ORIGINAL PAPER

The intuitive cooperation hypothesis revisited: a meta-analytic examination of effect size and between-study heterogeneity

Amanda Kvarven[3] · Eirik Strømland[3] · Conny Wollbrant[5] · David Andersson[1] ·
Magnus Johannesson[2] · Gustav Tinghög[1] · Daniel Västfjäll[1] ·
Kristian Ove R. Myrseth[4]

Abstract

The hypothesis that intuition promotes cooperation has attracted considerable attention. Although key results in this literature have failed to replicate in pre-registered studies, recent meta-analyses report an overall effect of intuition on cooperation. We address the question with a meta-analysis of 82 cooperation experiments, spanning four different types of intuition manipulations—time pressure, cognitive load, depletion, and induction—including 29,315 participants in total. We obtain a positive overall effect of intuition on cooperation, though substantially weaker than that reported in prior meta-analyses, and between studies the effect exhibits a high degree of systematic variation. We find that this overall effect depends exclusively on the inclusion of six experiments featuring emotion-induction manipulations, which prompt participants to rely on emotion over reason when making allocation decisions. Upon excluding from the total data set experiments featuring this class of manipulations, between-study variation in the meta-analysis is reduced substantially—and we observed no statistically discernable effect of intuition on cooperation. Overall, we fail to obtain compelling evidence for the intuitive cooperation hypothesis.

Keywords Cooperation · Dual-process · Intuition · Time pressure · Cognitive load

Amanda Kvarven and Eirik Strømland contributed equally to this work.

✉ Kristian Ove R. Myrseth
    kristian.myrseth@york.ac.uk

Extended author information available on the last page of the article

⚛ Springer

1 Introduction

The Social Heuristics Hypothesis (SHH) stipulates that intuitive decisions drive cooperative behavior and that reflective control overrides a cooperative 'default' behavior to produce selfish decisions (Bear and Rand 2016; Rand et al. 2014). According to the SHH, intuitive decisions tend to rely on experience from games encountered in everyday life, where interactions typically are repeated and involve opportunities for sanctions; deliberation adjusts behavior to the optimal self-interested response in the situation at hand.

The SHH, however, conflicts with suggestions elsewhere in the literature that deliberative processing supports pro-social decision making (e.g., Achtziger et al. 2015; Martinsson et al. 2012; Stevens and Hauser 2004). Moreover, several studies have failed to find a relationship between pro-social behavior and canonical manipulations of cognitive processes (e.g., Hauge et al. 2016; Tinghög et al. 2013, 2016; Verkoeijen and Bouwmeester 2014). This includes a recent registered replication report by Bouwmeester et al. (2017), which sought to replicate the keystone time-pressure study in Rand et al. (2012) but did not find an effect of time pressure on cooperation. Yet, recent meta-analyses present results consistent with an overall positive effect of intuitive decision processes on cooperation (Rand 2016, 2017a, b). In sum, the literature on intuitive cooperation has grown sharply since the publication of the original time-pressure study by Rand et al. (2012)—but without reaching a resolution.

This paper presents an updated meta-analysis to add clarity to the literature. While we obtain an overall meta-analytic effect of the intuition manipulations on cooperation, we can attribute this effect to a specific class of induction manipulations. These manipulations ask participants to rely on emotion over reason in determining their resource allocation (Gärtner et al. 2018; Levine et al. 2018). Thus, we identify a single source of variation in the effect size that may account for inconsistent conclusions in the literature; when we exclude the six experiments that feature this specific manipulation—comprising just 7% of our total data set—we obtain no effect of intuition on cooperation, and the exclusion also yields a substantial reduction in systematic between-study variation. These results are problematic for the SHH as emotion-induction manipulations are vulnerable to alternative interpretations—and the SHH gives no reason for favoring this class of manipulations over others. Moreover, the dramatic dissipation of systematic heterogeneity, following removal of emotion-induction manipulations, runs counter to the idea that the intuitive cooperation effect, if present, is highly heterogeneous (Rand 2016). We also note that our results cannot be explained by between-study variation in participant compliance rates; we find no evidence that studies with higher compliance rates yield systematically higher effect sizes, speaking against the claim in Rand (2017a, 2019) that non-compliance explains why many studies find no effect of intuition manipulations on cooperation.

Our paper proceeds as follows. First, we present our data set and methods, then the analysis, after which we offer concluding remarks on the cognitive foundations of cooperation and the state of the literature.

2 Data and methods

Our inclusion criteria largely follow those in Rand (2016), who presented a meta-analysis to examine the effect of intuitive decision making on cooperation. The inclusion criteria define relevant experimental games and intuition manipulations. To be included in our meta-analysis, a study has to feature a controlled experiment—with monetary incentives and no deception—that used time pressure, cognitive load, ego depletion, or induction to manipulate cooperation.[1] The required intuition manipulations follow Rand (2016), exactly.[2]

As for relevant experimental games, we depart slightly from Rand (2016) by focusing on games that capture cooperation in strategic interactions not contaminated by past or future choices, to ensure clear interpretation of the dependent variable. Therefore, we include only one-shot, simultaneous-move public goods games and prisoner's dilemmas. This differs from Rand (2016), who in addition to simultaneous-move public goods games and prisoner's dilemmas, also included second-player moves in sequential trust games and decisions from the last round of finitely repeated games. Nevertheless, to gauge how inclusion criteria affect our results, we perform robustness checks that also include sequential game decisions. Our final data set comprises 44 of 51 experiments included in the prior meta-analysis by Rand (2016), as most of his studies fit our inclusion criteria. In addition, we include 36 new experiments featuring 13,189 participants, an increase of 56.9% in the number of studies and an increase of 83.5% in the number of participants.[3] Table A.2 in Supplemental Online Material (SOM) A provides a full overview of the experiments comprising our data set, including the number of participants and details about game type and manipulation used.[4]

---

[1] Time-pressure studies typically ask participants either to decide within 10 s (time pressure) or to spend more than ten seconds on the decision (time delay). Studies with cognitive load ask participants to engage in some alternative task, while simultaneously engaging with the main decision problem. The idea is that participants will have less cognitive resources available for the main task and will hence rely more on their intuition when making a decision. Studies with depletion manipulations present an ancillary task prior to the main decision, to tire ('deplete') participants in advance of the main decision. Finally, induction manipulations ask participants to use their intuition while deciding, to rely on emotions or their 'heart' (emotion induction), to read the instructions in a foreign language or, prior to the decision, to recall a time in their lives when intuitive decisions paid off (recall induction).

[2] See (Myrseth et al. 2017) for a methodological critique of time-pressure experiments purporting to show evidence for 'intuitive cooperation.' This meta-analysis does not necessarily endorse the manipulations included; rather, it takes them on faith in an effort to address patterns discussed in the literature.

[3] Twenty-one of the experiments in our updated study originate from a recently registered replication report (Bouwmeester et al. 2017). Seven of these use time pressure, including: Strømland, et al. (2016), Everett et al. (2017), Isler et al. (2018), Capraro and Cococcioni (2015), Alós-Ferrer and Garagnani (2018), and Bird et al. (2018). We also added Madland (2017), Gärtner et al. (2018), Rand (2018), and Tinghög (2018), all of which are unpublished studies manipulating deliberation through a form of induction. Two field experiments (Artavia-Mora et al. 2017, 2018) do not satisfy the requirement that the study involve monetary incentives and were thus not included in our meta-analysis.

[4] We used Google Scholar to search for the following keywords in all possible combinations: "prisoner's dilemma" or "public goods game" combined with "cognitive load," "time pressure," "ego depletion," "intuition priming," "intuition recall," or "intuition conceptual priming." We also manually searched through articles citing (i) the original study by Rand et al. (2012) and (ii) the meta-analysis by Rand (2016).

Springer

Our inclusion decisions depart from Rand (2016) in two additional respects. First, our main analysis includes studies that provided in the experimental instructions information about time pressure. Rand (2016) argues that this introduces a potential comprehension confound—however, such challenges are inherent to these kinds of experiments regardless of when one introduces information about time pressure. Moreover, most of the data using this variation of the time-pressure manipulation originate from Tinghög et al. (2013), who successfully solved the issue of compliance plaguing other studies (e.g., Bouwmeester et al. 2017; Rand et al. 2012). For these reasons, we do not see adequate justification to exclude studies that, in the experimental instructions, inform participants about time pressure.

Second, all of our analyses include participants who did not comply with the experimental treatment, as excluding them would lead to selection bias. The meaning of 'compliance' depends on the specific manipulation type, and the compliance rate varies by type. Compliance is mostly an issue for the time-pressure manipulation (where non-compliance means not responding according to the time constraint) and induction manipulations (where non-compliance means that one has failed to follow instructions to write down something in an open field). Table A.5, in SOM A, displays compliance rates by manipulation type.

In his discussion of time-pressure experiments, Rand (2017a) argues that excluding non-compliers provides an improved picture of the effect and that such exclusion is justifiable due to the absence of correlation between observable factors and compliance with the time constraint. However, a re-analysis of Rand et al. (2014), Table A.1 in SOM A, shows that compliant participants are a selected subgroup—consistent with the argument that compliant-only analyses suffer from selection bias (Bouwmeester et al. 2017; Tinghög et al. 2013). Moreover, regardless of the outcome of balance tests, participants could self-select based on factors unobservable to the researcher. For this reason, we include non-compliers, and so all results must be interpreted as an 'intention-to-treat' effect. Still, the number of studies and participants featured in our meta-analysis allows for high statistical power to detect very small hypothesized population effect sizes (see SOM B for a detailed power analysis).[5]

We subject our data set to a random-effects meta-analysis, which allows for systematic variation between studies by assuming that each true effect is drawn from a normal population distribution, with a common mean and between-study variance (Higgins et al. 2009).[6] This modeling assumption seems reasonable a priori, as several papers argue that the effect is heterogeneous (Mischkowski and Glöckner 2016; Rand 2018; Rand et al. 2014; Strømland et al. 2016). In line with Rand (2016), we use as our dependent variable percentage contributed of the total endowment, ensuring that our results are directly comparable to those in the previous

---

[5] See Sect. 3.2 for further discussion on the implications of the compliance issue for our analysis and main conclusions.

[6] The normality assumption is only necessary for testing the null hypothesis of no mean effect in the random-effects distribution of true effects. For purposes of estimation, no distributional assumptions are required, and to construct valid confidence intervals, one may rely on asymptotic normality in the number of studies (Higgins et al. 2009).

meta-analysis. For decision problems with binary choice, such as the conventional prisoner's dilemma, the dependent variable takes the value 100 if the participant cooperates, and 0 otherwise.

Analytically, our study differs from Rand (2016) in that we pay particular attention to sources of heterogeneity—systematic inconsistency across experiments. When there is large systematic inconsistency across experiments, it is hard to interpret the weighted summary effect produced by a meta-analysis.

In our meta-analysis, each effect size is computed as the percentage point difference between the treatment (intuition condition) and control group (deliberation condition). This means that the effect-size measure is bounded between 0 and 100. For studies retrieved from Rand (2016), we use reported effect sizes and standard errors, directly. For Bouwmeester et al. (2017), we follow the same procedure and retrieve the standard errors from the data reported. For other studies not included in either of the aforementioned data sets, we retrieved the data from regression tables where the percentage point difference between the treatment group and the control group was reported, and we normalized the effect size to a scale ranging between 0 and 100. For studies where this was not possible (e.g., if the main analysis conditioned on participants' compliance status and the intention-to-treat effect was not reported), we downloaded the data and ran linear regressions of the normalized contribution rate on a dummy indicator for the intuition condition, using the estimated coefficient as a measure of the treatment effect (this estimator is equivalent to a simple mean difference between the intuition condition and the deliberation condition). We use robust standard errors in the regression and construct 95% confidence intervals (effect size ± 1.96SE, where SE is the standard error for the regression coefficient).

3 Results

We start by considering all experiments that meet our inclusion criteria. Figure 1 displays a forest plot of all experiments, including the overall effect with a corresponding 95% confidence interval. To the right of each estimate, we provide design details for the associated experiment.

As Fig. 1 shows, the magnitude of the overall effect of intuition manipulations on cooperation is 2.19 percentage points, and this effect is statistically significant ($p = 0.005$, Z test). However, the magnitude of the overall effect is only 35.7% of the main effect reported in a prior meta-analysis that excludes non-compliers (Rand 2016) and only 52.1% the size of the intention-to-treat effect reported in that meta-analysis. This reduction in effect size may reflect the addition of individual lab estimates featured in the large registered replication study by Bouwmeester et al. (2017), which finds no effect of time constraints on cooperation. This pattern, in turn, is consistent with the 'decline effect' (e.g., Fanelli et al. 2017), whereby the influence in a meta-analysis of publication bias in an initial study dissipates with the number of replication studies added with no effect.

The overall effect may nevertheless not capture a psychologically relevant parameter; we can attribute 62% of the variation in the above forest plot to systematic
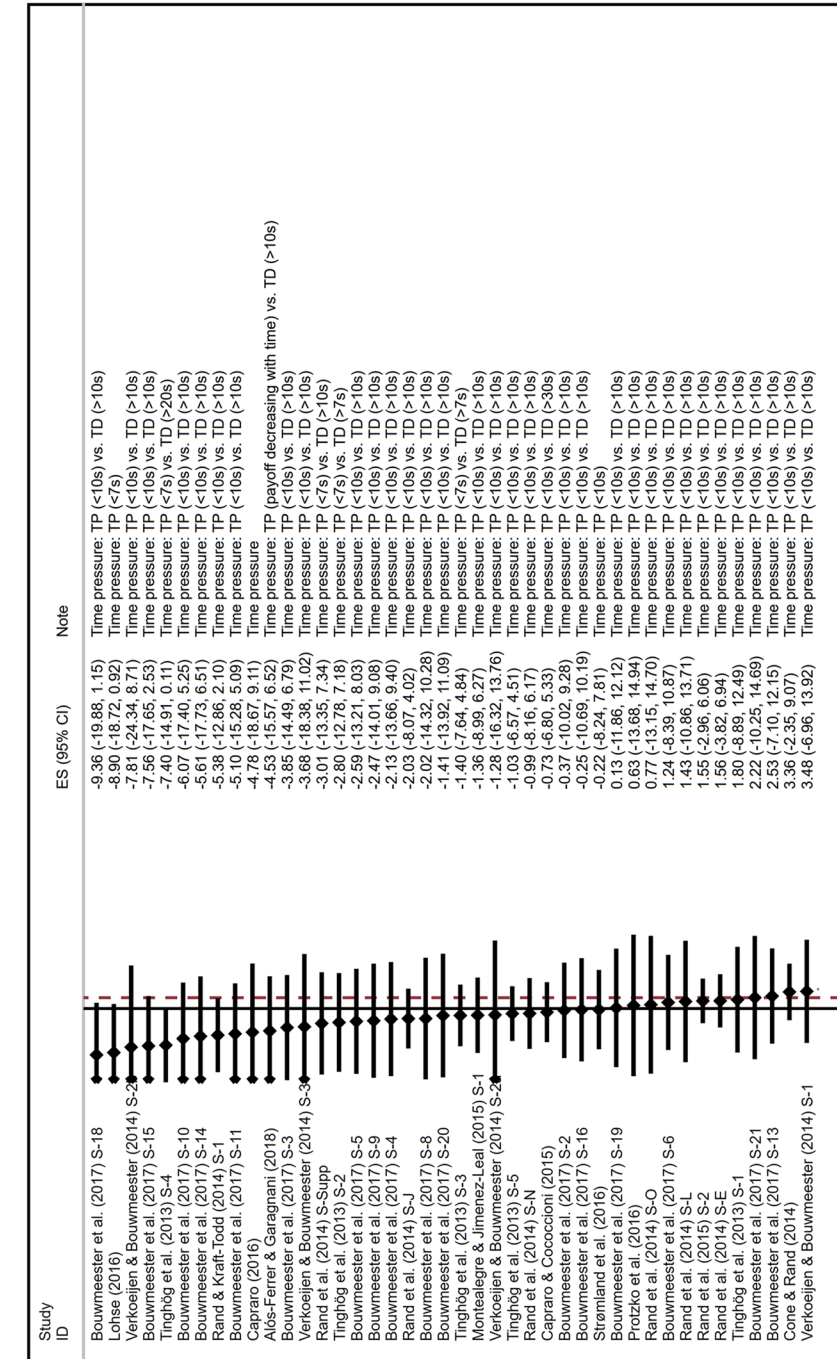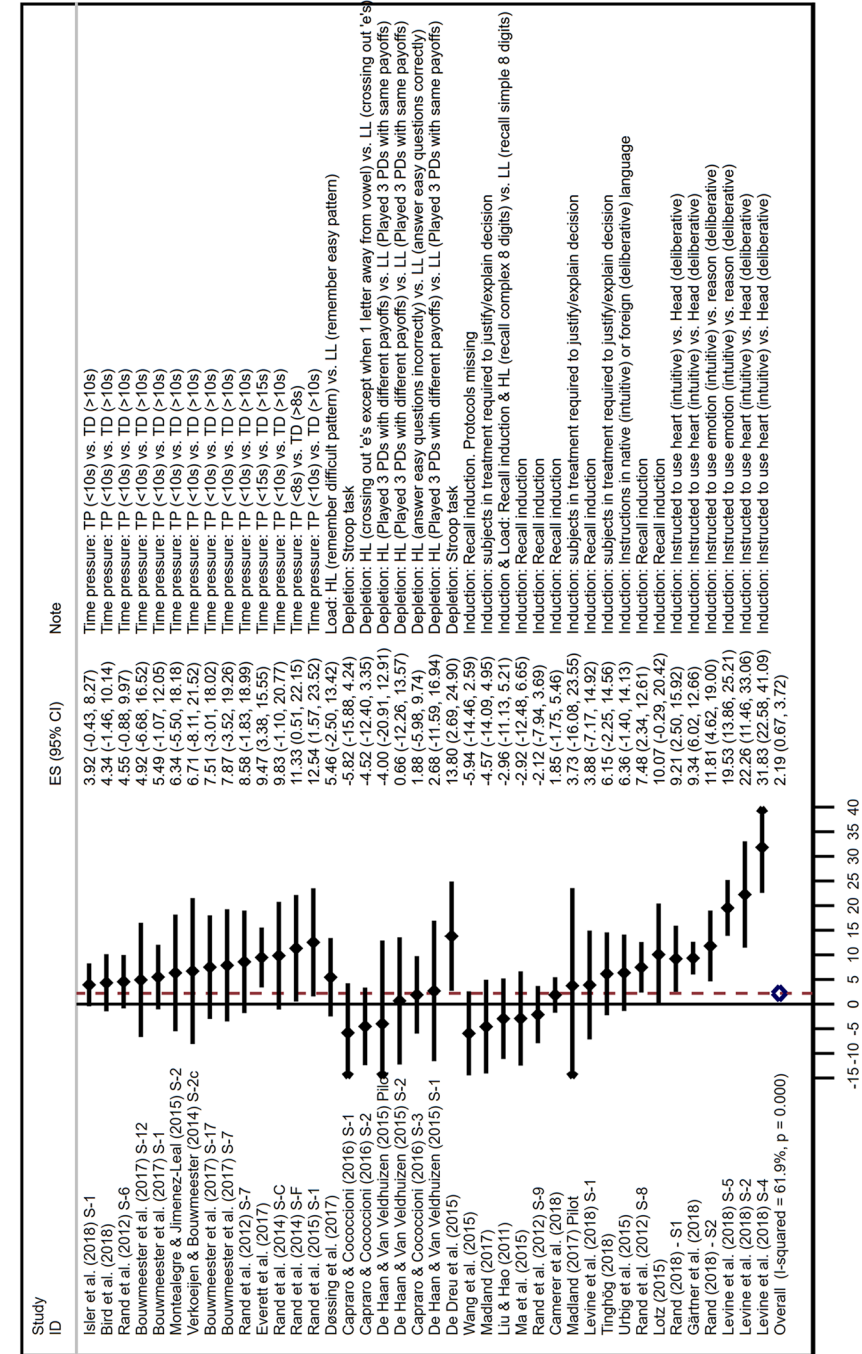
Fig. 1 Forest plot, all experiments

Fig. 1 (continued)

differences between experiments ($I^2 = 61.9\%$, $\chi^2(81) = 212.75$, $p < 0.001$).[7] Moreover, the estimated between-study variance is large ($\hat{\tau}^2 = 27.08$). As an illustration, note that the effect size varies from -9 percentage points to 32. In summary, the analysis suggests an overall positive effect, but the experiments included exhibit very large variation in effect sizes, and that variation may, to a large degree, be attributed to factors other than chance.[8] As an overall effect size provided by a random-effects analysis is insufficient to summarize a heterogeneous set of studies (Raudenbush and Bryk 1985), our summary estimate should be interpreted with caution.

When a meta-analysis suggests large between-study variation, it is common practice to search for the sources of that variation (Higgins et al. 2003a, b). In our case, the observed heterogeneity may have several explanations. One possibility is that the intuitive cooperation effect is contingent on various background factors, as suggested in several papers (Capraro and Cococcioni 2015; Mischkowski and Glöckner 2016; Rand et al. 2014; Strømland et al. 2016), including Rand's (2016) meta-analysis. Another possibility is that various manipulations, here grouped together as 'intuition manipulations', may work in different ways or even capture distinct psychological processes. That is, one may ask whether the observed inconsistency across studies is attributable to genuine and perhaps unpredictable variation in the underlying effect across study populations, or whether it is a by-product of the inclusion criteria. To distinguish between these possibilities, we turn to an analysis that separates manipulation types.

3.1 Comparing manipulations: meta-regressions

We use meta-regressions (see e.g., Thompson and Higgins 2002) to compare the intuitive cooperation effect across manipulation types. We take as a baseline experiments with time pressure, since time pressure is the manipulation type most frequently applied to induce cooperation. In SOM A (see Figures A.1–A.7), we provide meta-analyses specific to each manipulation type. In all individual meta-analyses but one, there is substantially less systematic between-study variation than there is in the overall analysis. The exception is that for induction manipulations (see Figure A.4), where the estimated heterogeneity is 83.1%—which is very high (Higgins et al. 2003a, b); this indicates that nearly all observed variation is attributable to genuine differences in the underlying effect across studies of this type. For this reason, we split induction manipulations into the following subcategories: (i) 'emotion-induction' manipulations instructing participants to rely on emotion over reason when making their choices, (ii) 'recall induction', and (iii) 'other induction' manipulations. The meta-regression results are displayed in Table 1. It is important to note that these regressions capture correlations, as we only have within-study randomization and no exogenous between-study variation.

---

[7] The estimated heterogeneity is similar to that found in Rand's (2016) meta-analysis.

[8] The distinction between I^2 and τ^2 is important. The former measure answers the question, "What proportion of the observed variation between studies is due to factors other than random chance?", while the second measure answers, "How large is the systematic inconsistency across studies?" (Borenstein et al. 2017). Thus, in a sufficiently large sample, I^2 may be large even in the absence of large between-study variation.

Table 1 Meta-regressions of effect size (intuitive cooperation effect) on manipulation type

	(1) Effect size	(2) Effect size	(3) Effect size	(4) Effect size
Depletion	− 0.177	− 15.06***		
	(2.567)	(3.145)		
Cognitive load	0.695	− 14.19***		
	(3.851)	(4.258)		
Recall induction	1.142	− 13.74***		
	(2.043)	(2.734)		
Emotion induction	14.88***		14.81***	
	(2.123)		(2.080)	
Other induction	2.565	− 12.32***		
	(3.148)	(3.635)		
Time pressure		− 14.88***		− 14.81***
		(2.123)		(2.080)
Pooled			0.978	− 13.83***
			(1.461)	(2.300)
Constant	0.619	15.50***	0.630	15.44***
	(0.777)	(1.976)	(0.765)	(1.934)
Observations	82	82	82	82

Standard errors in parentheses. (1) Meta-regressions on manipulation type (baseline: time pressure); (2) meta-regressions on manipulation type (baseline: emotion induction); (3) meta-regressions on manipulation type, all manipulations that are not emotion induction or time pressure pooled together (baseline: time pressure); and (4) meta-regressions on manipulation type, all manipulations that are not emotion induction or time pressure pooled together (baseline: emotion induction). 'Pooled' is a dummy for all manipulations that are not time pressure or emotion induction

*$p < 0.10$, ** $p < 0.05$, ***$p < 0.01$

The meta-regressions yield several noteworthy results. First, Column (1) shows that only experiments using emotion-induction manipulations are significantly more effective in promoting cooperation than are time-pressure studies (coefficient = 14.88 percentage points, $t(76) = 7.01$, $p < 0.001$); the other manipulations are not significantly different from the small and non-significant effect estimated for the time-pressure studies (coefficient = 0.619, $t(76) = 0.80$). It is also noteworthy that 'other induction' manipulations yield an estimated effect very close to that of time-pressure studies, a mere 2.57-percentage point difference ($t(76) = 0.81$). Column (2) takes emotion-induction manipulations as the baseline and shows that all other manipulations are significantly less effective in promoting cooperation. Consistent with this, in Column (4), both time-pressure ($t(79) = -7.12$, $p < 0.001$) and 'pooled' manipulations ($t(79) = -6.01$, $p < 0.001$) are estimated to reduce the effect size by about 14 percentage points relative to emotion-induction manipulations. Together, these results justify our subdivision of the wider class of induction manipulations.

A funnel plot of all studies in the main analysis (see SOM A, Fig. A.11) illustrates the relative effectiveness of manipulation types; five out of six experiments

using the emotion-induction manipulations appear as outliers, to the right of the 95%-confidence bar.

While Rand (2016) suggests that time-pressure manipulations are less effective than are other manipulations, our results indicate that only emotion-induction manipulations differ in their effect from other manipulations. We therefore proceed to test whether our overall meta-analytic effect depends on the emotion-induction manipulations, specifically; we conduct an alternative meta-analysis that includes all studies other than the six experiments using emotion-induction manipulations. This meta-analysis (see Fig. A.10) reveals no discernable overall effect on cooperation; the estimated meta-analytic effect is 1 percentage point ($p = 0.076$, Z test), and, judged by conventional classifications (Higgins et al. 2003a, b), heterogeneity is also quite low ($I^2(74) = 19.8\%$, $\chi^2(75) = 93.50$, $p = 0.073$, $\hat{\tau}^2 = 4.43$). Because time-pressure studies have been called into question, both for the size of their effect (Rand 2016) and their validity (Myrseth and Wollbrant 2017), we run a meta-analysis that excludes all emotion-induction and time-pressure manipulations, evaluating all other manipulations in the same test (Fig. A.9). In this meta-analysis, the estimated effect of the intuition manipulations is 1.62 percentage points—only 26.4% of the main effect reported in Rand (2016) and only 38.6% of that study's intention-to-treat estimate—and not significantly different from zero ($p = 0.177$, Z test).

To ensure that our conclusions are not sensitive to inclusion criteria, we undertake additional robustness checks, using various combinations of Rand's (2016) inclusion criteria while excluding the emotion-induction studies. In all tests, we follow Rand and include data on second movers and last-round moves in finitely repeated games. We also undertake robustness tests where we include data on trust game decisions, and tests where we include second-mover decisions only where the first mover contributed the maximum amount possible (as in Rand 2016). We carry out these robustness checks both for the specification excluding emotion-induction and time-pressure studies (Fig. A.9) and for the specification excluding only the emotion-induction studies (Fig. A.10). None of these robustness checks reveal a statistically significant overall effect; the estimated effect is consistently very small and insensitive to the inclusion criteria (see Table A.3 for details). Finally, it is worth noting that a separate meta-analysis of pre-registered studies (Bouwmeester et al. 2017; Camerer et al. 2018; Everett et al. 2017), only, leads to a similar conclusion; the effect size in this meta-analysis is just 0.79 percentage points and not statistically significant, and the estimated heterogeneity is low (see Fig. A.12).

A possible interpretation of our null result is that the 'true' effect size is very small, and that our result, when excluding emotion-induction manipulations, is a false negative. However, this interpretation would prove equally challenging to existing studies that report evidence for intuitive cooperation. Suppose that our upper bound on the effect size—1.8 percentage points in these eight specifications—represents the true effect size. Then, for a single study to have 80% power to detect the underlying effect, one would need a sample size of at least 15,486 participants (assuming a common standard deviation of 40 between treatment groups). Should the effect size instead be 1 percentage point, as in Fig A.10—which also corresponds closely to the effect size obtained using only pre-registered studies (see Fig. A.12)—one would need a sample size of at least 50,176 participants for a single

⚛ Springer

study to achieve 80% power. Thus, even if our main finding were a false negative, the mean effect size in this literature is so small that to meaningfully study it one would need sample sizes an order of magnitude larger than those typically used in experimental studies. Any statistically 'positive' finding in this literature, obtained in typical sample sizes, would therefore likely represent a major overestimate (Gelman and Carlin 2014).

3.2 Alternative explanations

Rand (2019) responds to a pre-print version of our analysis by undertaking his own updated meta-analysis. He uses a combination of the data from Rand (2016) and those from our paper. His main argument is that our choice to exclude sequential games from the main analysis is responsible for the null effect obtained when we exclude emotion-induction manipulations. However, this cannot be the reason for the discrepancies between his new findings and ours—Table A.3 in our supplementary materials shows that our results are insensitive to the differences in inclusion criteria between Rand (2016) and our study.

Rand (2019) argues further that poor experimental designs may account for why there are many null findings in the literature. He suggests that future studies should move towards experimental designs that increase the compliance rate and comprehension of the game, and he expects these design features to be associated with substantially larger treatment effects. Related to the latter point, we note that the registered replication report by Bouwmeester et al. (2017) undertook a high-powered test of the hypothesis that comprehension moderates the time-pressure effect; they found no time-pressure effect in the comprehending subgroup. As for the hypothesis that greater compliance is associated with greater effect size, we are not aware of prior tests in the literature, so we undertake it here. Because compliance varies between manipulation types, we also undertake a separate test for studies using time-pressure manipulations. Figure 2 presents a scatter plot of compliance rate and effect size, for all studies included in our meta-analysis.

As Fig. 2 shows, there is no obvious relationship between the compliance rate of the study and the observed effect size, neither in general for all manipulations nor specifically for the time-pressure manipulations. The correlation is estimated in a meta-regression to be small and positive, but not statistically significant, neither for the full sample nor for the sample of time-pressure studies (regression results in Table A.4). Based on evidence available, therefore, it seems unlikely that a movement towards studies with higher compliance rates will have a major impact on effect sizes in this literature.

An alternative way of addressing the role of study compliance is to run the meta-analysis for compliant participants only (so that the effect size is computed for each study only for participants who complied with the time allotted), as the main analysis in Rand (2016) was conducted. We report such an analysis in Fig. A.13, where we run the meta-analysis for all studies, including compliant participants only. This analysis yields a positive and statistically significant association between the intuition manipulations and cooperation, even when excluding emotion-induction
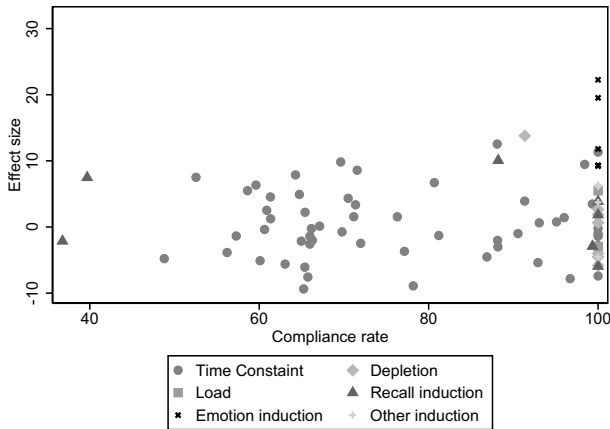
Fig. 2 Study-level compliance rate against observed effect size

manipulations. However, conditioning on compliance status amounts to a 'bad control problem', as the treatment effect conditional on potentially endogenous variables warrants causal interpretation only under quite restrictive assumptions (Montgomery et al. 2018). Specifically, the analysis assumes that compliance, which happens after randomization, does not affect systematically the relative distribution of participants in the treatment versus control groups. This assumption is unmerited, however, as conditioning on compliance may plausibly change the composition in the treatment versus control groups differentially, such that these groups no longer are directly comparable. And, as seen in Table A.1, there is empirical evidence for the selection-bias argument—data sets in this literature indicate that there is self-selection into who complies or fails to comply with the treatment assigned. Finally, we would also note that absence of imbalance would not in itself amount to evidence against the selection-bias argument, as balance tests do not have 100% statistical power—and not all factors imbalanced between treatments are measured. In choosing to include non-compliant participants in our main analysis, we also follow recent meta-analyses in this literature (Fromell et al. 2018; Köbis et al. 2019; Rand 2019).

4 Conclusion

We present an updated meta-analysis of experiments that attempt to manipulate intuitive decision-making processes in games of cooperation. Our analysis tests the Social Heuristics Hypothesis (SHH), which stipulates that intuitive decision-making processes facilitate cooperative behavior. In examining both the overall meta-analytic effect and the origin of the between-study heterogeneity, we fail to obtain robust evidence for the SHH. Although we find evidence in favor of an overall positive effect of intuitive decision processes on cooperation, we can attribute this effect to a particular class of emotion-induction manipulations—those asking participants to rely on emotion over reason when determining allocation. Other manipulation

🍃 Springer

types fail to yield a statistically discernable effect on cooperation. When we exclude the six studies with this manipulation type and conduct a meta-analysis on the remaining 76 studies, which comprise 93% of the observations in our full data set, we find that intuition manipulations have no effect on cooperation.

The consistency in findings across all manipulation types, save the emotion-induction manipulations, suggests that the latter produces a distinct effect. One possibility is that the transparency of the researchers' intention in this setting—asking people to rely on emotion over reason—is understood as a request that participants cooperate, akin to an experimenter demand effect. A request to use your 'heart' could be seen as encouragement to be 'nice', whereas a request to use your 'brain' may indicate that you should try to calculate personal consequences (and not be gullible). The demand effect is less likely to apply to the other intuition manipulations (e.g., time pressure) as the link in those cases, between the researcher's hypothesis of interest and the treatment, is less transparent. While a laboratory participant asked to decide within 10 s might suspect that the study is about the relationship between cooperation and making decisions fast or slow, the direction of the research hypothesis is not evident. Notably, direct requests that signal strongly potential underlying research objectives have been shown to strengthen experimenter demand effects (de Quidt et al. 2018).

An alternative, but perhaps less plausible possibility is that emotion induction is the only class of manipulations that successfully influences intuitive decision making. However, even if this alternative interpretation were true, it is worth noting that the SHH (Rand et al. 2014; Bear and Rand 2016) did not give reason a priori that this manipulation should work, whereas others should not. Related to this, one might wonder whether failure to comply with experimental instructions could account for our results, as compliance varies with study type. However, we do not find evidence for the hypothesis, put forward by Rand (2019), that studies with higher compliance exhibit higher effect sizes.

We also fail to find support for the idea that the underlying effect is highly heterogeneous (Rand 2016), as the removal of emotion-induction experiments from the meta-analysis reduces estimated between-study heterogeneity dramatically. This finding is consistent with the low between-study variation observed in the meta-analysis by Fromell et al. (2018), who study the effect of intuition manipulations on dictator game giving. We cannot rule out the possibility that we are underpowered to detect study-level heterogeneity, but it does appear that the meta-study by Rand (2016) overstates the importance of study-level heterogeneity for the effect of intuition manipulations. Nevertheless, tests for heterogeneity between studies will not necessarily pick up genuine individual-level heterogeneity, if such individual characteristics tend to be similar across study populations, and some studies argue that such individual-level heterogeneity is important for the link between intuition and cooperation (e.g., Alós-Ferrer and Garagnani 2018). One recent study on time-pressure effects in the dictator game tests more directly for such individual-level heterogeneity (across a large set of potentially relevant variables) and finds little evidence for it (Strømland and Torsvik 2019).

As our study focuses on cooperation, we cannot rule out that intuition influences other forms of pro-social behavior. According to Rand et al. (2016), the SHH also predicts intuitive altruism in women, but not men. While their meta-analysis finds support for this prediction, a more recent meta-analysis by Fromell et al. (2018) finds for men a negative effect of intuitive decision processes on altruism and for women no effect.

At a more general level, our findings also speak to the current discussion on heterogeneity in effect sizes in psychology and economics (DellaVigna and Pope 2018; Klein et al. 2014; McShane and Böckenholt 2014; van Aert et al. 2016). Meta-analyses in psychology typically suggest substantial systematic heterogeneity in effect size (Stanley et al. 2018), but the recent 'Many Labs' projects find relatively low systematic variation in effect size across various contexts and cultures (Klein et al. 2014, 2018). Consistent with this, studies by DellaVigna and Pope (2018) indicate that effect sizes tend to be more stable across settings than predicted by expert forecasts. Our meta-analysis is consistent with these findings, and it shows that estimated treatment effect heterogeneity in meta-analyses can be surprisingly sensitive to inclusion criteria; when we include the emotion-induction manipulations, heterogeneity is high—but when we exclude them, heterogeneity is low. Our evidence thus highlights the possibility that some of the heterogeneity reported in meta-analyses arises from researchers' inclusion decisions—as opposed to genuine variation in the effects under scrutiny.

References

Achtziger, A., Alós-Ferrer, C., & Wagner, A. K. (2015). Money, depletion, and prosociality in the dictator game. *Journal of Neuroscience, Psychology, and Economics, 8*(1), 1–14.

Alós-Ferrer, C., & Garagnani, M. (2018). *The cognitive foundations of cooperation*. (Working Paper No. 303). University of Zurich, Zurich, Switzerland. Retrieved from, https://www.econ.uzh.ch/dam/jcr:2e76448c-37da-45fd-a189-fc7f85e4e74a/econwp303.pdf.

Artavia-Mora, L., Bedi, A. S., & Rieger, M. (2017). Intuitive help and punishment in the field. *European Economic Review*, *92*, 133–145.

Artavia-Mora, L., Bedi, A. S., & Rieger, M. (2018). Help, prejudice and headscarves. Retrieved from, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3170249.

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences, 113*(4), 936–941.

Bird, B. M., Geniole, S. N., Procyshyn, T. L., Ortiz, T. L., Carré, J. M., & Watson, N. V. (2018). Effect of exogenous testosterone on cooperation depends on personality and time pressure. *Neuropsychopharmacology*. https://doi.org/10.1038/s41386-018-0220-8.

Borenstein, M., Higgins, J., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5–18.

Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., … & Evans, A. M. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, *12*(3), 527–542.

Capraro, V., & Cococcioni, G. (2015). Social setting, intuition and experience in laboratory experiments interact to shape cooperative decision-making. *Proceedings of the Royal Society B*. https://doi.org/10.1098/rspb.2015.0237.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644.

de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review, 108*(11), 3266–3302.

DellaVigna, S., & Pope, D. (2018). Stability of Experimental results: Forecasts and evidence. Working paper. Retrieved from, https://eml.berkeley.edu/~sdellavi/wp/StabilityDec2018.pdf.

Everett, J. A., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behavior: Even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *Journal of Experimental Social Psychology, 73,* 76–81.

Fanelli, D., Costas, R., & Ioannidis, J. P. (2017). Meta-assessment of bias in science. *Proceedings of the National academy of Sciences of the United States of America, 114*(14), 3714–3719.

Fromell, H., Nosenzo, D., & Owens, T. (2018). *Altruism, fast and slow? Evidence from a meta-analysis and a new experiment* (No. 2018–13).

Gärtner, M., Tinghög, G., & Västfjäll, D. (2018). *Inducing cooperation: Who is affected?* Unpublished manuscript.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651.

Hauge, K. E., Brekke, K. A., Johansson, L. O., Johansson-Stenman, O., & Svedsäter, H. (2016). Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Economics, 19*(3), 562–576.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003a). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557–560.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003b). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557.

Higgins, J. P., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172*(1), 137–159.

Isler, O., Maule, J., & Starmer, C. (2018). Is intuition really cooperative? Improved tests support the social heuristics hypothesis. *PloS one*. https://doi.org/10.1371/journal.pone.0190560.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). *Investigating variation in replicability: a "many labs" replication project. Social Psychology, 45,* 142–152.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., … Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.31234/osf.io/9654g.

Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619851778.

Levine, E. E., Barasch, A., Rand, D., Berman, J. Z., & Small, D. A. (2018). Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General, 147*(5), 702.

Madland, K. R. (2017). Do Cognitive Processes Influence Social Preferences? Testing the social heuristics hypothesis in a sequential prisoner's dilemma (Master's thesis). Bergen, Norway: The University of Bergen. Retrieved from http://bora.uib.no/bitstream/handle/1956/16098/Master-ferdi g.pdf?sequence=1&isAllowed=y.

Martinsson, P., Myrseth, K. O. R., & Wollbrant, C. E. (2012). Reconciling pro-social vs. selfish behavior: On the role of self-control. *Judgment and Decision Making, 7*(3), 304–315.

McShane, B. B., & Böckenholt, U. (2014). You cannot step into the same river twice: When power analyses are optimistic. *Perspectives on Psychological Science, 9*(6), 612–625.

Mischkowski, D., & Glöckner, A. (2016). Spontaneous cooperation for prosocials, but not for proselfs: Social value orientation moderates spontaneous cooperation behavior. *Scientific Reports, 6,* 21555.

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science, 62*(3), 760–775.

Myrseth, K. O. R., Wollbrant, C., & Cognitive, E. (2017). Cognitive foundations of cooperation revisited: Commentary on Rand et al. (2012, 2014). *Journal of Behavioral and Experimental Economics, 69,* 133–138.

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science, 27*(9), 1192–1206.

Rand, D. G. (2017a). Reflections on the time-pressure cooperation registered replication report. *Perspectives on Psychological Science, 12*(3), 543–547.

Rand, D. G. (2017b). Social dilemma cooperation (unlike Dictator Game giving) is intuitive for men as well as women. *Journal of Experimental Social Psychology, 73,* 164–168.

Rand, D. G. (2018). Non-naïvety may reduce the effect of intuition manipulations. *Nature Human Behaviour, 2*(9), 602.

Rand, D. G. (2019). Intuition, deliberation, and cooperation: Further meta-analytic evidence from 91 experiments on pure cooperation. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3390018.

Rand, D. G., Brescoll, V. L., Everett, J. A., Capraro, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General, 145*(4), 389.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., et al. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5,* 3677.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*(2), 75–98.

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin, 144*(12), 1325–1346.

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences, 8*(2), 60–65.

Strømland, E., Tjotta, S., & Torsvik, G. (2016). Cooperating, fast and slow: Testing the social heuristics hypothesis. CESifo Working Paper Series No. 5875. Retrieved from: https://www.cesifo.org/DocDL/cesifo1_wp5875.pdf.

Strømland, E., & Torsvik, G. (2019). Intuitive prosociality: Heterogeneous treatment effects or false positive? Retrieved from, https://osf.io/hrx2y.

Thompson, S. G., & Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*(11), 1559–1573.

Tinghög, G. (2018). Intuition induction in a Prisoner's Dilemma. Unpublished raw data.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., … & Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *498*(7452), E1–E2.

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., et al. (2016). Intuition and moral decision-making: The effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLoS ONE, 11*(10), e0164012.

van Aert, R. C., Wicherts, J. M., & van Assen, M. A. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*(5), 713–729.

Verkoeijen, P. P., & Bouwmeester, S. (2014). Does intuition cause cooperation? *PLoS ONE, 9*(5), e96654.

⩗ Springer

Affiliations

**Amanda Kvarven[3] · Eirik Strømland[3] · Conny Wollbrant[5] · David Andersson[1] ·
Magnus Johannesson[2] · Gustav Tinghög[1] · Daniel Västfjäll[1] ·
Kristian Ove R. Myrseth[4]** [ID]

[1] Department of Management and Engineering, Linköping University, Linköping, Sweden

[2] Department of Economics, Stockholm School of Economics, Stockholm, Sweden

[3] Department of Economics, University of Bergen, Bergen, Norway

[4] The York Management School, University of York, York, UK

[5] Economics Division, Stirling Management School, University of Stirling, Stirling, UK

Springer