, Erasmus University
Peter P. J. L. Verkoeijen, Erasmus University
OSF page: https://osf.io/xz7jr/

A total of 185 students were recruited from the psychology subject pool from the Erasmus University Rotterdam. Participants received course credits as a show-up fee and were tested in groups (group size ranged from 16 to 32 in multiples of 4). After protocol-based exclusions, our sample for the analysis consisted of 169 students (time pressure n = 87; forced delay n = 82). For English speaking students, we used the provided English Qualtrics scripts. For the Dutch students, we used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into Dutch, and (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€. In all other respects, we followed the official protocol.

Contributing Labs
(Alphabetical by last name of first author)

Balazs Aczel, Eotvos Lorand University
Bence Palfi, Eotvos Lorand University
Barnabas Szaszi, Eotvos Lorand University
Aba Szollosi, Eotvos Lorand University
OSF page: https://osf.io/f6jtm/

A total of 204 students (time pressure n = 102; forced delay n = 102) were recruited from the psychology subject pool at the Eotvos Lorand University, Budapest, Hungary. Participants received course credit as show-up fee and were tested in groups (group size of 12 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into Hungarian, and (b) participants made their contributions in Hungarian Forint with a maximum contribution for each participant of 550 HUF. Although we used the preregistered instruction, which specified the maximum possible contribution as 1,100 HUF, we decreased it to 550 HUF for the game to reflect the local economic circumstances. In all other respects, we followed the official protocol.

Thorsten G. H. Chmura, University of Nottingham
Roberto Hernan-Gonzalez, University of Nottingham
OSF page: https://osf.io/h9gxm/

A total of 192 students (time pressure n = 96; forced delay n = 96) were recruited from the CRIBS and CEDEX subject pool at the University of Nottingham. Participants were paid a show-up fee of £2.10 and were tested in groups (group size ranged from 35 to 40). We used the provided Qualtrics scripts with one change: Participants made their contributions in dollars and their final earnings were paid in GBP using the following exchange rate of $1.00 = £0.70. In all other respects, we followed the official protocol. Participants were recruited by offering a show-up fee of $3 = £2.10.

Antonio M. Espín, Middlesex University
Pablo Brañas-Garza, Middlesex University
Praveen Kujal, Middlesex University.
OSF page: https://osf.io/3auwr/

A total of 161 students (time pressure n = 79; forced delay n = 82) were recruited from the subject pool at Middlesex University London. Participants were paid a show-up fee of £5 and were tested in groups (group size ranged from 12 to 20 in multiples of 4). We used the provided Qualtrics scripts with one change: Participants made their contributions in GBP, instead of USD, with a maximum contribution for each participant of £4. In all other respects, we followed the official protocol.

Anthony M. Evans, Tilburg University
Anna E. van 't Veer, Leiden University
OSF page: https://osf.io/c765h/

A total of 152 students (time pressure n = 76; forced delay n = 76) were recruited from the psychology subject pool at Tilburg University. Participants received 30 min of participation credit and were tested in groups (group size ranged from 8 to 12 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into Dutch, and (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€. In all other respects, we followed the official protocol.

Fernando Ferreira-Santos, University of Porto
Tiago O. Paiva, University of Porto
Eva C. Martins, Maia University Institute ISMAI/CPUP
Carlos Mauro, Catholic University of Portugal
Fernando Barbosa, University of Porto
OSF page: https://osf.io/z6jsu/

A total of 171 students (time pressure n = 85; forced delay n = 86) were recruited from the student body of the University of Porto, Maia University Institute–ISMAI, and the Catholic University of Portugal, Porto. Participants were paid a show-up fee of 2.50€ and were tested in groups (group size ranged from 8 to 12 in multiples of 4). We used the provided Qualtrics scripts with three changes: (a) Our study materials were translated into Portuguese, (b) participants made their contributions in Euros with a maximum contribution for each participant of 2.00€, and (c) before the beginning of the session, one of the researchers

entered the computer ID into the Qualtrics survey (because all computers share one external Internet IP, making it impossible to identify individual entries in Qualtrics); participants did not see this question. In all other respects, we followed the official protocol.

Susann Fiedler, Max Planck Institute for Research on Collective Goods
Rima-Maria Rahal, Max Planck Institute for Research on Collective Goods
Minou Ghaffari, Max Planck Institute for Research on Collective Goods
OSF page: https://osf.io/hsdf3/
A total of 196 students (time pressure n = 99; forced delay n = 97) were recruited from the subject pool of the Max Planck DecisionLab. Participants were paid a show-up fee of 5€ and were tested in groups (group size ranged from 8 to 12 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into German, and (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€. In all other respects, we followed the official protocol. After the exclusion of students who were not native speakers of German (n = 18), those who did not move the slider to provide a contribution (n = 5), those who were younger than 18 (n = 5) or older than 35 (n = 2), and students of economics and psychology (n = 12), our total sample for the analysis consisted of 154 participants (time pressure n = 79, forced delay n = 75). Data from 2 additional participants in the forced-delay condition were not included in the analysis due to a coding error that removed their age from the data file. The missing information was only recovered after the data for the RRR had been finalized. Their data are provided on OSF.

Jennifer S. Trueblood, Vanderbilt University
Lisa Guo, University of California, Irvine
OSF page: https://osf.io/3km2q/
A total of 156 students (time pressure n = 78; forced delay n = 78) were recruited from the Department of Psychology human subject pool at Vanderbilt. Participants were paid a show-up fee of $5 USD and were tested in groups (group size ranged from 8 to 24 in multiples of 4). We used the provided Qualtrics scripts without changes. The lab we used (Wilson Hall 120) was an open computer lab without dividers between computers (see photo on OSF). However, the computers were spaced far apart, and we do not think participants felt observed by other participants or the experimenter. The lab could accommodate up to 30 participants in one sitting. Although our preregistered plan specified that we would recruit at least 160 participants, we were unable to recruit enough people to meet our target sample size before the end of the academic semester, ending with a total of 156 participants.

Oliver P. Hauser, Harvard University
OSF page: https://osf.io/5hza7/
A total of 166 students (time pressure n = 84; forced delay n = 82) were recruited from the Harvard Decision Science Laboratory subject pool at Harvard University. Participants were tested in groups (group size ranged from 12 to 36 in multiples of 4). The provided Qualtrics scripts were used with one change: Before the beginning of the session, a research assistant entered the computer ID (to help with distributing payments) into the Qualtrics survey; participants did not see this question, and the remaining part of the study followed the official protocol. Although our preregistered plan specified that participants would receive a show-up fee of $5, we were unable to recruit enough people to meet our target sample size with that method, so 110 participants were recruited by offering a show-up fee of $8.

Tei Laine, Université Grenoble Alpes, France
Laurent Bègue, Université Grenoble Alpes, France
Anthony Herrero, Université Grenoble Alpes, France
OSF page: https://osf.io/2z4rg/
A total of 223 students (time pressure n = 109; forced delay n = 114) were recruited at Université Grenoble Alpes campus. Participants were paid a show-up fee of 5€ and were tested in groups (group size ranged from 8 to 12 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into French and (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€. In all other respects, we followed the official protocol.

Johannes Lohse, University of Birmingham
Timo Goeschl, University of Heidelberg
OSF page: https://osf.io/6xdzp/
A total of 163 students (time pressure n = 81; forced delay n = 82) were recruited from the general subject pool of volunteers at the University of Heidelberg "AWI Lab" using Hroot (Bock et al., 2012). Participants were paid a show-up fee of 3€ and were tested in groups. Group size was either 12 or 16 participants, apart from one session in which we had to run with 8 participants due to no-shows. We used the Qualtrics script provided, but with three changes: (a) Our study materials were translated into German in accordance with the other German labs, (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€, and (c) at the end of the experiment participants entered a personal code that was used to ensure anonymous payment. In all other respects, we followed the official protocol.

Dorothee Mischkowski, University of Hagen
Andreas Glöckner, University of Hagen and Max Planck Institute for Research on Collective Goods
OSF page: https://osf.io/3mwta/

We collected data from a total of 212 participants, from which n = 188 fulfilled the inclusion criteria (complete participation, not missing the contribution variable, below the age limit of 34 years) that were included in the reported analyses (time pressure n = 97; forced delay n = 91). Participants were recruited from the social psychology subject pool at University of Göttingen, Germany. Participants were paid a show-up fee of 5€ and were tested in groups (group size ranged from 8 to 12 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into German, and (b) participants made their contributions in Euros with a maximum contribution for each participant of 4€. In all other respects, we followed the official protocol. For a different project, after the public goods game, we measured social value orientation (SVO) using the 15 item SVO Slider Measure (Murphy, Ackermann, & Handgraaf, 2011), which was not included in the original study. Thereby SVO was preregistered and tested as potential moderator of the spontaneous cooperation effect.

Tess M.S. Neal, Arizona State University
Megan Warner, Arizona State University
OSF page: https://osf.io/bkmd7/
A total of 170 students were recruited from the undergraduate student subject pool at Arizona State University. After we applied the exclusion criteria, 165 students comprised the final sample (time pressure n = 81; forced delay n = 79; missing n = 5). Most of the students were psychology majors (n = 126), but non-psychology majors were also allowed to participate (n = 39). Psychology majors were provided with 2 research credits in our psychology subject pool in lieu of a monetary show-up fee, and the non-psychology majors were provided a $5 show-up fee in lieu of the research credits. Participants were tested in groups, with groups ranging in size from 4 to 16 (M = 9.38). As described in the "Differences From the Official Protocol" section of our lab's Open Science Framework (OSF) implementation page, we decided to run in groups that sometimes were not multiples of 4 given the unique constraints of data collection on our campus (i.e., a small subject pool on the ASU West Campus and a tight timeline for data collection). We altered the formula for calculating payments to correspond with the number of people in the smaller groups in each session (further description on our OSF page). We used the provided Qualtrics script with one change: We created an additional question on the first page that asked "What is your computer number?" into which we could indicate the number affixed to the computer rather than the IP address as per the official protocol. This change enabled us to track an individual participant through the study and assign them to groups within the session in order to calculate payouts. In all other respects, we followed the official protocol.

Julie Novakova, Charles University, Prague
Petr Houdek, University of Economics, Prague
Jaroslav Flegr, Charles University, Prague
OSF page: https://osf.io/a56y4/
A total of 203 students were recruited from the subject pool at CEBEX Laboratory (belonging to the CEVRO Institute; however, the study was conducted by the Charles University). One participant in the time-pressure condition did not meet the age inclusion requirements, leaving 101 participants in each condition. Participants were paid a show-up fee of 50 CZK (or credits in case of students of classes taught at the Charles University in Prague and the University of Economics in Prague) and were tested in groups (group size ranged from 8 to 16 in multiples of 4). We used the provided Qualtrics scripts with the following changes: (a) Our study materials were translated into Czech, (b) participants made their contributions in Czech crowns (CZK) with a maximum contribution for each participant of 65 CZK, and (c) as all of the computers in our lab had a shared IP address, we added a question asking the computer's number (which was visible next to each computer) so that we could use these instead of IP addresses in paying the participants their rewards. In one instance, a student had to leave just after the experiment had commenced. It did not disturb the other 15 participants, so we decided to continue the session and therefore had an odd number of participants in the group that one time. In all other respects, we followed the official protocol.

Roger Pagà, Pompeu Fabra University
Gert Cornelissen, Pompeu Fabra University
Daniel Navarro-Martínez, Pompeu Fabra University
OSF page: https://osf.io/9dpjy/
A total of 157 students (time pressure n = 79; forced delay n = 78) were recruited from the subject pool of the Behavioral Sciences Laboratory (BESLab) at Pompeu Fabra University. Participants were paid a show-up fee of 3.50€ and were tested in groups (group size ranged from 12 to 28 in multiples of 4). We used the provided Qualtrics script with four changes: (a) Our study materials were translated into Spanish, (b) participants made their contributions in Euros with a maximum contribution for each participant of 2.50€, (c) two of the questions that assess participants' understanding of the public goods game and that specifically ask participants which contributions would result in the maximum group and individual gains used an 11-point scale instead of a 9-point scale, and (d) participants were asked to type the ID of the computer that they used to perform the study. In all other respects, we followed the official protocol. Neither the use of a longer scale in the two comprehension questions mentioned above nor the inclusion of the computer ID question were anticipated in our preregistered plan; the longer scale was a side-effect of using Euros instead of USD; the closest approximation to a 9-point dollar scale from 0 cents to 400 cents in increments of 50 cents was an 11-point Euro scale ranging from 0 cents to 250 cents in increments of 25 cents. The question asking the ID of participants' computers had to be added because the Qualtrics script failed to detect the individual IP addresses of the computers used in the study. The

computer IP addresses were necessary to determine the final payments for each participant. As the IP addresses could not be obtained, we were forced to replace them with an alternative identifier: the computer ID variable.

Marco Piovesan, University of Copenhagen
Felix S. Døssing, University of Copenhagen
Erik Wengström, Lund University
Karoline Ø. Thor, University of Copenhagen
OSF page: https://osf.io/b4ra6/
A total of 227 participants (time pressure n = 113; forced delay n = 114) were recruited. Because we used a different currency (DKK), we changed the amount of possible answers in the two comprehension questions from nine to seven. We did this because it made sense to have seven possible answers going from 0 to 3000 øre (Danish cent) with increments of 500 øre. This meant that the correct value in the first question was "7" rather than "9," and this had to be corrected for in the data analysis scripts.

Erika Salomon, University of Illinois
Zachary Horne, University of Illinois
OSF page: https://osf.io/6j4rb/
A total of 202 participants (time pressure n = 101; forced delay n = 101) were recruited from the paid subject pool at University of Illinois. Participants were paid a show-up fee of $4.00 USD and were tested in groups (group size ranged from 8 to 20 in multiples of 4). The replication deviated from the protocol in the following ways. The University of Illinois Paid Subject Pool does not use study descriptions in its recruitment ads. Participants only saw a study number, duration, location, and minimum compensation ($5.00 USD) before arriving. Therefore, the recruitment ad did not describe the study as a "study of decision making." In all other respects, we followed the official protocol. In one testing session, due to a counting error, 14 rather than 16 participants were run at once. This was discovered at the conclusion of the session, and the participants in the final group were paid as if they had been in a four person group where the two missing participants had each contributed all of their money. As this error was not noticed until the end of the testing session, these participants are included in the analyses as required in the protocol.

Narayanan Srinivasan, University of Allahabad
Ajita Srivastava, University of Allahabad
Sumitava Mukherjee, Indian Institute of Management Ahmedabad
OSF page: https://osf.io/pfzkb/
A total of 204 students (time pressure n = 103; forced delay n = 101) were recruited from University of Allahabad and Motilal Nehru National Institute of Technology, Allahabad. Participants were paid a show-up fee of 50 INR and were tested in groups 8, 12, or 16. We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into Hindi, and (b) participants made their contributions in Rupees with a maximum contribution for each participant of 80 INR. In all other respects, we followed the official protocol. Nine participants were excluded from the final analysis because they did not enter a monetary contribution, resulting in 195 participants (time pressure n = 98; forced delay n = 97). After removing participants based on age, the final sample consisted of 177 participants (time pressure n = 88, forced delay n = 89).

Gustav Tinghög, Linköping University
Lina Koppel, Linköping University
Magnus Johannesson, Stockholm School of Economics
Daniel Västfjäll, Linköping University & Decision Research, Eugene, OR
OSF page: https://osf.io/6qn3m/
A total of 168 students were recruited from the subject pool at Linköping University. After protocol-based exclusions, our sample for the analysis consisted of 164 students (time pressure n = 83; forced delay n = 81). Participants were paid a show-up fee of 50 SEK and were tested in groups (group size ranged from 12 to 16 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) Our study materials were translated into Swedish, and (b) participants made their contributions in SEK with a maximum contribution for each participant of 40 SEK. In all other respects, we followed the official protocol. Our study did not deviate from the preregistered plan.

Julian Wills, New York University
Jay J. Van Bavel, New York University
OSF page: https://osf.io/p9s2d/
A total of 148 students (time pressure n = 76; forced delay n = 72) were recruited from the Center of Experimental Social Science subject pool at New York University. Participants were paid a show-up fee of $5 USD and were tested in groups (group size ranged from 8 to 32 in multiples of 4). We used the provided Qualtrics scripts with two changes: (a) We added two questions that asked "What data collection session is this?" and "What computer station is this?"—these questions were answered by the experimenters before the participants arrived to the study, and (b) after completing all the measures in the protocol, participants then completed four additional questionnaires. Our protocol also differed in four additional ways. First, if the number of participants who arrived were not divisible by four, then any remainder participants were assigned to computer stations at the back of the room to complete a separate task of similar length and pay as the one described in the protocol. Second, 2 additional participants did not enter a contribution so their data were excluded before any analyses. Third, 3 participants were excluded because it could not be determined that they met the age requirements: (a) One participant reported that he or she was born in 1933; (b) one reported he or she was born in 1998; and (c) one completed the procedure at a much slower pace than the rest of the sample, so the experimenters

had to terminate the session before he/she could provide the year they were born. Finally, in one session, the experimenters erroneously tested 14 participants. The presence of a 14th participant could potentially undermine the instructions about payments, thereby affecting performance in the task. Consequently, we have excluded all data from that session (at the recommendation of the Editor who was blind to any results). Data from these additional participants are posted on our OSF page. Otherwise our procedure was exactly as stated in the protocol.

Conny E. Wollbrant, University of Gothenburg, and NTNU Business School, Norway
Kristian Ove R. Myrseth, Trinity College Dublin
OSF page: https://osf.io/cynbz/
A total of 156 students (time pressure $n = 79$; forced delay $n = 77$) were recruited from the subject pool at the University of Gothenburg. Participants were paid a show-up fee of SEK 50 and were tested in groups of 8. We used the Qualtrics scripts provided, but with two changes: (a) Our study materials were translated into Swedish, and (b) participants made their contributions in SEK, with a maximum contribution for each participant of SEK 40. In all other respects, we followed the official protocol.

Acknowledgments

Thanks to the American Psychological Society (APS) and the Arnold Foundation who provided funding to participating laboratories to defray the costs of running the study. Thanks to David Rand for providing materials and helping to ensure the accuracy of the protocol. Thanks to Edison Choe for coding the analysis scripts and to Courtney Soderberg at the Center for Open Science for verifying their accuracy.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. We found the "you" versus "your" typographical error only after some labs had begun data collection. We decided not to correct the error at that stage because we did not want to change the procedures and we felt it would not be confusing.
2. Participants from one lab (Srinivasan) had substantially longer response times in the time-pressure condition than did those in all other labs, leading to more exclusions due to noncompliance. On the OSF page, we provide the same analyses excluding results from that one lab. The overall pattern of results does not change.
3. The result reported in the original paper excluded noncompliant participants and found a 15.31 percentage point difference in the amount contributed between the time-pressure ($M = 58\%$) and forced-delay ($M = 42\%$) conditions.

References

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

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*, *144*, 951–966.

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6*, Article 7455. doi:10.1038/ncomms8455

Lohse, J. (2016). Smart or selfish—When smart guys finish nice. *Journal of Behavioral and Experimental Economics*, *64*, 28–40.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation (SVO). *Judgment and Decision Making, 6,* 771–781.

Myrseth, K. O. R., & Wollbrant, C. (2016). Models inconsistent with altruism cannot explain the evolution of human cooperation. *Proceedings of the National Academy of Sciences, USA*, *113*, E2472.

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. doi:10.1177/0956797616654455

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

Rand, D. G., Greene, J. D., & Nowak, M. A. (2013). Rand et al. reply. *Nature*, *498*, E2–E3.

Rand, D. G., & Kraft-Todd, G. T. (2014). Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience*, *8*, Article 300. doi:10.3389/fnbeh.2014.00300

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

Rubin, M., Paolini, S., & Crisp, R. J. (2010). A processing fluency explanation of bias against migrants. *Journal of Experimental Social Psychology*, *46*, 21–28. doi:10.1016/j.jesp.2009.09.006

Singelis, T. M., Triandis, H. C., Bhawuk, D. P., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-cultural Research*, *29*, 240–275.

Strømland, E., Tjøtta, S., & Torsvik, G. (2016). *Cooperating, fast and slow: Testing the social heuristics hypothesis*. Retrieved from http://ssrn.com/abstract=2780877

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). doi:10.1371/journal.pone.0096654