



HAL
open science

Intuition rather than deliberation determines selfish and prosocial choices

Bence Bago, Jean-François Bonnefon, Wim De Neys

► **To cite this version:**

Bence Bago, Jean-François Bonnefon, Wim De Neys. Intuition rather than deliberation determines selfish and prosocial choices. 2023. hal-04164426

HAL Id: hal-04164426

<https://hal.science/hal-04164426>

Preprint submitted on 18 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The logo consists of the letters 'W', 'W', and 'D' in a large, white, sans-serif font. The first 'W' is partially obscured by a dark blue horizontal bar that contains the text 'WORKING PAPERS' in white, uppercase, sans-serif font.

WORKING PAPERS

N° 1213

May 2021

“Intuition Rather Than Deliberation Determines Selfish and Prosocial Choices”

Bence Bago, Jean-François Bonnefon, Wim De Neys

Intuition Rather Than Deliberation Determines Selfish and Prosocial Choices

Bence Bago¹, Jean-François Bonnefon², Wim De Neys^{3*}

Affiliations:

¹ Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse 31015, France.

² Toulouse School of Economics (TSM-Research), Centre National de la Recherche Scientifique, University of Toulouse Capitole, Toulouse 31015, France.

³ Université de Paris, LaPsyDÉ, CNRS, F-75005 Paris, France.

(In Press) - Journal of Experimental Psychology : General

*Corresponding author:

Wim De Neys

LaPsyDE (CNRS & Université de Paris)

Sorbonne - Labo A. Binet

46, rue Saint Jacques

75005 Paris

France

wim.de-neys@parisdescartes.fr

ABSTRACT

Human interactions often involve a choice between acting selfishly (in ones' own interest) and acting prosocially (in the interest of others). Fast-and-slow models of prosociality posit that people intuitively favour one of these choices (the selfish choice in some models, the prosocial choice in other models), and need to correct this intuition through deliberation in order to make the other choice. We present 7 studies that force us to reconsider this longstanding “corrective” dual process view. Participants played various economic games in which they had to choose between a prosocial and a selfish option. We used a two-response paradigm in which participants had to give their first, initial response under time-pressure and cognitive load. Next, participants could take all the time they wanted to reflect on the problem and give a final response. This allowed us to identify the intuitively generated response that preceded the final response given after deliberation. Results consistently showed that *both* prosocial and selfish responses were predominantly made intuitively rather than after deliberate correction. Pace the deliberate correction view, the findings indicate that making prosocial and selfish choices does typically not rely on different types of reasoning modes (intuition vs deliberation) but rather on different types of intuitions.

INTRODUCTION

Humans show unique prosocial behavior. Often we are willing to pay a personal cost to help strangers who we never expect to encounter again (Jensen, Call, & Tomasello, 2007). This prosocial tendency is a key factor that allows human societies and economies to prosper (Fehr & Fischbacher, 2003; Tomasello, 2019). Economists, psychologists and other social scientists have long tried to pinpoint the mechanisms underlying human prosociality. Recent experimental work has been heavily influenced by the dual process framework that conceives human thinking as resulting from an interplay between fast intuition and slow deliberation (Kahneman, 2011). For example, one popular *deliberate prosociality* view entails that making prosocial choices requires deliberate control of our intuitive selfish impulses (e.g., DeWall, Baumeister, Gailliot, & Maner, 2008; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Zaki & Mitchell, 2013). An alternative *deliberate selfishness* view entails that our intuitions favor prosocial choices and it is only after deliberation that we will seek to maximize our self-interest (e.g., Rand, Greene, & Nowak, 2012; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003).

Despite the differences with respect to which behavior is assumed to be favored by deliberation and the—extensive—resulting debate (e.g., Baumeister, Vohs, & Tice, 2007; Bouwmeester et al., 2017; Capraro & Cococcioni, 2016; Rand, 2016, 2019; Tinghög et al., 2013, 2016; see Capraro, 2019 for a review), both the deliberate prosociality and deliberate selfishness view are built on the same underlying corrective deliberation model: Intuition will favor one type of behavior whereas making the competing choice will require slow, deliberate processes to control and correct the initial intuitive impulse (Wills, Hackel, & Van Bavel, 2019). If correct, such deliberate correction model holds interesting policy implications and promises since it implies that we can steer people towards more/less prosocial behavior by simply pushing them to deliberate more/less (Krajbich, Bartling, Hare, & Fehr, 2015).

Here we present a critical direct test of the underlying deliberate correction model. In our studies participants played incentivized, anonymous, one-shot economic games in which they had to choose to maximize their self-interest or opt for a prosocial choice. These classic games vary in their specifics but in all of them the prosocial choice consisted in paying a personal cost to give a benefit to others (Dictator Game, DG; Public Goods Game, PG) or to punish others' selfish behavior (second-mover Ultimatum Game, UG). Each game involves a specific sub-type of prosocial behavior (e.g., altruism, DG; costly punishment, UG; cooperation, PG). By including the various games we aimed to get the broadest possible test of the deliberate correction model of prosocial behavior.

We used a two-response paradigm (e.g., Kessler, Kivimaki, & Niederle, 2017; Krawczyk & Sylwestrzak, 2018; Thompson, Turner, & Pennycook, 2011) in which participants initially had to respond as fast as possible with the first intuitive answer that came to mind. Immediately afterwards they were given all the time they wanted to reflect and give a final answer. To make maximally sure that the initial response was

generated intuitively, possible deliberation was experimentally “knocked-out” by having people make their initial choice under time-pressure (i.e., a 2 s to 6 s response deadline) and concurrent cognitive load (i.e., memorization of a complex visual pattern). Given that deliberate processing is assumed to be more time and cognitive resource demanding than intuitive processing, forcing people to respond under time-pressure and load minimizes the possibility that people will deliberate during the initial response phase (Bago & De Neys, 2017, 2019). If the dual process correction model is correct, we expect that the alleged deliberate response—be it selfish or prosocial—will only (or at least predominantly) be observed in the final response stage after correction of one’s initial response (i.e., selfish-to-prosocial or prosocial-to-selfish changes).

We ran a total of seven studies. The various studies tested the generality and robustness of the findings across different games and methodological refinements.

STUDY 1: DICTATOR GAME

METHOD

Participants

112 Hungarian students (79 female, Mean age = 22.1 years, SD = 2.23 years) from the Eotvos Lorand University of Budapest were tested. A total of 86.7% of participants reported high school as the highest completed educational level, while 13.3% reported having a postsecondary education degree. Participants received course credit in return for participation and a possible monetary bonus payment (see further).

The sample size decision was based on previous two-response studies in the logical and moral reasoning field (Bago & De Neys, 2017, 2019) in which approximately 100 participants per condition were tested.

Participants in Study 1-4 were recruited from a pool which was renewed each semester and consisted of approximately 400 individuals. Study 1 and 4 were run in the same semester, Study 2 and 3 in the subsequent semester. All participants participated online. Recruitment numbers were regularly checked and manually stopped when at least 100 participants were tested or the pool was depleted. Each semester participants for the one-response pretests (see further) were recruited after the main two-response studies were completed.

Materials

Dictator Game (DG). We designed the two-choice DG following Schulz, Fischbacher, Thöni, and Utikal (2014). Participants were presented with a total of 16 DG trials (8 standard and 8 control, see further). In each trial, participants played the role of Dictator and were asked to divide 100 points (which were later

converted into money, at a rate of 1 point = 3 HUF, approximately \$0.01) between a new anonymous other player and themselves. Each trial presented two allocations of the 100 points. One allocation was always more fair (equal) than the other. For example:

- a) You keep 61, and the other player gets 39.
- b) You keep 51, and the other player gets 49.

For the fair option, we used four types of quasi-equal allocations: 51/49, 52/48, 53/47, 54/46. To avoid an “equality heuristic” (Schulz et al., 2014), we did not use the absolute fair 50/50 allocation and always allotted the largest share of the quasi-equal fair split to the participant. For the unfair option, we used both high difference trials (81/19, 82/18, 83/17, or 84/16) and low difference trials (61/39, 62/38, 63/37, or 64/36). Fair and unfair options were paired together such that the difference between the participant’s gains in all high and low difference trials remained constant (e.g., the unfair 81/19 and 82/18 options were paired with fair options 51/49 and 52/48, respectively). Note that the high and low difference factor did not affect the correction findings and was not further considered. The presentation order of the fair and unfair option was counterbalanced.

In addition to 8 standard DG trials, we also presented 8 control DG trials. In the standard problems the participant maximized self-interest by selecting the unfair split. Thus, selfish and prosocial fairness considerations cue conflicting choices. In the control trials, the participant gained more by selecting the quasi-equal fair split. This was achieved by changing the pay-off structure of the unfair option; in control trials the participants received the smallest share of the unfair split. For example:

- a) You keep 39, and the other player gets 61.
- b) You keep 51, and the other player gets 49.

Thus, in control trials selfish and prosocial fairness considerations cued the same response (i.e., the fair split). This should result in a high, non-random selection rate of this response option (Engel, 2011). For convenience, we label the response as the selfish choice. The presentation order of the two response options was counterbalanced and the same allocation combinations as in the standard trials were used. The eight standard and control trials were presented in random order.

Control trials allowed us to test for a guessing confound and are reported in this context. Unless otherwise noticed all reported results in the main text and supplementary material concern the critical standard trials.

Load task. We wanted to make maximally sure that participants' initial response was intuitive (i.e., deliberation is minimized). Therefore, we used a cognitive load task (i.e., the dot memorization task, see Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) to burden participants' cognitive resources while they were making their initial choice. The rationale behind the load manipulation is simple; dual process theories assume that deliberation requires more cognitive resources than intuition (Evans & Stanovich, 2013). Consequently, if we burden someone's cognitive resources, it will be less likely that they can engage in deliberation. We opted for the dot memorization task because it has been repeatedly shown to hinder deliberation during logical reasoning and economic game decision making (De Neys, Novitskiy, Geeraerts, Ramautar, & Wagemans, 2011; De Neys & Schaeken, 2007; De Neys & Verschueren, 2006; Franssens & De Neys, 2009). Before each game, participants were presented with a 3 x 3 grid, in which four grid squares were filled with crosses (see Figure 1 and S1). Participants were instructed that it was critical to memorize the location of the crosses even though it might be hard while making a game decision. After the DG choice was made, participants were shown four different matrices and they had to choose the correct, to-be-memorized pattern. They received feedback as to whether they chose the correct or incorrect pattern. The load was only applied during the initial response stage and not during the subsequent final response stage in which participants were allowed to deliberate (see further).

Procedure

Reading pretest. Before we ran the main study we recruited an independent convenience sample of 6 participants for a reading pretest (2 female, Mean age = 34.7 years, SD = 13.1 years). The basic goal of the reading pretest was to determine the response deadline which could be applied in the main DG study. The idea was to base the response deadline on the average reading time in the reading test (e.g., Bago & De Neys, 2017). The rationale was that if people are allotted the time they need to simply read the problem, we can expect that any further deliberate reasoning will be minimized (Bago & De Neys, 2017, 2019). Thus, in the reading pretest participants were presented with the same items as in the main DG. They were instructed to simply read the problems and randomly click on one of the answer options (see Supplementary Material A for literal instructions).

Each of the 16 problems was presented to the participants. Before each trial, a fixation cross was presented for one second. Once participants clicked on one of the response options, they were automatically taken to the next trial.

The average reading time was $M = 3.4$ s ($SD = 1.7$). Note that raw reaction time (RT) data were first logarithmically transformed prior to analysis. Mean and standard deviation were calculated on the transformed data, and then they were back-transformed into seconds (geometrical mean and SD). We wanted to give the participants some minimal leeway, thus we rounded the average reading time to the closest higher natural number (Bago & De Neys, 2017); the response deadline was therefore set to 4 s.

One-response (deliberative-only) pretest. We also ran a traditional “one-response” version of our Dictator game (without deadline or load). The same material as in the main two-response study was used. The only difference was that as in traditional economic game studies, participants were simply asked to give one single answer for which they could take all the time they wanted. We recruited an independent sample of 82 participants (61 female, Mean age = 22.4 years, SD = 3.1 years) from the Eotvos Lorand University of Budapest who received course credit and a possible monetary bonus payment depending on the game outcome. A total of 86.6% of the participants reported high school as highest completed educational level, while 13.4% reported having a postsecondary education degree.

The one-response version allowed us to check for a possible consistency confound in the two-response paradigm. When people are asked to give two consecutive responses, they might be influenced by a desire to look consistent. Thereby, the paradigm might underestimate the correction rate. Previous two-response work in other fields already argued against such a confound (De Neys & Bago, 2017, 2019; Thompson et al., 2011). Here we tested for it by contrasting the proportion of selfish responses on the standard trials in the one-response pretest and the final two-response stage in the main study. A consistency confound would result in a clear discrepancy. However, results showed that the rates in the one-response (43.4%, SD = 49.6) and final two-response stage were very similar (50.4%, SD = 50, see also Table S1). A mixed-effect logistic regression model (accounting for the random intercept of participants and items) indicated that the difference was not significant, $\chi^2(1) = 1.7, p = 0.19, b = 0.65$.

We also contrasted response latencies in the one-response pretest and main two-response study to check whether participants were under time-pressure when giving their initial response in the main study. Raw reaction time (RT) data were first logarithmically transformed prior to analysis (and again back-transformed into s afterwards). We focused on the critical standard trials. Results confirmed that participants responded faster in the initial two-response stage (Mean = 2.02 s, SD = 2.15) than in the unrestricted one-response condition (Mean = 2.53, SD = 1.86). A mixed-effect linear regression model (accounting for the random intercept of participants and items) indicated that the difference was significant, $\chi^2(1) = 9.79, p < 0.0001, b = -0.07$.

Two-response Dictator Game. The experiment was run online. Participants were specifically instructed at the beginning that we were interested in their very first, initial answer that came to mind. They were also told that they would have additional time afterwards to reflect on the problem and could take as much time as they needed to provide a final answer. After this general introduction, participants were presented with a more specific instruction page which explained them the upcoming game. Participants were informed that the game was incentivized and anonymous. Participants were told that each game was played with a different

anonymous person and that their identity would never be revealed (see Supplementary Material A for full instructions and pay-out structure).

After the specific instruction page, participants were presented with an attention check. They saw a hypothetical trial on screen and were asked to enter how many points they and the other player would make when they chose option A or B. Right amounts could be simply copied from the screen information. If a mistake was made, a second, similar attention check was given. If the second attention check was failed, the study immediately ended. After the attention check, participants started a practice session to familiarize them with the experimental procedure. First, they were presented with two practice DG trials in which they simply had to respond before the deadline. Next, they solved two practice dot matrix load problems (without concurrent DG). Finally, at the end of the practice, they had to solve the one earlier practice game and a new one under cognitive load just as in the main study.

Figure 1 illustrates the full trial structure. Each trial started with the presentation of a fixation cross for 1 s followed by the load matrix that stayed on the screen for 2 s. Next, the Dictator proposal choice appeared. From this point onward participants had 4 s to enter their answer; after 3 s the background of the screen turned yellow to warn participants about the upcoming deadline. If they did not provide an answer before the deadline, they were asked to pay attention to provide an answer within the deadline on subsequent trials. After the initial response, participants were asked to rate their experienced decision difficulty (“*How hard did you find it to make a decision? Please type a number from 0 (absolutely not hard) to 100 (very hard)*”). Next, they were presented with four matrix pattern options, from which they had to choose the correct, to-be-memorized pattern. If the answer was not correct, they were asked to pay more attention to memorize the correct pattern on subsequent trials. Finally, the full problem was presented again, and participants were asked to provide a final response. After the response was entered, participants again rated the experienced difficulty.

The colour of the answer options was green during the first response, and blue during the final response phase, to visually remind participants which question they were answering. Therefore, right under the question we also presented a reminder sentence: “Please indicate your very first, intuitive answer!” and “Please give your final answer”, respectively, which was also coloured as the answer options. At the end of the study participants completed a page with standard demographic questions. As in Rand et al. (2012), at this point, we also asked about their prior experience participating in similar economic game studies.

Exclusion criteria. All participants passed the attention check questions. Participants failed to provide a first response before the deadline in 1.4% of the trials. In addition, in 8.4% of the trials participants responded incorrectly to the dot memorization load task. All these trials were removed from the analysis because it cannot be guaranteed that the initial response resulted from intuitive processing: If participants took longer than the deadline or neglected the load task, they might have engaged in deliberation. Removing trials that did not meet the inclusion criteria gives us the purest possible test of our hypothesis. In total, 9.6%

of trials were excluded and 1620 trials (out of 1792) were further analysed (initial and final response for the same item counted as 1 trial).

Statistical analysis and Confidence Intervals (CI). Unless otherwise noticed, throughout the article we used mixed-effect regression models to analyze our results (Baayen, Davidson, & Bates, 2008), accounting for the random intercept of participants and items. For this, we used the lme4 package (Bates, Sarkar, Bates, & Matrix, 2007) in R (R Core Team, 2018). For binary choice data, we used logistic regression while for continuous data (RT and confidence) we used linear regression. When we had singular fit issues, the random intercept of items was excluded (these tests are shown with an * next to them).

CI's were computed using the DescTools package in R (Signorell, 2016). For binomial distributions (correction indexes, see further) we used the default “Wilson” confidence interval (Brown, Cai, & DasGupta, 2001) and for multinomial distributions (initial-to-final SS, PP, SP, PS averages, see further) we used the default “Sison-Glaz” approach (Sison & Glaz, 1995).

RESULTS AND DISCUSSION

Given that there are two response stages and two possible choices (selfish or prosocial) there are four possible combinations in which a subject can change their answer or not (initial and final selfish, SS; initial and final prosocial, PP; Initial selfish and final prosocial, SP; initial prosocial and final selfish, PS). Figure 2A plots an overview of the Study 1 results for the critical standard trials (see Table S2 for table format). It will be clear that we find evidence for corrective selfish (PS, 7.6% of total) and prosocial (SP, 8.8% of total) responses in which an initial choice is changed after deliberation in the final response stage. In and of itself, this indicates that deliberate correction exists. However, the problem is that these corrective cases are relatively rare. As Figure 2A indicates, it is far more likely that one's final selfish (SS, 42.8%) or prosocial (PP, 39.6%) choice was already selected in the initial response stage. Hence, both selfish and prosocial choices are typically already made intuitively in the absence of any deliberation. To quantify this, the last two bars in Figure 2A present the correction indexes (selfish correction index, CorS = the proportion of corrective selfish trials (PS), out of all final selfish choice trials (PS+SS); and the prosocial correction index, CorP = the proportion of corrective prosocial trials (SP) out of all final prosocial choice trials (SP+PP)). If the deliberate correction model is appropriate, this ratio should be at ceiling—or at the very least be above 50% and account for the majority of cases. Indeed, a low absolute proportion of PS or SP trials is not necessarily problematic for the corrective model. The key prediction is that in case the alleged (demanding) deliberate choice is made, it should only be made as final response. It is this critical ratio that is captured by the correction indexes.

However, as Figure 2A indicates, the selfish and prosocial correction indexes were far from ceiling or even the 50% mark (average CorS = 15.1%; average CorP = 20.2%). To test this statistically, for the selfish correction rate, we recoded SS trials as 0 and PS trials as 1. Then we excluded all PP and SP trials and calculated the average (which is equal to the CorS rate) and 95% CI. We used mixed-effect linear regression models to test whether the average of this variable differed from 50%. For the prosocial correction rate we coded PP as 0 and SP as 1, excluded all SS and SP trials, and applied the same statistics. Results showed that both the CorS, $b = -2.43, p < 0.0001^*$ and CorP, $b = -1.67, p < 0.0001$, indexes were significantly below 50%.

A critic might argue that the strong cognitive constraints at the initial response stage led people to make random first responses. Such guessing could overestimate the true initial selfish/prosocial response rate and drive the correction rate down. To test for such a general confound, half of the trials in our game were control trials in which there was no conflict between the selfish and prosocial choice. This was achieved by varying the pay-off structure such that a prosocial decision also maximized personal payoff. If people refrain from blind guessing, one expects a non-random preference for the prosocial/selfish choice on these control problems. Results showed that this was indeed the case. In Study 1, the selection rate of the selfish/prosocial response option (i.e., the fair split) in the control trials reached 77% (SD = 50%, see also Table S1) which significantly differed from chance, $b = -2.19, p < 0.0001$. We also calculated a stability index on the standard trials: For each participant, we calculated on how many standard trials they showed the same direction of change pattern (i.e., SS, PP, SP, or PS). The average stability in Study 1 was 76.3% (SD = 23.1) which was significantly higher than 50% chance, $t(111) = 12.06, p < 0.0001$ (in all our studies we used a one-sample, one-sided t-test for the stability analysis). This response consistency further indicates that participants were not giving random responses.

In sum, the Study 1 Dictator Game findings directly argue against the corrective deliberation model. In the vast majority of cases participants arrived at their selfish (prosocial) answer without any deliberation. In Study 2 we tested the generality of the findings with a two-response version of the Ultimatum Game.

STUDY 2: ULTIMATUM GAME

METHOD

Participants

101 Hungarian students (77 female, Mean age = 21.1 years, SD = 1.8 years) from the Eotvos Lorand University of Budapest were tested. A total of 85.7% of participants reported high school as the highest completed educational level, while 14.3% reported having a postsecondary education degree. Participants received course credit in return for participation and a possible monetary bonus payment.

Materials

Ultimatum Game (UG). Participants were presented with 16 UG trials (8 standard and 8 control, see further). The procedure was based on the work of Sanfey and colleagues (Sanfey et al., 2003; Van't Wout, Kahn, Sanfey, & Aleman, 2006). Participants played the role of responder (second-mover). In every trial, participants were faced with an offer made by a new, anonymous proposer to split an allocated sum of 10 points (points were later converted to HUF at a rate of 1 point = 10 HUF, approximately \$0.035). If participants accepted the offer, both players received the proposed amount. If they refused, neither player received anything. The item format looked as follows:

You received the following offer:

Player "A" keeps 9, and you get 1.

Do you accept this offer?

Yes

No

We presented 8 standard and 8 control trials in random order. In the standard trials the participant was presented with an unfair offer in which the proposer kept the larger part (9:1, 8:2, 7:3, and 6:4). Participants were presented every unfair offer twice. In the control trials the proposer always offered a fair (5:5) split. In all trials, participants always maximized their self-interest by accepting the offer. In the standard trials participants could make a prosocial choice and pay a personal cost to punish the other's player selfish behaviour by rejecting the offer (Rand et al., 2012). In control trials, the other player did not behave selfishly so there was no prosocial motivation to punish the other and reject the offer. Hence, a high rate of accepted control offers can be expected (Sanfey et al., 2003). For convenience, we will refer to accepted control trials as "selfish" responses.

Procedure

Reading pretest. Before we ran the main study we recruited an independent sample of 28 participants for a reading pretest (20 female, Mean age = 21.8 years, SD = 4.2 years) following the same procedure and logic as in Study 1. Each of the 16 UG items was presented to the participants. They were asked to simply read the sentences and click on one of the answers to advance to the next item.

The average reading time was 3.3 s (SD = 1.8). Note that raw RT data were first logarithmically transformed prior to analysis. Mean and standard deviation was calculated on the transformed data, and then they were back-transformed into seconds (geometrical mean and SD). In Study 1, results showed that on

average participants managed to give a response in the initial response stage well before the 4 s deadline. Therefore, in Study 2 we decided to round the reading time down to the lower integer and set the response deadline to 3 s.

One-response (deliberative-only) pretest. Following the same logic as in Study 1, we also ran a one-response pretest of the Ultimatum Game. We recruited an independent sample of 58 participants (44 female, Mean age = 21.5 years, SD = 2.5 years) from the Eotvos Lorand University of Budapest who received course credit and a possible monetary bonus payment depending on the game outcome. A total of 89.7% of the participants reported high school as highest completed educational level, while 10.3% reported having a postsecondary education degree.

We again observed a fairly similar rate of selfish choices on standard trials in the one-response pretest (39%, SD = 48.8) and the final stage of the main two-response study 41.9% (SD = 42.4), $\chi^2(1) = 0.07$, $p = 0.79$, $b = 0.14$. Latency analysis indicated that participants responded faster on standard trials in the initial two-response stage (Mean = 1.87 s, SD = 1.29) than in the unrestricted one-response condition (Mean = 2.74 s, SD = 1.84), $\chi^2(1) = 153.56$ $p < 0.0001$, $b = -0.29$.

Two-response paradigm. As in Study 1, participants were specifically instructed at the beginning that we were interested in their very first, initial answer that came to mind. They were also told that they would have additional time afterwards to reflect on the problem and could take as much time as they needed to provide a final answer. After the general two-response introduction (see Study 1), participants were presented with a more specific instruction page which explained them the upcoming game. Participants were informed that the game was incentivized and anonymous (see Supplementary Material A for full instructions and pay-out structure). The remainder of the two-response procedure was similar to Study 1. After the specific instruction page, participants were presented with an attention check, familiarized with the deadline and load procedure, and started the game. The only difference was that the initial response deadline was set to 3 s in the UG.

Exclusion criteria. Three participants failed the introductory attention check questions and could not participate further. The remaining participants failed to provide a first response before the deadline in 2.6% of the trials. In addition, in 8.6% of the trials participants responded incorrectly to the dot memorization load task. All missed deadline and load trials were excluded (10.8% of trials; 1399 trials out of 1568 were analysed).

RESULTS

Figure 2B and Table S2 show the results. We report an overview of the statistical analyses here. See Study 1 results for background.

As Figure 2B indicates, selfish (PS, 8.9%) and prosocial (SP, 5.5%) deliberate correction was rare and the corresponding correction indexes were low. Our statistical analysis confirmed that the selfish, 21.2%, $b = -1.57$, $p = 0.006$, and prosocial, 9.4%, $b = -2.6$, $p < 0.0001$, correction index rate for the standard UG were significantly lower than 50%.

The selection rate of the selfish/prosocial response option (i.e., accepting the fair split) in the control trials was 95.3% (SD = 21.2%) which significantly differed from chance, $b = 10.52$, $p < 0.0001^*$. The average standard trial stability in Study 2 was 74.8% (SD = 21.5) which was significantly higher than 50% chance, $t(97) = 11.43$, $p < 0.0001$.

Study 2 showed that deliberate correction in the Ultimatum Game was the exception rather than the rule. As in Study 1, correction indexes were low and people typically arrived at their selfish (prosocial) answer without deliberation. In Study 3, we further tested the generality of the findings with a two-response version of a two-player Public Good Game.

STUDY 3: PUBLIC GOOD GAME

METHOD

Participants

95 Hungarian students (58 female, Mean age = 21.2 years, SD = 1.95 years) from the Eotvos Lorand University of Budapest were tested. A total of 86.6% of participants reported high school as the highest completed educational level, while 13.4% reported having a postsecondary education degree. Participants received course credit in return for participation and a possible monetary bonus payment.

Materials

Public Good Game (PG). Participants played 16 PG trials (8 standard and 8 control trials). In each trial, participants were randomly paired with another, anonymous participant. In each trial participants were given 100 points and were told that they could both decide to contribute this allocation to a common pool or not. They were informed that the number of points in the common pool would be multiplied by a factor 1.2 to 2.9 and that the resulting total points would be equally split between the two participants. In each trial participants were presented with the actual pay-off matrix on screen. For example:

	Player B contributes	Player B doesn't contribute
You contribute	120 / 120	60 / 160

You don't contribute	160 / 60	100 / 100
----------------------	----------	-----------

Do you contribute to the common pool?

- a) Yes
- b) No

The participants' pay-off was shown in black, the other players' pay-off in grey. The order of the matrix rows (and response options) was counterbalanced; for half of the participants the top row ("You contribute") stated the pay-offs in case they contributed, for the other half the top row stated the pay-offs in case of non-contribution. Response options were ordered similarly (e.g., if first row stated contribution pay-off, the upper response option stated "Yes").

We presented 8 standard and 8 control trials. In standard trials we multiplied the number of points with a multiplier between 1.2 and 1.9. In these trials, self-interest is maximized by not contributing whereas the group pay-off is maximized by contributing. Hence, selfish considerations cue a "No" decision, whereas prosocial considerations cue a "Yes" decision. In control trials, we selected multipliers between 2.2 and 2.9. In these trials, contributing always yields a higher pay-off compared to non-contribution both for oneself and for the group. Hence, both selfish and prosocial motivations cue the same "Yes" response. Consequently, a high, non-random proportion of "Yes" responses is expected on the control trials. For convenience, we label such "Yes" control responses as "selfish" choices.

Procedure

Reading pretest. Before we ran the main study we recruited an independent sample of 28 participants for a reading pretest (20 female, Mean age = 21.3 years, SD = 2.3 years) following the same procedure and logic as in Study 1-2. Each of the 16 PG items was presented to the participants. They were asked to simply read the information and click on one of the answers to advance to the next item.

The average reading was 6.1 s (SD = 2.5). Note that raw reaction time (RT) data were first logarithmically transformed prior to analysis. Mean and standard deviation were calculated on the transformed data and back-transformed into seconds. Following the logic of Study 2, we rounded the average reading time to the closest lower natural number; the response deadline was therefore set to 6 s.

One-response (deliberative-only) pretest. Following the same logic as in Study 1-2, we also ran a one-response pretest of our Public Good Game. We recruited an independent sample of 65 participants (48 female, Mean age = 21.5 years, SD = 2 years) from the Eotvos Lorand University of Budapest who received

course credit and a possible monetary bonus payment depending on the game outcome. No specific education level information was recorded but the participants came from the same university sample as the ones in the main study.

We again observed a fairly similar mean rate of selfish choices on standard trials in the one-response pretest (70.6%, SD = 45.6%) and the final stage of the main two-response study (66.2%, SD = 47.3%), $\chi^2(1) = 0.04$, $p = 0.84$, $b = -0.1$. Latency analysis indicated that participants responded faster on standard trials in the initial two-response stage (Mean = 2.94 s, SD = 1.5) than in the unrestricted one-response condition (Mean = 6.92 s, SD = 2.25), $\chi^2(1) = 96.93$, $p < 0.0001$, $b = -0.37$.

Two-response paradigm. As in Study 1-2, participants were specifically instructed at the beginning that we were interested in their very first, initial answer that came to mind. They were also told that they would have additional time afterwards to reflect on the problem and could take as much time as they needed to provide a final answer. After the general two-response introduction (see Study 1), participants were presented with a more specific instruction page which explained them the upcoming game. Participants were informed that the game was incentivized and anonymous (see Supplementary Material A for full instructions and payout structure).

The remainder of the two-response procedure was similar to Study 1. After the specific instruction page, participants were presented with an attention check, familiarized with the deadline and load procedure, and started the game. The only difference was that the initial response deadline was set to 6 s in the PG.

Exclusion criteria. Three participants failed the introductory attention check questions and could not participate further. The remaining participants failed to provide a first response before the deadline in 1.7% of the trials. In addition, in 8.7% of the trials participants responded incorrectly to the dot memorization load task. All missed deadline and load trials were excluded (10.1% of trials; 1180 trials out of 1312 were analysed).

RESULTS

Figure 2C and Table S2 show the results. As Figure 2C indicates, selfish (PS, 10.6%) and prosocial (SP, 6.1%) deliberate correction was rare and the corresponding correction indexes were low. The selfish, 16%, $b = -2.2$, $p < 0.0001^*$, and prosocial, 18.2%, $b = -2.06$, $p < 0.0001^*$, correction indexes in the standard PGG trials were significantly lower than 50%.

The selection rate of the selfish/prosocial response option (i.e., “Yes” responses) in the control trials was 86% (SD = 34.7) which significantly differed from chance, $b = 10.5$, $p < 0.0001^*$. The average standard trial stability was 78.7% (SD = 22) which was significantly higher than 50% chance, $t(81) = 11.83$, $p < 0.0001$.

In sum, the Study 3 Public Goods Game findings were consistent with the results we obtained for the Dictator Game (Study 2) and Ultimatum Game (Study 2) and showed that prosocial and selfish responses are typically driven by intuitive processing rather than deliberate correction.

In Study 4-7 we addressed various alternative explanations and possible experimental confounds. One might note that although the two-response paradigm allows participants to deliberate in the final response stage, this does not guarantee they actually do so. This is also a common issue for previous studies that only use a single (one-response), deliberative stage. A common solution when examining the impact of deliberation on prosocial choice is to force people to think for at least 10 s before they can enter a response (Rand et al., 2012). Therefore, in Study 4 participants played the same Dictator Game as in Study 1 but with forced deliberation in the final response stage.

STUDY 4: DICTATOR GAME WITH FORCED DELIBERATION

METHOD

Participants

112 Hungarian students (79 female, Mean age = 22.13 years, SD = 2.4 years) from the Eotvos Lorand University of Budapest were tested. A total of 89.2% of participants reported high school as the highest completed educational level, while 10.8% reported having a postsecondary education degree. Participants received course credit in return for participation and a possible monetary bonus payment.

Materials & Procedure

Delayed final response. We used the same DG trials with the same two-response paradigm as in Study 1. The only change we made was the forced deliberation at the final response stage. Participants were instructed to reflect for at least 10 s in the final response stage and could not enter a response before 10 s had passed (see Supplementary Material A for full instructions).

Exclusion criteria. One participant failed the introductory attention check questions and could not participate further. The remaining participants failed to provide a first response before the deadline in 2.9% of the trials. In addition, in 10.3% of the trials participants responded incorrectly to the dot memorization load task. All missed deadline and load trials were excluded (12.5% of trials; 1554 trials out of 1776 were analysed).

RESULTS

Figure 2D and Table S2 show the results. As Figure 2 indicates, despite the forced deliberation in Study 4 results were similar to Study 1. Deliberate correction remained the rare exception. The statistical analysis confirmed that the selfish, 11.7%, $b = -2.64$, $p < 0.0001$, and prosocial, 20.2%, $b = -2.51$, $p < 0.0001^*$, correction index rates on the standard trials were significantly lower than 50% in Study 4.

With respect to guessing, the selection rate of the selfish/prosocial response option in the control trials was 81.2% (SD = 39.1%) which significantly differed from chance, $b = 2.85$, $p < 0.0001^*$. The average standard trial stability in Study 4 was 80.3% (SD = 21.7) which was significantly higher than 50% chance, $t(110) = 14.73$, $p < 0.0001$.

In sum, Study 4 confirmed the Study 1 results. In study 5-7 we further refined the methodology. Although our instructions, load, and time-pressure manipulations were based on established procedures, one could argue that they were not demanding enough and still allowed participants to deliberate during the initial response stage (Kessler et al., 2017; Krawczyk & Sylwestrzak, 2018). Therefore, we tuned up our manipulations in Study 5-7. In addition to the forced deliberation we also used more stringent response deadlines and more extreme load. In Study 1-4 participants memorized a 4-dot pattern in a 3x3 grid while making their initial game choices whereas this was increased to a 5-dot pattern in a 4x4 grid in study 5-7. In addition, the response deadline for each of the games in Study 5-7 was set at the average initial response time observed in Study 1-4.

STUDY 5: DICTATOR GAME WITH FORCED DELIBERATION AND EXTREME TIME PRESSURE/LOAD

METHOD

Participants

119 participants were recruited (56 female, Mean age = 32.5 years, SD = 10.7 years) on Prolific Academic. Only native English speakers from Canada, Australia, New Zealand, the United States of America or the United Kingdom were allowed to take part in the study. Among them 47.1% reported high school as the highest completed educational level, while 52.1% reported having a postsecondary education degree. Participants received a participation fee (£5/hour) and a possible bonus payment depending on the game outcome.

Because we expected a higher number of excluded trials in the extreme deadline/load treatment we recruited a slightly higher number of participants in Study 5-7. For each study the goal was to get at approximately 120 analyzable data sets (vs approximately 100 in Study 1-4).

Materials & Procedure

Delayed final response. The exact same Dictator Game with forced deliberation procedure as in Study 4 was used.

Extreme load. In Studies 1–4, participants had to memorize a complex four-dot pattern in a 3 x 3 grid during the initial response stage. In Study 5 we presented an even more complex five-dot pattern in a 4 x 4 grid (e.g., Bialek & De Neys, 2017; Trémolière & Bonnefon, 2014, see Figure 1 and S1). Except for the precise load pattern, the load procedure was similar to Study 1–4.

Response deadline. In Study 1 and 4 the response deadline for the initial DG response was 4 s and the average initial response latency on standard trials was 2.02 s ($SD = 2.15$). In Study 5 we decreased the deadline to this average (rounded to nearest integer) and set it to 2 s. The screen background turned yellow to alert the participant to enter their response 500 ms before the deadline.

Difficulty rating. In Study 1-4 participants were asked to indicate how hard they found it to make a decision after each response. In Study 5-7 this difficulty rating was dropped.

Exclusion criteria. All participants passed the introductory attention check questions. Participants failed to provide a first response before the deadline in 8.9% of the trials. In addition, in 14.6% of the trials participants responded incorrectly to the dot memorization load task. All missed deadline and load trials were excluded (21% of trials; 1505 trials out of 1904 were analysed). Note that—consistent with the stronger constraints—the percentage of discarded trials in Study 5(-7) was generally higher than in Study 1(-4).

RESULTS

Figure 2E and Table S2 show the Study 5 results. As Figure 2 indicates, despite the stronger constraints, results were similar to Study 1 and 4. Deliberate correction remained exceptional and never became the dominant pattern. Statistical testing indicated that the selfish, 8.3%, $b = -3.18$, $p < 0.0001^*$, and prosocial, 27.7%, $b = -1.32$, $p < 0.0001^*$, correction index rates were significantly lower than 50%.

Given the stronger cognitive constraints in Study 5 it is important to check for a possible random responding or guessing confound. However, the selection rate of the selfish/prosocial response option in the control trials reached 86.6% ($SD = 34.1$) which significantly differed from chance, $b = 3.15$, $p < 0.0001$. The average standard trial stability in Study 5 was 83.2% ($SD = 20.03$) which was significantly higher than 50% chance, $t(117) = 17.996$, $p < 0.0001$. This establishes that despite the stronger constraints, participants were not responding randomly.

The Study 5 Dictator Game findings are consistent with the Study 1 and 4 Dictator Game results. In Study 6 we adopted the same strong constraints as in Study 5 with the Ultimatum Game.

STUDY 6: ULTIMATUM GAME WITH FORCED DELIBERATION AND EXTREME TIME PRESSURE/LOAD

METHOD

Participants

126 participants (55 female, Mean age = 33.8 years, SD = 10.4 years) were recruited on Prolific Academic. Only native English speakers from Canada, Australia, New Zealand, the United States of America or the United Kingdom were allowed to take part in the study. 35.3% of participants reported high school as the highest completed educational level, while 63.8% reported having a postsecondary education degree. Participants received a participation fee (£5/hour) and a possible bonus payment depending on the game outcome.

Materials & Procedure

Delayed final response. We used the same ultimatum game trials as in Study 2. The delayed response procedure was completely similar to the one adopted in Study 4-5.

Extreme load. We used the same extreme load task as described in Study 5.

Response deadline. In Study 2 the response deadline for the initial DG response was 3 s and the average initial response latency on standard trials was 1.87 s ($SD = 1.29$). In Study 6 we decreased the deadline to this average (rounded to nearest integer) and set it to 2 s. The screen background turned yellow to alert the participant to enter their response 500 ms before the deadline.

Exclusion criteria. Five participants failed the introductory attention check questions and could not participate further. The remaining participants failed to provide a first response before the deadline in 14.6% of the trials. In addition, in 17.7% of the trials participants responded incorrectly to the memorization load task. All missed deadline and load trials were excluded (27% of trials; 1514 trials out of 1936 were analysed).

RESULTS

Figure 2F and Table S2 show the Study 6 results. As Figure 2 indicates, despite the stronger constraints, key results were similar to Study 2. Deliberate correction remained exceptional and never became the dominant pattern. The selfish, 6.3%, $b = -10.87$, $p < 0.0001$, and prosocial, 23.1%, $b = -1.96$, $p < 0.0001^*$, correction index rates on the standard trials were both significantly lower than 50%.

The selection rate of the selfish/prosocial response option in the control trials reached 96.7% (SD = 17.9) which significantly differed from chance, $b = 8.72$, $p < 0.0001^*$. The average standard trial stability in Study 6 was 78.1% (SD = 20.8) which was significantly higher than 50% chance, $t(113) = 14.43$, $p < 0.0001$. This again indicates that the stronger cognitive constraints did not underestimate the correction indexes by prompting a random response tendency.

Study 6 confirmed our previous findings. In the final Study 7 we also tested the Public Good Game with the strong constraints that we adopted in Study 5-6.

STUDY 7: PUBLIC GOODS GAME WITH FORCED DELIBERATION AND EXTREME TIME PRESSURE/LOAD

METHOD

Participants

131 participants (63 female, Mean age = 30.3 years, SD = 10.5 years) were recruited on Prolific Academic. Only native English speakers from Canada, Australia, New Zealand, the United States of America or the United Kingdom were allowed to take part in the study. 47.3% of participants reported high school as the highest completed educational level, while 50% reported having a postsecondary education degree. Participants received a participation fee (£5/hour) and a possible bonus payment depending on the game outcome.

Materials & Procedure

Delayed final response. We used the same Public Good Game trials as in Study 3. The delayed response procedure was completely similar to the one adopted in Study 4-6.

Extreme load. We used the same extreme load task as described in Study 5-6.

Response deadline. In Study 3 the response deadline for the initial PG response was 6 s and the average initial response latency on standard trials was 2.94 s (SD = 1.5). Consistent with the Study 5-6 rationale, in Study 7 we decreased the deadline to this average (rounded to nearest integer) and set it to 3 s. The screen background turned yellow 1 s before the deadline to alert the participant to enter their response.

Exclusion criteria. 19 participants failed the introductory attention check questions and could not participate further. The remaining participants failed to provide a first response before the deadline in 11.5% of the trials. In addition, in 13.8% of the trials participants responded incorrectly to the dot memorization load task. All missed deadline and load trials were excluded (15.8% of trials; 1509 trials out of 1792 were analysed).

RESULTS

Figure 2G and Table S2 show the Study 7 results. As Figure 2 indicates, despite the stronger constraints, critical findings are consistent with our Public Good Game findings in Study 3 and our other studies. The correction indexes indicate that deliberate correction never became the dominant pattern. The statistical analysis confirmed that both the selfish, 25.2%, $b = -1.73$, $p < 0.0001$, and prosocial, 25.2%, $b = -1.26$, $p = 0.002$, correction index rate on the standard trials were significantly lower than 50%.

The selection rate of the selfish/prosocial response option in the control trials in Study 7 was 85.6% (SD = 35.1) which significantly differed from chance, $b = 3.06$, $p < 0.0001$. The average standard trial stability was 72.2% (SD = 20.8) which was significantly higher than 50% chance, $t(111) = 11.67$, $p < 0.0001$. This again indicates that participants were not responding randomly.

Our results were very similar in the different studies that we ran. For convenience, we also ran an analysis on the combined Study 1-7 data. These results are shown in Figure 2H (see Supplementary Material B for statistical tests).

GENERAL DISCUSSION

The present study tested the corrective dual process model of prosocial behavior. This influential model entails that intuition will favor one type of behavior whereas making the competing choice will require slow, deliberate processing to control and correct the initial intuitive impulse. Pace the deliberate correction prediction, in each of our seven studies we consistently observed that both prosocial and selfish responses were predominantly made intuitively rather than after deliberate correction. Although corrective patterns were observed, they were the rare exception rather than the rule. In general, it seems that the corrective dual process model—and the deliberate prosociality and deliberate selfishness views that are built on it—has given too much weight to deliberation. Making prosocial and selfish choices does not typically rely on different types of reasoning systems (i.e., an intuitive “System 1” and deliberate “System 2” or an impulsive and controlled system) but rather on different types of intuitions.

Clearly, the conclusion only holds in so far as there are no alternative explanations or experimental confounds. In the consecutive studies and pretests we controlled for various possible issues. We summarize

the key points here and discuss some additional control analyses. First, the cognitive constraints at the first stage of the two-response paradigm might have led people to make random first responses, which would lead to an underestimation of the frequency and systematicity of their corrections. However, our “control” trials (on which prosocial decisions also maximized personal payoff) and stability analysis consistently indicated that responses at the intuitive stage were not random.

Second, the two-response paradigm might introduce experimental demands for consistency or an anchoring effect that would artificially reduce the rate at which participants correct their decisions. To rule out this concern, Studies 1–3 included an additional “one-response” pretest in which participants only gave deliberate responses, skipping the first intuitive stage. Results showed that choices in the final two-response stage and pretest were very consistent. For none of the games, choices differed by more than 7%. Hence, a possible consistency confound cannot explain the very low correction rates.

Third, the two-response paradigm allows people to deliberate in the final response stage but does not guarantee they actually do so. Since previous studies that examined the impact of deliberation on prosocial choice often forced people to think for at least 10 s before they could enter a response, Studies 4-7 adopted this forced deliberation in the final response stage. However, even with forced deliberation, the correction rates remained floored.

Fourth, even if our instructions, time pressure and cognitive load manipulations had all been largely validated by previous research, it is still possible that they were not demanding enough, and thus allowed participants to engage in deliberation at the first, intuitive stage. To rule out this concern, we used more extreme load and time pressure manipulations in Study 5-7. Deliberate correction remained infrequent and never became the modal pattern.

Fifth, contrasting of the unrestricted one-response pretests latencies in Study 1-3 and the corresponding initial two-response latencies also established that even in our least restrictive two-response studies 1-3, participants were under time pressure. In addition, we also ran an extra control analysis in which we contrasted the correction index rate for fast and slow initial responses (based on a median split on initial response time in each of the studies). If our deadlines were not stringent enough and allowed (some) participants to engage in deliberation in the initial response stage, one expects that the correction rate should be higher for fast vs slow initial responses (since slow responders had more time to engage in possible preemptive deliberation). Results showed this was not the case (see supplementary Material B, Table S2-S4). Even for the fastest half of responses, the correction rate remained floored.

Sixth, the literature has pointed to possible moderators of the prevalence of selfish/prosocial choice such as gender and experience in participating in economic games (McAuliffe, Forster, Pedersen, & McCullough, 2018; Rand, 2016; Rand, Brescoll, Everett, Capraro, & Barcelo, 2016; Rand et al., 2012). In additional control tests we also analyzed results separately for both genders (Table S5-S6) and for subjects who indicated they had previously participated in economic game studies or not (Table S7-S8). Given that

participants played multiple one-shot games in each study we also ran an analysis on their first game trial only (Table S9) to sidestep a possible repeated testing confound (McAuliffe et al. 2018). However, none of these factors altered the conclusions. For none of the control analyses there was a single study in which deliberate correction became the modal pattern. Irrespective of gender and experience, selfish and prosocial choices were predominantly made intuitively.

Seventh, we always discarded from analyses the trials in which participants failed to comply with the deadline or failed their memorization task. This eliminates the possibility that participants neglected and traded-off the time-pressure or load task to deliberate during the initial game choice. However, although the number of discarded trials was acceptable (overall 12% of all trials), in theory, this might lead to a selection confound (Bouwmeester et al., 2017). Table S10 shows the data without application of the load exclusion criterion (missed deadline trials were not recorded). Results were fully consistent with the original analyses.

In sum, our results show strong evidence that deliberation has a very limited impact on prosocial decisions. Both prosocial and selfish choices are typically generated intuitively and rarely require deliberation. Previous research already established that different individuals have different yet stable inclinations to make prosocial decisions in the economic games we used in this article, at least when they only know about the payoffs of the game (and not, e.g., about the other player). Different labels exist for one's inclination to be prosocial in this context, for example, one's social value orientation (Murphy & Ackermann, 2014), cooperative phenotype (Peysakhovich, Nowak, & Rand, 2014), or other-regarding preferences (Cooper & Kagel, 2016). When people express their orientations, phenotypes, or preferences in the context of economic games, they may rely either on intuitive or deliberate processing. Dual-process models typically assume that the people who deliberate have an opportunity to correct their intuitive response, in a systematic direction. Our results suggest that few people take advantage of this opportunity, and that their corrections do not follow a systematic pattern. Accordingly, social value orientations, cooperative phenotypes and other-regarding preferences are likely to largely reflect peoples' intuitive reactions, rather than deliberate corrections. By and large, the individuals who are intuitively prosocial remain prosocial when they deliberate, and the individuals who are intuitively selfish remain selfish when they deliberate.

This is not to say that people cannot change their decision through deliberation. In every study, we observed instances of deliberate correction. The problem is one of prevalence and impact, not of existence. Neither do we argue that deliberation has no possible role in prosocial contexts—for example, it may very well allow people to give ex post justifications for their prosocial or selfish decisions (Mercier & Sperber, 2011). What seems clear, though, is that deliberation plays little role in shaping these decisions. Deliberation can also play a role in prosocial games when people receive additional information about the context of the game, in addition to its payoffs. For example, people may receive information about the reputation of the other player, they may have direct information about the behavior of the other player in the case of repeated games, they may be able to exchange messages with the other player, they may be able to see a picture of the

other player, and so forth. There are infinite variations on the kind of additional information people may access, that may change their decision. Depending on the kind of information, this change may result from intuitive processing, or from deliberate processing. Our findings concern the minimal, prototypical setting in which people only know about the payoffs of the game, and do not speak to variants in which additional information may involve deliberate processing.

Some aspects of our data are consistent with specific conjectures about human prosociality. For example, based on the Social Heuristic Hypothesis (SHH, Rand et al., 2012, 2014) one might predict that in case deliberation alters our intuitive choices in the context of pure cooperation games (as tracked by the Public Good Game, but not the Dictator or Ultimatum Game), it will lead to more selfish than prosocial choices. Our two Public Good Games studies (Study 3 and 7) indeed showed that prosocial-to-selfish (PS) changes were more likely than selfish-to-prosocial (SP) changes (See Figure 2). Hence, this specific aspect of the SHH about the direction of deliberate correction is supported by our work. However, the key problem is the prevalence of deliberate correction per se. Although deliberate correction—when it occurs—might make people more selfish in the Public Good Game, it is far more likely that a selfish Public Good Game decision is generated intuitively than that it follows from deliberation. The point is that the dual process correction mechanism is only accounting for a small fraction of our behavior. As such it does not present a viable psychological model of prosocial behavior.

Interestingly, the low prevalence of deliberate correction was also observed in previous studies on prosociality that attempted to track how people changed (or didn't change) their answer after prolonged deliberation (e.g., Kessler et al., 2019; Krawczyk & Sylwestrzak, 2018). Although the design of these studies did not prevent participants from deliberating when they generated their first answer, the present findings validate the observed trends. The low prevalence of deliberate correction can also help to explain some of the discrepant findings in the prosociality literature. As we noted, there is a longstanding discussion as to whether deliberation promotes prosocial (vs selfish) behavior. Experimental manipulations designed to favor intuitive processing in prosocial choice tasks have resulted in mixed results: some studies reported null-findings, some support the deliberate selfishness account, and others support the deliberate prosociality account (see Capraro, 2019, for a review). Meta-analyses have indicated that if anything, the effect is small (e.g., Rand, 2017, 2019). For example, Rand (2019) found that experimental manipulations that favored intuition/limited deliberation led on average to an increase of 3.1% prosocial cooperative choices. The present findings can help to put this in perspective. Deliberate correction clearly exists. However, the prevalence is low and there are always corrections in both directions. This implies that any experimental procedure aimed to “knock-out” deliberation is bound to have small effects. In the vast majority of cases, people will manage to generate the same response intuitively. It is only the small fraction of corrective cases that can be affected. Hence, small effect sizes might not point to a “failed” manipulation but simply reflect the low prevalence of the corrective behavior in question. Likewise, given that the base-rate of corrective cases is so low, even minimal individual

sample composition variation (i.e., slightly higher SP or PS cases) might result in opposite effects of one's manipulation.

As we noted, at the theoretical level the findings suggest that dual-process, fast-and-slow models of prosociality have given too much weight to the role of deliberative correction. This fits with recent calls to upgrade the view of intuitive reasoning in dual process models (e.g., De Neys, 2017). Just as in the current case of prosocial choice, in other fields evidence is also amassing against the deliberate correction view. From moral reasoning to logical inferencing, emerging evidence indicates that the response that traditional dual process models assume to require deliberation, is often already generated intuitively (e.g., Bago & De Neys, 2017, 2019; Gürçay & Baron, 2017; Newman, Gibb, & Thompson, 2017; Pennycook, Fugelsang, & Koehler, 2015; Trippas & Handley, 2017). Newer models posit that fast-and-slow decision making is mostly a conflict between fast intuitions, with a limited role for deliberate correction. Under this view, decisions are driven by the activation levels of competing intuitions, and deliberate correction only occurs in difficult cases when competing intuitions have maximally similar activation levels (e.g., Bago & De Neys, 2017; De Neys & Pennycook, 2019). Interestingly, there is some evidence in the current studies that fits with this account. In Study 1-4 we asked participants to rate how hard they found it to make a decision. Results showed that for the rare trials on which the initial response was subsequently corrected after deliberation, the initial decision was experienced as more difficult than for trials that were not changed (Figure S2, Supplementary Material B). Although further validation will be required, this framework might present the various psychologists, economists, and philosophers interested in human prosociality with a more viable alternative to the traditional dual process model.

It is perhaps not surprising that the deliberate correction dual process model has attracted the attention of various policy makers (e.g., Chater, 2018; Melnikoff & Bargh, 2018; Thaler & Sunstein, 2008; World Bank, 2015). Clearly, if the model is correct, it holds the promise of steering people's behavior in one direction or the other by simple interventions that prime intuitive vs deliberate processing. But even if such interventions are deemed ethically desirable, the present results suggest that they will have overall limited impact on people's prosocial behavior. Given that both prosocial and selfish choices are predominantly generated intuitively, stimulating intuition or deliberation per se will not be very effective. Based on the current results it seems more promising for policy makers to try to target different types of intuitions rather than intuitive vs deliberate thinking per se.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the Agence Nationale de la Recherche (DIAGNOR ANR-16-CE28-0010-01). BB and JFB were supported by ANR grant ANR-17-EURE-0010 (Investissements d'Avenir program), and ANR Labex IAST. BB was further supported by the Scientific Research Fund Flanders (FWO-

Vlaanderen). We would like to thank Balazs Aczel for providing us with the participant pool for this research.

Raw data can be retrieved from https://osf.io/cwhuz/?view_only=3aa64419c2c44a5783b1bf44d02d36b2.

REFERENCES

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109.
- Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801.
- Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R Package Version*, *2*(1), 74.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355.
- Bialek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, *12*(2), 148–167.
- Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., ... Espín, A. M. (2017). Registered Replication Report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, *12*(3), 527–542.
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 101–117.
- Capraro, V. (2019). The dual-process approach to human sociality: A review. *Available at SSRN 3409146*.
- Capraro, V., & Cococcioni, G. (2016). Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports*, *6*, 27219.
- Chater, N. (2018). Is the Type 1/Type 2 Distinction Important for Behavioral Policy? *Trends in Cognitive Sciences*, *22*(5), 369–371.

- Cooper, D. J., & Kagel, J. H. (2016). Other-regarding preferences. J. H. Kagel and A. E. Roth (Eds), *The Handbook of Experimental Economics*, 217-290. Princeton University Press.
- De Neys, W. (2017). Bias, conflict, and fast logic: Towards a hybrid dual process future? In W. De Neys (Ed.), *Dual Process Theory 2.0*. Oxon, UK: Routledge.
- De Neys, W., Novitskiy, N., Geeraerts, L., Ramautar, J., & Wagemans, J. (2011). Cognitive control and individual differences in economic ultimatum decision-making. *PLoS One*, 6(11), e27107.
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28, 503–509.
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie)*, 54(2), 128–133.
- De Neys, W., & Verschueren, N. (2006). Working memory capacity and a notorious brain teaser: The case of the Monty Hall Dilemma. *Experimental Psychology*, 53(2), 123–131.
- DeWall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin*, 34(12), 1653–1662.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583–610.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791.
- Franssens, S., & De Neys, W. (2009). The effortless nature of conflict detection during thinking. *Thinking & Reasoning*, 15(2), 105–128.
- Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, 23(1), 49–80. <https://doi.org/10.1080/13546783.2016.1216011>
- Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science*, 318(5847), 107–109.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

- Kessler, J., Kivimaki, H., & Niederle, M. (2017). *Thinking fast and slow: generosity over time*. Retrieved from http://assets.wharton.upenn.edu/~juddk/papers/KesslerKivimakiNiederle_GenerosityOverTime.pdf
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.
- Krajibich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*, 7455.
- Krawczyk, M., & Sylwestrzak, M. (2018). Exploring the role of deliberation time in non-selfish behavior: The double response method. *Journal of Behavioral and Experimental Economics*, *72*, 121–134.
- McAuliffe, W. H., Forster, D. E., Pedersen, E. J., & McCullough, M. E. (2018). Experience with anonymous interactions reduces intuitive cooperation. *Nature Human Behaviour*, *2*(12), 909–914.
- Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(02), 57–74.
- Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640.
- Newman, I., Gibb, M., & Thompson, V. A. (2017). Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief -bias and base-rate neglect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1154–1170.
- Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.
- Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a ‘cooperative phenotype’ that is domain general and temporally stable. *Nature Communications*, *5*, 4939.

- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>
- 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. (2017). 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. (2019). Intuition, Deliberation, and Cooperation: Further Meta-Analytic Evidence from 91 Experiments on Pure Cooperation. Available at SSRN 3390018. 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., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755–1758.
- Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, 41, 77–87.
- Signorell, A. (2016). DescTools: Tools for descriptive statistics. R Package Version 0.99, 17.
- Sison, C. P., & Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, 90(429), 366–369.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107–140.

- 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., & Västfjäll, D. (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.
- Tomasello, M. (2019). *Why we cooperate*. Cambridge, MA: MIT Press.
- Trémolière, B., & Bonnefon, J.-F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*(7), 923–930.
- Trippas, D., & Handley, S. (2017). The parallel processing model of belief bias: review and extensions. In W. De Neys (Ed.), *Dual Process Theory 2.0* (pp. 28–46). Oxon, UK: Routledge.
- Van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research*, *169*(4), 564–568.
- Wills, J., Hackel, L., & Van Bavel, J. J. (2019). *Shifting prosocial intuitions: Neurocognitive evidence for a value based account of group-based cooperation*. Retrieved from; <https://doi.org/10.31234/osf.io/u736d>
- World Bank. (2015). *World development report 2015: Mind, society, and behavior*. World Bank Publications.
- Zaki, J., & Mitchell, J. P. (2013). Intuitive prosociality. *Current Directions in Psychological Science*, *22*(6), 466–470.

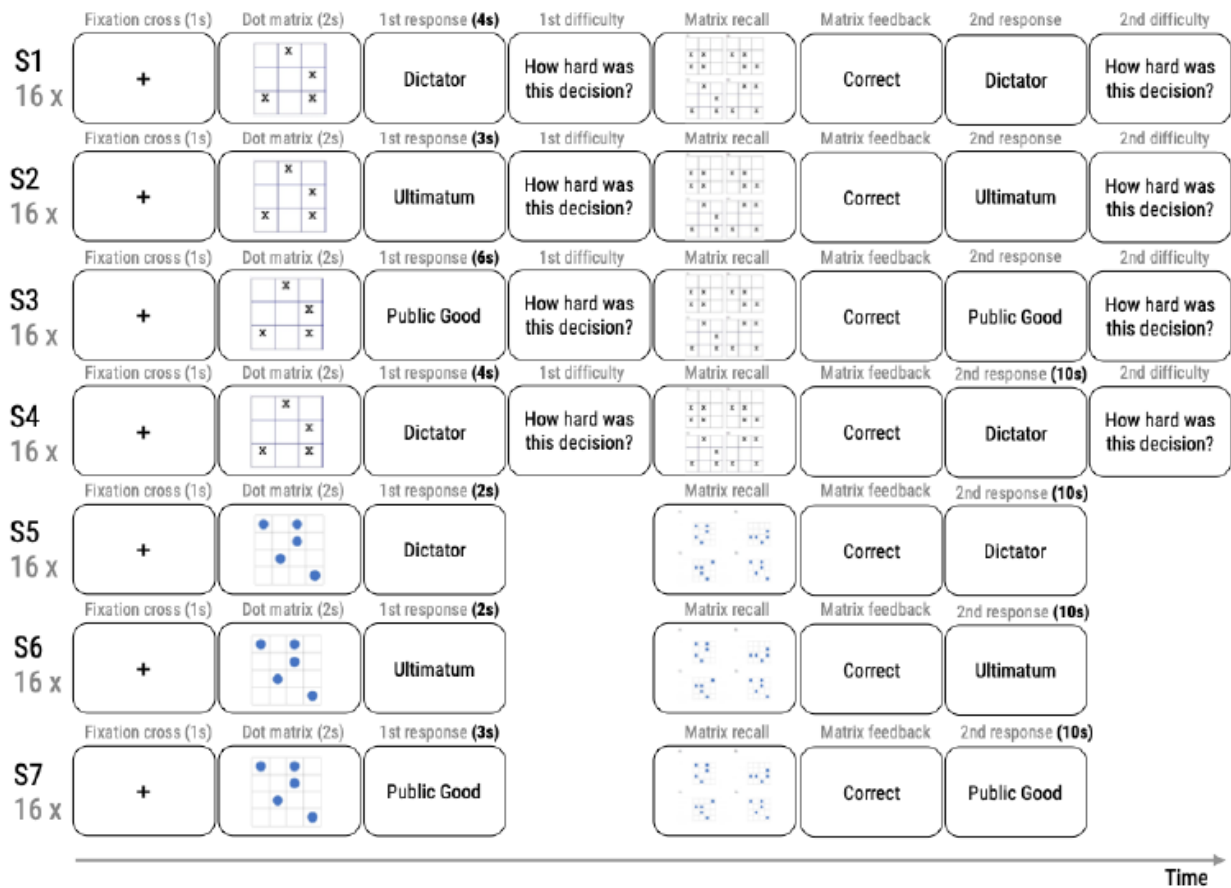


Figure 1. Overview of study procedures. In each study, each trial featured a different dot matrix to be memorized, and different payoffs for the game used in the study. See method sections for a detailed description of the games and instructions.

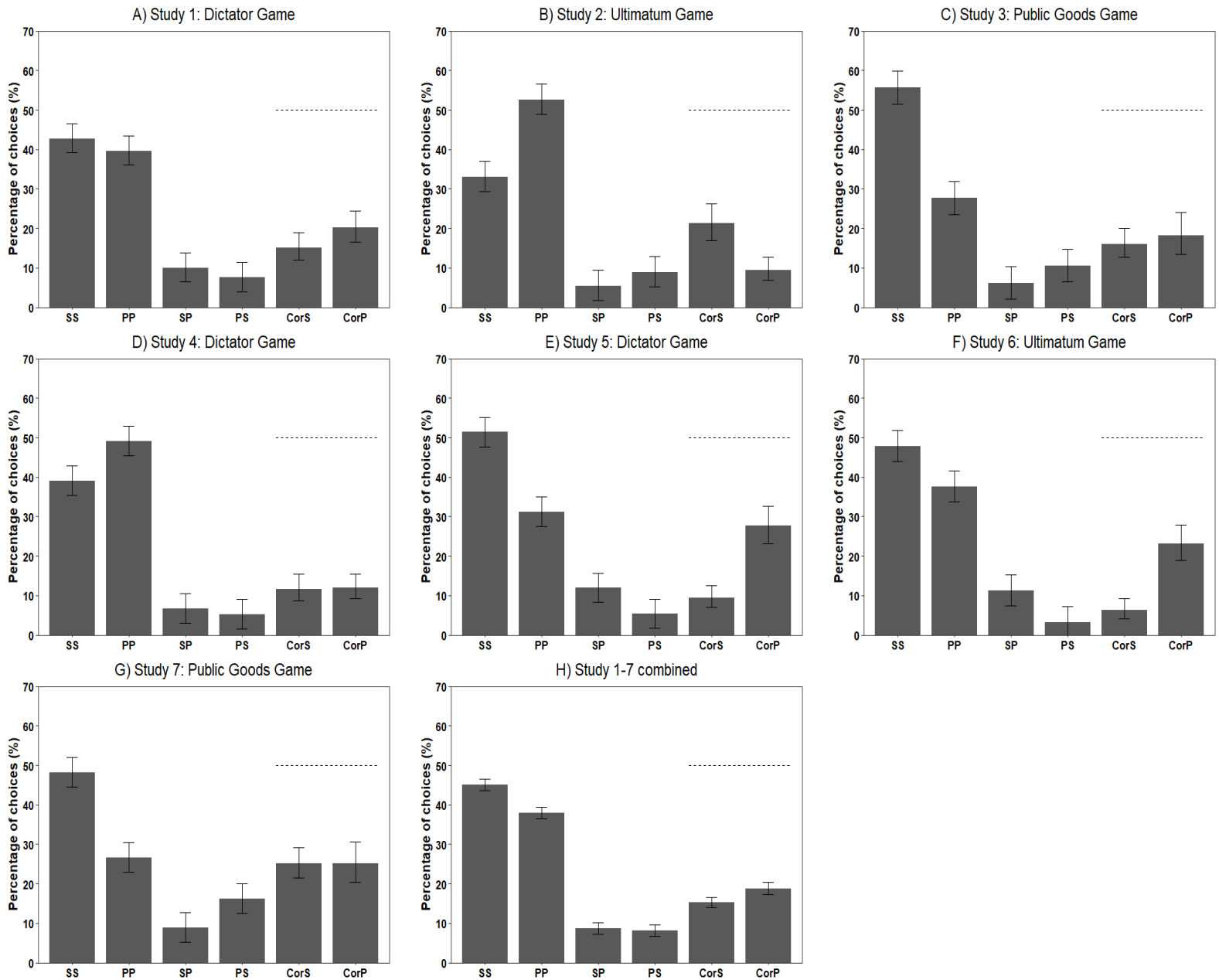


Figure 2. Overview of study results. Percentage of four possible initial-to-final choice patterns and correction indexes. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. Error bars present 95% confidence intervals.

SUPPLEMENTARY MATERIAL

A. Instructions and pay-out structure

Study 1 reading pretest

The literal instructions that were used stated the following (translated from Hungarian):

“In this experiment, your only task will be to read the tasks we will present you. You do not have to think, just read each and every sentence in the task. Then, once you finished reading, you can randomly click on one of the two responses. We simply would like to know how much time people need to read these tasks. There will be 16 questions, and the experiment will take approximately 5 minutes.”

Click on “Next” to start the experiment.”

Study 1 (Dictator Game) main two-response study

The literal instructions that were used stated the following (translated from Hungarian):

“Welcome to the experiment!

Please read these instructions carefully!

This experiment is composed of 16 questions and a couple of practice questions. It will take about 12 minutes to complete and it demands your full attention. You can only do this experiment once. In this task, we'll present you with a set of decision-making problems. We would like you to read every problem carefully and enter your response by clicking on it. We want to know what your initial, intuitive decision is and how you respond after you have thought about the problem for some more time. Hence, as soon as the problem is presented, we will ask you to enter your initial response. We want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you can enter your final response. You will have as much time as you need to indicate your second response.

After you have entered your first and final answer we will also ask you to indicate how hard you found it to make a decision.

In sum, keep in mind that it is really crucial that you give your first, initial response as fast as possible. Afterwards, you can take as much time as you want to reflect on the problem and select your final response.

Please confirm below that you read these instructions carefully and then press the "Next" button.”

After this general introduction, participants were presented with a more specific instruction page which explained them the upcoming task and informed them about the response deadline. The literal instructions were as follows:

“In this experiment, you have the opportunity to earn money up to 300 HUF. During the experiment, you can collect points. The points you earn during the experiment will be converted at the following rate: 1 point = 3 HUF.

This experiment will include 16 trials. In every trial, you will make two decisions, and at the end of the experiment we randomly choose one of your decisions and you will receive money corresponding to the number of points you earned.

In each trial, we will ask you to play a simple game in which there are always two players (Player 1 and 2). In each trial, you are going to play as player 1. In every trial, you are going to be allocated 100 points. Your task will be to divide this amount into two parts; you can decide how much points (out of the 100) you want to keep, and how much you want to give to the other player. The game is completely anonymous; the other player will never know your identity, neither will

you know who you are playing with. The players are randomly assigned to each other; in every trial, you will play with a new opponent.

In every trial, you will always have two response options; both of the options state a proposed split-up of the offered 100 points. For example:

- A) You keep 60; the other player gets 40
- B) You keep 30; the other player gets 70

You will have to click on the option you want to go with. Let's say you go with option A. At the end of the experiment we randomly choose one trial out of the 16 you will play. If this was the randomly chosen trial you will receive 60 points meaning \$0.6, while the other player will receive 40 points; which means \$0.4. Keep this in mind, while doing the experiment.

All trials and players are real. To assure this, for each participant we will randomly determine at the end of the experiment whether they will be paid out as player 1 or player 2. Half of the participants will be paid out as player 1, the other half as player 2. If you are paid out as player 1 you receive the money you decided to keep for yourself on one randomly chosen decision you made. If you are paid out as player 2, you will be randomly paired with another participant and receive the amount they gave to the other player in one of their randomly chosen trials. You will never know who the other players were and whether you were paid as player 1 or 2. The computer will randomly pair participants and select trials at the end of the experiment. At no point in time will you be informed about the identity of another participant. Likewise, the other participants will not be informed about your identity. All payments will be made anonymously. That is, the other participants do not learn how much you earned in the experiment.

As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. To assure this, a time limit was set for the first response, which is going to be 4 seconds. When there is one second left, the background colour will turn to yellow to let you know that the deadline is approaching. Please make sure to answer before the deadline passes. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind, you can enter your final response for which you will have as much time as you need.

Both your initial and final decision will be included in the lottery to determine how much money you and the other player will receive.

Press "Next" if you are ready to start the practice session!"

Pay-out structure: As it is clear from the instructions, the game was incentivized and anonymous. All participants played the game as Dictator (Player 1). To determine the pay-out, we randomly paired participants and decided who was paid as Player 1 and 2. Then, we randomly selected a decision (an initial or final response in one trial) from Player 1 and we paid both players accordingly.

Study 2 (Ultimatum Game) main two-response study

The literal instructions that were used stated the following (translated from Hungarian):

“In this experiment, you have the opportunity to earn money up to 300 HUF. During the experiment, you can collect points. The points you earn during the experiment will be converted at the following rate: 1 point = 3 HUF.

This experiment will include 16 trials. In every trial, you will make two decisions, and at the end of the experiment we randomly choose one of your decisions and you will receive money corresponding to the number of points you earned. In each trial, we will ask you to play a simple game in which there are always two players (Player A and B). In each trial you are going to play as player B. In every trial, the person playing Player A is going to be offered 100 points. Player A then will be asked to divide this amount between the two of you; he/she is allowed to make any proposal he/she wants.

Then, you will be shown Player A's proposal and you can decide to accept or reject it. If you decide to accept, both of you will get the proposed amount. If you reject, both of you will receive 0 points. The game is completely anonymous; the other player will never know your identity, neither will you know who you are playing with. The players are randomly assigned to each other; in every trial, you will play with a new opponent.

For example, Player A makes the following proposition for you:

He/she keeps 60 points, and you get 40

Do you accept this offer?

Yes

No

If you decide to reject (and hence, click on the "No" response), both of you will receive 0 points. However, if you decide to accept, you will receive 40 points and Player A will receive 60.

You will have to click on the option you want to go with. Let's say you accept this offer. At the end of the experiment we randomly choose one trial out of the 16 you will play. If this was the randomly chosen trial Player A will receive 60 points meaning 180 Forint, while the other player will receive 40 points; which means 120 Forints. Keep this in mind, while doing the experiment.

As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. To assure this, a time limit was set for the first response, which is going to be 3 seconds. When there is 1 second left, the background colour will turn to yellow to let you know that the deadline is approaching. Please make sure to answer before the deadline passes. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you enter your final response.

Both your initial and final decision will be included in the lottery to determine how much money you and the other player will receive.

At no point in time will you be informed about the identity of another participant. Likewise, the other participants will not be informed about your identity. Thus, all payments will be made anonymously. That is, the other participants do not learn how much you earned in the experiment.

After you made your choice and clicked on it, you will be automatically taken to the next page. After you have entered your first and final answer we will also ask you to indicate how hard you found it to make a decision.

Press "Next" if you are ready to start the practice session!"

Pay-out structure: As it is clear from the instructions, the game was incentivized and anonymous. All participants played the game as second-mover (Player B). However, to determine the pay-out, we randomly paired participants and decided who would be paid as Player A and B. Then, we randomly selected a decision from Player B and we paid both players accordingly. Hence, although—as in Sanfey et al. (2003)—deception was used in that the offers from Player A were predetermined, the participants' decisions could have real monetary repercussions for another player.

Study 3 (Public Good Game) main two-response study

The literal instructions that were used stated the following (translated from Hungarian):

"In this experiment, you have the opportunity to earn money up to \$1. During the experiment, you can collect points. The points you earn during the experiment will be converted at the following rate: 1 point = 0.30 cents.

This experiment will include 16 trials. In every trial, you will make two decisions, and at the end of the experiment we

randomly choose one of your decisions and you will receive money corresponding to the number of points you earned (see details below).

In each trial, we will ask you to play a simple game in which there are always two players (Player A and B). You are always going to play as player A, and you are always going to be allocated 100 points (just like Player B). You can make a decision whether to put this number of points into a “common pool”. Your opponent will be presented with the same decision. In every game, we multiply the number of points contributed to the common pool (by a factor between 1.2 and 2.9) and will divide this amount equally between the two players. In every game, you are going to be paired with a new opponent.

Imagine the following situation: You decided to contribute the 100 points you received, and your opponent decided to do so too. Thus, there will be 200 points in the common pool (100 of each player). Imagine that in this game the multiplier is 1.5. Hence, the 200 points get multiplied by 1.5 such that there will be 300 points in the common pool. This amount gets split evenly between the two of you. Hence, both of you will end up with 150 points after this game. Both of you will gain 150 points. However, imagine that the other player didn't contribute to the pool, only you did. This will mean that there will be $100 \times 1.5 = 150$ points in the common pool. The common pool always gets split in two so you will end up with 75 points in this game. Player B will also receive 75 points from the common pool in addition to their 100 points they decided to keep for themselves. Hence, Player B will make 175 points in this game. If neither one of you contributes to the pool, you both end up with your original 100 points.

The game is completely anonymous; the other player will never know your identity, neither will you know who you are playing with. The players are randomly assigned to each other; in every trial, you will play with a new opponent. All players (both A and B) are fully informed about the rules of the game.

To help you grasp the payoff structure, in every game you are going to be presented with a payoff matrix. The numbers in black (numbers before the slash) always represent what you would gain in a given situation. The numbers in grey (numbers after the slash) tell you what Player “B” would win. Here's an example:

	Player B contributes	Player B doesn't contribute
You contribute	300 / 300	150 / 250
You don't contribute	250 / 150	100 / 100

So, if you decide to contribute in this example (and hence, click on the “Yes” response), and Player B decides to contribute as well, both of you will receive 300 points. However, if you decide not to contribute, but Player B does contribute, you will receive 250 points and Player B will receive 150.

You will have to click on the option you want to go with. Let's say you contributed, but Player B didn't. At the end of the experiment we randomly choose one trial out of the 16 you will play. If this was the randomly chosen trial Player B will receive 150 points meaning \$0.45, while you will receive 250 points; which means \$0.75. Keep this in mind, while doing the experiment.

As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. To assure this, a time limit was set for the first response, which is going to be 6 seconds. When there is 1 second left, the background colour will turn to yellow to let you know that the deadline is approaching. Please make sure to answer before the deadline passes. Next, the problem will be presented again and you can take all the time you want to actively reflect on it and enter your final response.

Both your initial and final decision will be included in the lottery to determine how much money you and the other player will receive.

At no point in time will you be informed about the identity of another participant. Likewise, the other participants will not be informed about your identity. Thus, all payments will be made anonymously. That is, the other participants do not learn how much you earned in the experiment.

After you made your choice and clicked on it, you will be automatically taken to the next page.

Press "Next" if you are ready to start the practice session!"

Pay-out structure: As it is clear from the instructions, the game was incentivized and anonymous. Participants were told that each game was played with a different anonymous person and that their identity would never be revealed. To determine the pay-out, we randomly paired participants and we randomly selected a trial for each pair (either initial or final response on the trial). Participants were paid based on the decision made on that specific trial.

Study 4 forced deliberation two-response instructions

The literal instructions that were used stated the following (translated from Hungarian):

“As we told you we are interested in your initial, intuitive response. First, we want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. To assure this, a time limit was set for the first response, which is going to be 4 seconds. When there is 1 second left, the background colour will turn to yellow to let you know that the deadline is approaching. Please make sure to answer before the deadline passes. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. To make sure that you are actually reflecting on the problem, you will not be able to provide a response in the first 10 seconds. Once the 10 seconds have passed, and you have made up your mind, you can enter your final response for which you will have as much time as you need.”

B. Supplementary analyses

Combined Study 1-7 results and analyses

To get the most general picture of the results we also ran an analysis on the combined Study 1-7 data. For all statistical random effect mixed model analyses on the combined data we included the individual studies as a random effect to the model (in addition to the random effect of participants and items) whenever possible. In case we ran into singularity problems, the random effect of study was excluded (marked with *) or the random effect of study and items were both excluded (marked with **).

Selfish and prosocial correction rate test. The selfish, $b = -2.46$, $p < 0.0001$, and prosocial, $b = -1.86$, $p < 0.0001$, correction rate on the standard trials were significantly lower than 50%.

Guessing. The selection rate of the selfish/prosocial response option in the control trials was 86.7% (SD = 34) which significantly differed from chance, $b = 3.67$, $p < 0.0001$. The average standard trial stability was 77.7% (SD = 21.5) which was significantly higher than 50% chance, $t(746) = 35.25$, $p < 0.0001$.

Fast/Slow trial analysis

Based on a median split on the initial response time we calculated the correction rates for the fastest and slowest half of standard trials in each individual study. Table S3 (fast) and S4 (slow) give an overview of the results (see Table S2 for the overall figures). If our deadline was not stringent enough and allowed (some) participants to engage in deliberation in the initial response stage, one expects that the correction rate should be higher for fast vs slow initial responses (since slow responders had more time to engage in possible preemptive deliberation). Results showed this was not the case. Even for the fastest half of responses, the correction rate remained floored.

Note that given the reduced sample size in the fast/slow analysis (and moderator analyses below, i.e., half of the original trials or less), we ran the statistical mixed model test on the combined Study 1-7 data. Exploratory analyses on the individual study data pointed to frequent singularity issues or were not possible because of the need to remove the random effect of subjects. Descriptive data and CIs for each individual study can be found in the corresponding tables and indicate that the findings are robust across studies. Results of the statistical analysis showed that correction rates for both fast (CorS, $b = -7.53$, $p < 0.0001^*$; CorP, $b = -2.1$, $p < 0.0001$) and slow (CorS, $b = -2.2$, $p < 0.0001$; CorP, $b = -1.77$, $p < 0.0001$) trials were statistically lower than 50%.

Moderators

Gender. Table S2 shows the direction of chance and correction index results across gender. Table S5 (males) and S6 (females) show results for both genders separately. For the statistical analyses, we again focused on the combined Study 1-7 data. Results showed that correction rates for both females (CorS, $b = -2.3$, $p < 0.0001$; CorP, $b = -1.8$, $p < 0.0001$) and males (CorS, $b = -2.91$, $p < 0.0001^*$; CorP, $b = -2.1$, $p < 0.0001$) were statistically lower than 50%. Given that it has been suggested that gender might specifically affect correction in the Dictator Game (but not other games, Rand, 2017) we also ran a separate analysis that focused on Dictator Game studies only (Study 1, 4, and 5). Here too correction rates for both females (CorS, $b = -2.81$, $p < 0.0001^{**}$; CorP, $b = -1.75$, $p < 0.0001^{**}$) and males (CorS, $b = -2.58$, $p < 0.0001^{**}$; CorP, $b = -2.23$, $p < 0.0001^{**}$) were statistically lower than 50%.

Experience. At the end of each study we asked participants to indicate how frequently they had participated in similar experiments. Following Rand et al. (2012) we analysed results for inexperienced participants who responded they had never participated in similar studies (Table S7) and all other experienced participants (Table S8). We ran statistical analyses on the combined Study 1-7 data. The analyses showed that the correction rates for inexperienced participants (CorS, $b = -2.22$, $p < 0.0001$; CorP, $b = -1.71$, $p < 0.0001$) and experienced participants (CorS, $b = -3.36$, $p < 0.0001^*$; CorP, $b = -2.02$, $p < 0.0001$) were statistically lower than 50%.

First trial. We also ran an analysis (standard logistic regression – every participant contributed with one item to this analysis) that was restricted to the first standard trial that participants were presented with. Table S9 shows the results. Analyses showed that correction rates on the first trial (CorS, $b = -1.64$, $p < 0.0001$; CorP, $b = -1.21$, $p < 0.0001$) were statistically lower than 50%.

No load exclusion analysis

In all our studies we discarded trials on which the load memorization task was failed. This eliminates the possibility that participants neglected and traded-off the load task to deliberate during the initial economic game choice. However, although the number of discarded trials was acceptable (overall 12% of all trials), in theory, this might lead to a selection confound (Bouwmeester et al., 2017). Table S10 shows the data without application of the load exclusion criterion. Results are consistent with the original analysis (deviation average CorS = 0.1%, CorP = 1.1%).

Difficulty rating analysis

In Study 1-4 participants were asked to indicate how hard they found it to make a decision after each initial and final response (“*How hard did you find it to make a decision? Please type a number from 0 (absolutely not hard) to 100 (very hard)*”). Figure S2 shows the results for each of the three games combined in each of the two response stages on the standard trials. Our key interests concerned the initial experienced difficulty for trials in which the initial choice was corrected after deliberation or not. Previous two response studies in the logical and moral reasoning field (Bago & De Neys, 2017, 2019; Thompson et al., 2011) established that when people corrected their initial response after deliberation, they indicated to have experienced more difficulties to arrive at their initial decision (i.e., feel more conflicted or be less certain) than when they did not correct their response. Visual inspection of Figure S2 points to a similar trend. The corrective SP and PS trials are initially experienced as more difficult than the non-corrective SS and PP trials. A mixed-effect linear regression model (accounting for the random intercept of participants and items) that contrasted the experienced difficulty rating for the initial response in the non-corrective and corrective response categories indicated that this trend was significant, $\chi^2(1) = 62.87$, $p < 0.0001$, $b = 9.35$. The same pattern was observed for each of the individual games: Dictator Game (Study 1 and 4 combined), $\chi^2(1) = 24$, $p < 0.0001$, $b = 7.17$, Ultimatum Game (Study 2), $\chi^2(1) = 36.84$, $p < 0.0001$, $b = 16.81$, and the Public Goods Game (Study 3), $\chi^2(1) = 5.65$, $p = 0.018$, $b = 6.1$.

C. Supplementary References

- Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *48*, 1782-1801.
- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90-109.
- Bouwmeester, S., Verkoeijen, P. P. J. L., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., . . . Wollbrant, C. E. (2017). Registered Replication Report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, *12*, 527-542.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107-140.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755-1758.
- Rand, D. G. (2017). 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., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427-430.

A

	X	
X		X
		X

B

●		●	
		●	
	●		
			●

Figure S1. Examples of the to-be-memorized load patterns in Study 1-4 (A) and Study 5-7 (B).

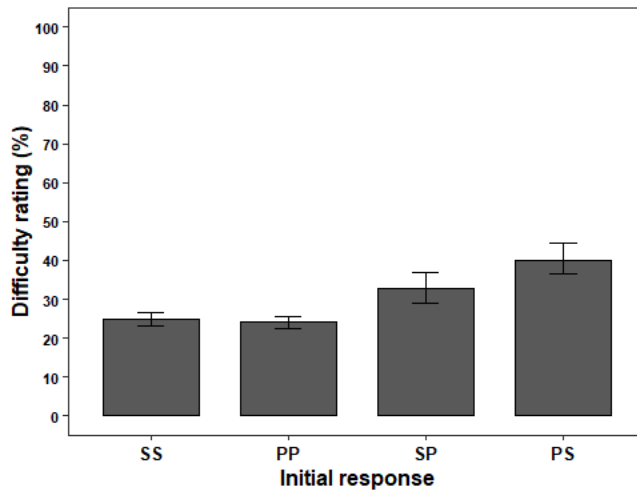
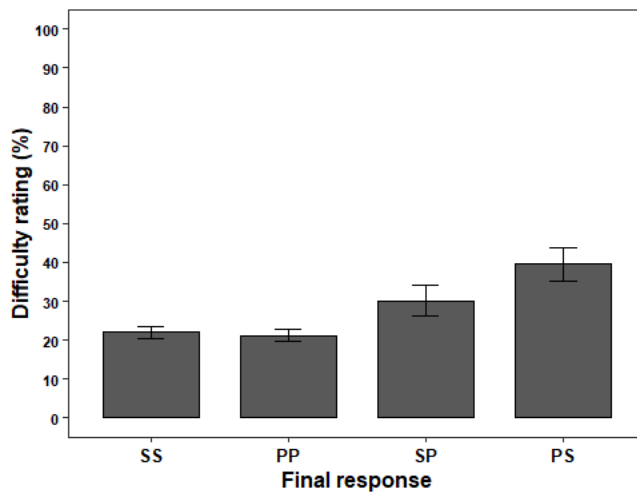
A Experienced difficulty at initial response**B Experienced difficulty at final response**

Figure S2. Average experienced difficulty on standard trials for different initial-to-final choice types at the initial (A) and final (B) response stage. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. Combined data from Study 1-4. Error bars are 95% confidence intervals.

	Standard		Control	
	Initial response	Final response	Initial response	Final response
Study 1– Dictator Game	52.8% (50)	50.4% (50)	77% (50)	78.5% (50)
Study 2 – Ultimatum Game	38.5% (48.7)	42% (49.4)	95.3% (21.2)	95.6% (20.5)
Study 3 – Public Goods Game	61.8% (48.6)	66.2% (47.3)	86% (34.7)	83.8% (36.8)
Study 4 – Dictator Game	45.7% (49.9)	44.2% (49.7)	81.2% (39.1)	81.9% (38.5)
Study 5 – Dictator Game	63.3% (48.2)	56.8% (49.6)	86.6% (34.1)	89.5% (30.7)
Study 6 – Ultimatum Game	59.2% (49.2)	51% (50)	96.7% (17.9)	98.6% (11.7)
Study 7 – Public Goods Game	57.1% (49.5)	64.4% (47.9)	85.6% (35.1)	87% (33.7)
Overall average	53.9% (49.9)	53.2% (49.9)	86.7% (34)	87.7% (32.8)

Table S1. Frequency of selfish choices on standard and control trials. On control trials selfish and prosocial considerations cue the same response. Standard deviations are in brackets.

	SS	PP	SP	PS	CorS	CorP
Study 1 – Dictator Game	42.8% [39.2; 46.5] (n=349)	39.6% [36; 43.5] (n=323)	10% [6.5; 13.8] (n=82)	7.6% [4; 11.4] (n=62)	15.1% [11.9; 18.9]	20.2% [16.6; 24.4]
Study 2 – Ultimatum Game	33% [29.3; 37.1] (n=230)	52.6% [48.9; 56.6] (n=366)	5.5% [1.7; 9.5] (n=38)	8.9% [5.2; 12.9] (n=62)	21.2% [16.9; 25.3]	9.4% [6.9; 12.6]
Study 3 – Public Goods Game	55.6% [51.5; 59.8] (n=326)	27.6% [23.5; 31.8] (n=162)	6.1% [2; 10.3] (n=36)	10.6% [6.5; 14.8] (n=62)	16% [12.7; 20]	18.2% [13.4; 24.1]
Study 4 – Dictator Game	39% [35.4; 42.9] (n=303)	49.1% [45.5; 52.9] (n=381)	6.7% [3.1; 10.5] (n=52)	5.2% [1.5; 9] (n=40)	11.7% [8.7; 15.5]	20.2% [16.6; 24.4]
Study 5 – Dictator Game	51.4% [47.4; 55.1] (n=391)	31.3% [27.6; 35] (n=238)	12% [8.3; 15.7] (n=91)	5.4% [1.7; 9.2] (n=41)	8.3% [6.1; 11.3]	27.7% [23.1; 32.7]
Study 6 – Ultimatum Game	47.8% [43.9; 51.8] (n=329)	37.7% [33.7; 41.7] (n=259)	11.3% [7.4; 15.4] (n=78)	3.2% [0; 7.2] (n=22)	6.3% [4.2; 9.3]	23.1% [19%; 27.9]
Study 7 – Public Goods Game	48.2% [44.5; 51.9] (n=365)	26.7% [23; 30.4] (n=202)	9% [5.3; 12.8] (n=68)	16.2% [12.5; 20] (n=123)	25.2% [21.6; 29.2]	25.2% [20.4; 30.7]
Overall Average	45.1% [43.7; 46.6] (n=2293)	38% [36.5; 39.5] (n=1931)	8.8% [7.3; 10.2] (n=445)	8.1% [6.6; 9.6] (n=412)	15.2% [13.9; 16.6]	18.7% [17.2; 20.4]

Table S2. Frequency of four initial-to-final choice patterns and correction indexes on standard trials. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	SS	PP	SP	PS	CorS	CorP
Study 1 a – Dictator Game	47.3% [42.2; 52.6] (n=193)	36.3% [31.4; 41.6] (n=148)	9.1% [4.2; 14.4] (n=37)	7.4% [2.5; 12.7] (n=30)	13.5% [9.6; 18.6]	20% [14.9; 26.3]
Study 2 – Ultimatum Game	37.9% [32.5; 43.4] (n=132)	52% [46.6; 57.5] (n=181)	4.9% [0; 10.4] (n=17)	5.2% [0; 10.6] (n=18)	12% [7.7; 18.2]	8.6% [5.4; 13.3]
Study 3 – Public Goods Game	49.5% [43.7; 55.7] (n=145)	35.8% [30; 42] (n=105)	6.8% [1; 13] (n=20)	7.8% [2; 14] (n=23)	13.7% [9.3; 19.7]	16% [10.6; 23.4]
Study 4 – Dictator Game	46.3% [41.1; 51.6] (n=179)	43.7% [38.5; 49] (n=169)	5.7% [0.5; 11] (n=22)	4.4% [0; 9.7] (n=17)	8.7% [5.5; 13.5]	11.5% [7.7; 16.8]
Study 5 – Dictator Game	52.3% [46.3; 56.8] (n=195)	31.1% [26.1; 36.5] (n=118)	11.6% [6.6; 17] (n=44)	6.1% [1.1; 11.5] (n=23)	10.6% [7.1; 15.3]	27.2% [20.9; 34.5]
Study 6 – Ultimatum Game	53.9% [48.7; 59.6] (n=185)	29.7% [24.5; 35.4] (n=102)	14% [8.7; 19.7] (n=48)	2.3% [0; 8] (n=8)	4.1% [2.1; 8]	32% [25.1; 39.8]
Study 7 – Public Goods Game	39.8% [34.6; 45.3] (n=151)	36.4% [31.1; 41.8] (n=138)	8.7% [3.4; 14.1] (n=33)	15% [9.8; 20.5] (n=57)	27.4% [21.8; 33.8]	19.3% [14.1; 25.9]
Overall Average	46.5% [44.4; 48.6] (n=1180)	37.9% [35.8; 40] (n=961)	8.7% [6.7; 10.8] (n=221)	6.9% [4.9; 9.1] (n=176)	13% [11.3; 14.9]	18.7% [16.6; 21]

Table S3. Frequency of four initial-to-final choice patterns and correction indexes for fastest half of initial standard trials. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	SS	PP	SP	PS	CorS	CorP
Study 1 a – Dictator Game	38.2% [33.3; 43.6] (n=156)	42.9% [38; 48.3] (n=175)	11% [6.1; 16.4] (n=45)	7.8% [2.9; 13.2] (n=32)	17% [12.3; 23]	20.5% [15.7; 26.3]
Study 2 – Ultimatum Game	28.2% [23; 33.8] (n=98)	53.2% [48; 58.8] (n=185)	6% [0.9; 11.7] (n=21)	12.6% [7.5; 18.3] (n=44)	31% [24; 39]	10.2% [6.8; 15.1]
Study 3 – Public Goods Game	61.8% [56.3; 67.4] (n=181)	19.5% [14; 25] (n=57)	5.5% [0; 11.1] (n=16)	13.3% [7.8; 18.9] (n=39)	17.7% [13.2; 23.2]	21.9% [14; 32.7]
Study 4 – Dictator Game	31.9% [27; 37.2] (n=124)	54.5% [49.6; 59.8] (n=212)	7.7% [2.8; 13] (n=30)	5.9% [1; 11.2] (n=23)	15.6% [10.7; 22.4]	12.4% [8.8; 17.2]
Study 5 – Dictator Game	51.4% [46.5; 56.9] (n=196)	31.5% [26.5; 37] (n=120)	12.3% [7.3; 17.8] (n=47)	4.7% [0; 10.2] (n=18)	8.4% [5.4; 12.9]	28.1% [21.9; 35.4]
Study 6 – Ultimatum Game	41.7% [36.2; 47.4] (n=144)	45.5% [40; 51.1] (n=157)	8.7% [3.2; 14.3] (n=30)	4.1% [0; 9.7] (n=14)	8.9% [5.4; 14.3]	16% [11.5; 22]
Study 7 – Public Goods Game	56.5% [51.5; 61.5] (n=214)	16.9% [11.9; 21.9] (n=64)	9.2% [4.2; 14.2] (n=35)	17.4% [12.4; 22.4] (n=66)	23.6% [19; 28.9]	35.4% [26.6; 45.2]
Overall Average	43.8% [41.7; 45.9] (n=1113)	38.1% [36.1; 40.2] (n=970)	8.8% [6.7; 10.9] (n=224)	9.3% [7.2; 11.4] (n=236)	17.5% [15.6; 19.6]	18.8% [16.6; 21.1]

Table S4. Frequency of four initial-to-final choice patterns and correction indexes for slowest half of initial standard trials. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	SS	PP	SP	PS	CorS	CorP
Study 1 – Dictator Game	43% [36.5; 49.7] (n=105)	42.2% [35.7; 48.9] (n=103)	8.6% [2; 15.3] (n=21)	6.1% [0; 12.8] (n=15)	12.5% [7.7; 19.6]	16.9% [11.4; 24.5]
Study 2 – Ultimatum Game	32.9% [25.3; 41] (n=52)	57% [49.4; 65.1] (n=90)	3.8% [0; 11.9] (n=6)	6.3% [0; 14.4] (n=10)	16.1% [9; 27.2]	6.2% [2.9; 13]
Study 3 – Public Goods Game	69.1% [62.9; 76.2] (n=121)	20.6% [14.3; 27.6] (n=36)	3.4% [0; 10.5] (n=6)	6.9% [0.6; 13.9] (n=12)	9% [5.2; 15.1]	14.3% [6.7; 27.8]
Study 4 – Dictator Game	32.9% [26.4; 39.6] (n=76)	54.5% [48.1; 61.2] (n=126)	6.9% [0.4; 13.6] (n=16)	5.6% [0; 12.3] (n=13)	14.6% [8.7; 23.4]	11.3% [7.1; 17.5]
Study 5 – Dictator Game	54.9% [49.9; 60] (n=218)	32% [27; 37.1] (n=127)	7.6% [2.5; 12.6] (n=30)	5.5% [0.5; 10.6] (n=22)	9.2% [6.1; 13.5]	19.1% [13.7; 26]
Study 6 – Ultimatum Game	46.8% [41.4; 52.2] (n=174)	36.6% [31.2; 42] (n=136)	14.5% [9.1; 19.9] (n=54)	2.2% [0; 7.6] (n=8)	4.4% [2.2; 8.4]	28.4% [22.5; 35.2]
Study 7 – Public Goods Game	55.6% [50.3; 61] (n=190)	22.8% [17.5; 28.2] (n=78)	7.3% [2; 12.7] (n=25)	14.3% [9.1; 19.8] (n=49)	20.5% [15.9; 26.1]	24.3% [17; 33.4]
Overall Average	48.8% [46.4; 51.2] (n=936)	36.3% [33.9; 38.7] (n=696)	8.2% [5.9; 10.7] (n=158)	6.7% [4.4; 9.1] (n=129)	12.1% [10.3; 14.2]	18.5% [16; 21.2]

Table S5. Frequency of four initial-to-final choice patterns and correction indexes on standard trials for males. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	SS	PP	SP	PS	CorS	CorP
Study 1 – Dictator Game	42.7% [38.5; 47.2] (n=244)	38.5% [34.3; 43] (n=220)	10.7% [6.5; 15.2] (n=61)	8.2% [4; 12.7] (n=47)	16.2% [12.4; 20.8]	21.7% [17.3; 26.9]
Study 2 – Ultimatum Game	33.1% [28.8; 37.6] (n=178)	51.3% [47; 55.9] (n=276)	6% [1.7; 10.5] (n=32)	9.7% [5.4; 17.2] (n=52)	22.6% [17.7; 28.4]	10.4% [7.5; 14.3]
Study 3 – Public Goods Game	49.9% [45; 55.1] (n=205)	30.7% [25.8; 35.9] (n=126)	7.3% [2.4; 12.5] (n=30)	12.3% [7.3; 17.4] (n=50)	19.6% [15.2; 24.9]	19.2% [13.8; 26.1]
Study 4 – Dictator Game	41.7% [37.2; 46.1] (n=227)	46.8% [42.4; 51.3] (n=255)	6.6% [2.2; 11.1] (n=36)	5% [0.6; 9.4] (n=27)	10.6% [7.4; 15]	12.4% [9.1; 16.7]
Study 5 – Dictator Game	47.5% [42.3; 53.1] (n=173)	30.5% [25.3; 36.1] (n=111)	16.8% [11.5; 22.4] (n=61)	5.2% [0; 10.8] (n=19)	9.9% [6.4; 14.9]	35.5% [28.7; 42.9]
Study 6 – Ultimatum Game	48.2% [42.4; 54.1] (n=149)	39.5% [33.7; 45.4] (n=122)	7.9% [1.9; 13.6] (n=24)	4.5% [0; 10.4] (n=14)	8.6% [5.2; 13.9]	16.4% [11.3; 23.3]
Study 7 – Public Goods Game	42.1% [37; 47.2] (n=175)	29.8% [24.8; 34.9] (n=124)	10.3% [5.3; 15.5] (n=43)	17.8% [12.7; 22.9] (n=74)	19.7% [24.4; 35.7]	25.7% [19.7; 32.9]
Overall Average	42.8% [41; 44.7] (n=1351)	39.1% [37.2; 41] (n=1234)	9.1% [7.2; 11] (n=287)	9% [7.1; 10.9] (n=283)	17.3% [15.6; 19.2]	18.9% [17; 20.9]

Table S6. Frequency of four initial-to-final choice patterns and correction indexes on standard trials for females. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	N	SS	PP	SP	PS	CorS	CorP
Study 1 – Dictator Game	92	44.1% [40.1; 48.2] (n=297)	39.1% [35.1; 43.1] (n=263)	9.8% [5.8; 13.8] (n=66)	7% [3; 11] (n=47)	13.7% [10.4; 17.7]	20.1% [16.1; 24.7]
Study 2 – Ultimatum Game	74	33.1% [28.8; 37.7] (n=176)	52.7% [48.4; 57.3] (n=280)	5.5% [1.1; 10] (n=29)	8.7% [4.3; 13.2] (n=46)	20.7% [15.9; 26.6]	9.4% [6.6; 13.2]
Study 3 – Public Goods Game	65	56.5% [51.9; 61.2] (n=261)	27.1% [22.5; 31.8] (n=125)	5.2% [0.6; 9.9] (n=24)	11.3% [6.7; 16] (n=52)	16.6% [12.9; 21.1]	16.1% [11.1; 22.8]
Study 4 – Dictator Game	80	38.4% [34.2; 42.9] (n=216)	50% [45.7; 54.4] (n=281)	7.5% [3.2; 11.9] (n=42)	4.1% [0; 8.5] (n=23)	9.6% [6.5; 14]	13% [9.8; 17.1]
Study 5 – Dictator Game	53	56.5% [51.2; 61.8] (n=192)	20.9% [15.6; 26.2] (n=71)	15.6% [10.3; 20.9] (n=53)	7.1% [1.8; 12.4] (n=24)	11.1% [7.6; 16]	42.7% [34.4; 51.5]
Study 6 – Ultimatum Game	56	38.1% [32.7; 44.1] (n=127)	44.4% [39; 50.4] (n=148)	13.2% [7.8; 19.2] (n=44)	4.2% [0; 10.2] (n=14)	9.9% [6; 16]	22.9% [17.5; 29.4]
Study 7 – Public Goods Game	58	45.4% [40.3; 50.7] (n=178)	25.8% [20.7; 31.1] (n=101)	8.9% [3.8; 14.3] (n=35)	19.9% [14.8; 25.2] (n=78)	30.5% [25.2; 36.4]	25.7% [19.1; 33.7]
Overall Average	478	43.9% [42.1; 45.8] (n=1447)	38.5% [36.7; 40.4] (n=1269)	8.9% [7.1; 10.7] (n=293)	8.6% [6.8; 10.5] (n=284)	16.4% [14.7; 18.2]	18.8% [16.9; 20.8]

Table S7. Frequency of four initial-to-final choice patterns and correction indexes on standard trials for inexperienced participants. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval. N is the number of inexperienced participants in the given study.

	N	SS	PP	SP	PS	CorS	CorP
Study 1 – Dictator Game	20	36.4% [28; 45.1] (n=52)	42% [33.6; 50.7] (n=60)	11.2% [2.8; 19.9] (n=16)	10.5% [2.1; 19.2] (n=15)	22.1% [14.1; 33.7]	21.1% [13.4; 31.5]
Study 2 – Ultimatum Game	24	32.7% [25.5; 41.1] (n=54)	52.1% [44.8; 60.5] (n=86)	5.5% [0; 13.8] (n=9)	9.7% [2.4; 18] (n=16)	22.9% [14.6; 34]	9.5% [5.1; 17]
Study 3 – Public Goods Game	17	52.4% [43.5; 61.3] (n=65)	29.8% [21; 38.8] (n=37)	9.7% [0.8; 18.6] (n=12)	8.1% [0; 17] (n=10)	13.3% [7.4; 22.8]	24.5% [14.6; 38.1]
Study 4 – Dictator Game	31	40.7% [34.1; 48.1] (n=87)	46.7% [40.2; 54.2] (n=100)	4.7% [0; 12.1] (n=10)	7.9% [1.4; 15.4] (n=17)	16.3% [10.5; 24.6]	9.1% [5; 15.9]
Study 5 – Dictator Game	65	47.3% [42.3; 52.4] (n=199)	39.7% [34.7; 44.8] (n=167)	9% [4; 14.1] (n=38)	4% [0; 9.1] (n=17)	7.9% [5; 12.2]	18.5% [13.8; 24.4]
Study 6 – Ultimatum Game	59	56.3% [51.1; 61.9] (n=196)	31.6% [26.4; 37.2] (n=110)	9.8% [4.6; 15.3] (n=34)	2.3% [0; 7.9] (n=8)	3.9% [2; 7.5]	23.6% [17.4; 31.2]
Study 7 – Public Goods Game	54	51.1% [45.9; 56.5] (n=187)	27.6% [22.4; 33] (n=101)	9% [3.8; 14.4] (n=33)	12.3% [7.1; 17.7] (n=45)	19.4% [14.8; 25]	24.6% [18.1; 32.6]
Overall Average	270	47.2% [44.7; 49.7] (n=840)	37.1% [34.6; 39.6] (n=661)	8.5% [6.1; 11] (n=152)	7.2% [4.7; 9.7] (n=128)	13.2% [11.2; 15.5]	18.7% [16.2; 21.5]

Table S8. Frequency of four initial-to-final choice patterns and correction indexes on standard trials for experienced participants. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval. N is the number of inexperienced participants in the given study.

	SS	PP	SP	PS	CorS	CorP
Study 1 a – Dictator Game	42.9% [33.9; 53.2] (n=48)	39.3% [30.4; 49.6] (n=44)	9.8% [0.9; 20.1] (n=11)	8% [0; 18.3] (n=9)	15.8% [8.5; 27.4]	20% [11.6; 32.4]
Study 2 – Ultimatum Game	36.7% [27.6; 47.8] (n=36)	49% [39.8; 60] (n=48)	7.1% [0; 18.2] (n=7)	7.1% [0; 18.2] (n=7)	16.3% [8.1; 30]	12.7% [6.3; 24]
Study 3 – Public Goods Game	50% [40.2; 62] (n=41)	25.6% [16.9; 37.6] (n=21)	8.5% [0; 20.5] (n=7)	15.9% [6.1; 27.8] (n=13)	24.1% [14.6; 36.9]	25% [12.7; 43.4]
Study 4 – Dictator Game	36% [27; 46.2] (n=40)	47.8% [38.7; 57.9] (n=53)	9% [0; 19.2] (n=10)	7.2% [0; 17.4] (n=8)	16.7% [16.7; 29.6]	15.9% [8.9; 26.8]
Study 5 – Dictator Game	50% [41.5; 59.9] (n=59)	25.4% [16.9; 35.3] (n=30)	16.9% [8.5; 26.8] (n=20)	7.6% [0; 17.5] (n=9)	13.2% [7.1; 23.3]	40% [27.6; 53.8]
Study 6 – Ultimatum Game	47% [38.5; 57.3] (n=55)	38.5% [29.9; 48.7] (n=45)	13.7% [5.1; 23.9] (n=16)	0.9% [0; 11.1] (n=1)	1.8% [0.3; 9.4]	26.2% [16.8; 38.4]
Study 7 – Public Goods Game	50.9% [42; 60.7] (n=57)	25% [16.1; 34.8] (n=28)	8% [0; 17.9] (n=9)	16.1% [7.1; 25.9] (n=18)	24% [15.8; 34.8]	24.3% [13.4; 40.1]
Overall Average	44.8% [41.1; 48.7] (n=336)	35.9% [32.1; 39.7] (n=269)	10.7% [6.9; 14.5] (n=80)	8.7% [4.9; 12.5] (n=65)	16.2% [12.9; 20.1]	22.9% [18.8; 27.6]

Table S9. Frequency of four initial-to-final choice patterns and correction indexes on the first standard trial. First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.

	SS	PP	SP	PS	CorS	CorP
Study 1 a – Dictator Game	43.1% [39.6; 46.7] (n=380)	38.7% [35.2; 42.2] (n=341)	10.1% [6.6; 13.6] (n=89)	8.1% [4.5; 11.6] (n=71)	15.7% [12.7; 19.4]	20.7% [17.1; 24.8]
Study 2 – Ultimatum Game	32.9% [29.2; 36.7] (n=249)	52.4% [48.7; 56.1] (n=396)	6.2% [2.5; 10] (n=47)	8.5% [4.8; 12.2] (n=64)	20.4% [16.4; 25.3]	10.6% [8.1; 13.8]
Study 3 – Public Goods Game	56.2% [52.3; 60.2] (n=363)	27.4% [23.5; 31.4] (n=177)	5.9% [2; 9.9] (n=38)	10.5% [6.7; 14.5] (n=68)	15.8% [12.6; 19.5]	17.7% [13.2; 23.3]
Study 4 – Dictator Game	38.6% [35.2; 42.2] (n=333)	49.3% [45.8; 52.9] (n=425)	6.7% [3.2; 10.3] (n=58)	5.3% [1.9; 8.9] (n=46)	12.1% [9.2; 15.8]	12% [9.4; 15.2]
Study 5 – Dictator Game	50.8% [47.4; 54.3] (n=444)	31% [27.6; 34.5] (n=271)	12.4% [8.9; 15.9] (n=108)	5.8% [2.4; 9.4] (n=51)	10.3% [7.9; 13.3]	28.5% [24.2; 33.2]
Study 6 – Ultimatum Game	47.4% [44.1; 51.5] (n=387)	36.6% [33; 40.4] (n=297)	12.5% [8.9; 16.2] (n=101)	3.2% [0; 7%] (n=26)	6.3% [4.3; 9.1]	25.4% [21.4; 29.9]
Study 7 – Public Goods Game	47.7% [44.2; 51.2] (n=418)	26.6% [23.1; 30.1] (n=233)	9.9% [6.5; 13.5] (n=87)	15.8% [12.4; 19.4] (n=139)	25% [21.5; 28.7]	27.2% [22.6; 32.3]
Overall Average	45.1% [43.7; 46.5] (n=2570)	37.5% [36.1; 38.9] (n=2140)	9.3% [7.9; 10.7] (n=528)	8.1% [6.8; 9.6] (n=465)	15.3% [14.1; 16.6]	19.8% [18.3; 21.3]

Table S10. Frequency of four initial-to-final choice patterns and correction indexes on standard trials when trials with incorrect response on the load question were not excluded.

First/second letter refers to initial/final choice. S = Selfish choice. P = Prosocial choice. CorS = Selfish correction rate, CorP = Prosocial correction rate. n = number of trials. Values between square brackets present the 95% confidence interval.