ONLINE SUPPLEMENTARY MATERIALS

FOR THE PAPER

"ALTRUISM, FAST AND SLOW? EVIDENCE FROM A META-ANALYSIS AND A NEW EXPERIMENT"

BY HANNA FROMELL, DANIELE NOSENZO, AND TRUDY OWENS

- Appendix A contains details of the design of the meta-analysis presented in Section 2 of the paper.
- Appendix B contains details of the procedures used in the experiment presented in Section 3.
- Appendix C contains the instructions used in the experiment of Section 3.
- Appendix D contains additional analysis of the experimental data.

Appendix A: Meta-Study

In this appendix, we discuss additional design details and results of the meta-study. Table A.1 provides a list of studies included in the meta-study. For each study, the table reports the number of experiments contained in the study and the number of subjects who participated in those experiments, the method used to manipulate cognitive resources (cognitive load, priming, ego depletion or time pressure), and the type of game used to measure altruism (dictator game or donation game). Table A.2 contains descriptive statistics of the variables used in the mediator analysis.

A.1 Design

We searched the literature for studies to include in the meta-study, initially using Google Scholar and New Economic Papers (NEP) alerts, and then following up on relevant papers cited in the set of papers that we initially found in our literature search.[1] We looked for experimental studies that manipulated cognitive resources of participants using one of the four standard types of interventions, i.e. cognitive load, priming, ego depletion, or time pressure. Since our focus is on altruistic behavior, we restricted the search to experiments in which subjects were asked to allocate money between themselves and another passive player (i.e., dictator games or donation games). We excluded all types of experimental games in which the recipient of the money allocated by the decision-maker was not completely passive (i.e., we excluded ultimatum games; ring measures of Social Value Orientation; trust games; etc.). We allow designs that use role uncertainty (both the decision-maker and the passive player make one decision, but only one decision, randomly selected at the end of the experiment, counts for payment).

We require that the allocation decisions have real monetary consequences for participants, i.e. we only include non-hypothetical studies. We also require that the studies follow the methodology of experimental economics (although we do not restrict the analysis to economics studies), in particular that they do not use deception in their designs. This excludes a handful of studies, including two studies previously incorporated in the meta-study by Rand et al. (2016).[2]

[1] In addition, we further cross-checked our literature search with the recent literature review of the dual-process literature by Capraro (2019).

[2] The excluded studies due to use of hypothetical stakes or deception are: Janssen et al. (2008); Cornelissen et al. (2011); Xu et al. (2012); Kinnunen and Windmann (2013); and one treatment of Banker et al. (2017).

TABLE A.1

List of studies included in the meta-study

Study	Acronym	N	Method	N. experiments contributed to study	Game
Benjamin et al. 2013 (*)	BBS	97	CL	2 (pilot / main study)	DG
Grossman and v.d. Weele 2017 (*)	GVdW	224	CL	2 (high / low affect)	Donation
Hauge et al. 2016 (*)	HBJJS	192	CL	5 (exp 1 / exp 2 give / exp 2 take / exp 3 give / exp 3 take)	DG / Donation
Kessler and Meier. 2014 (*)	KM	405	CL	5 (study 1 / replication 1-4)	Donation
Schulz et al. 2014 (*)	SFTU	135	CL	1	DG
Banker et al. 2017	BABAV	459	PR	2 (exp 2 / exp 4b)	DG
Rand et al. 2016 (*)	RBECB	2,788	PR	7 (study 1 D / E / F / G / H / V / study 2)	DG
Achtziger et al. 2015 (*)	ACW	128	DPL	1	DG
Banker et al. 2017	BABAV	795	DPL	3 (exp 1 / exp 3 low / high anchor)	DG
Friehe and Schildberg-Hörisch 2017	FSH	90	DPL	1	DG
Halali et al. 2013 (*)	HBMO	68	DPL	1	DG
Itzchakov et al. 2018 (*)	IUW	104	DPL	2 (with / without persuasion message)	Donation
Tinghög et al. 2016 (*)	TABJKKW	309	DPL/ CL	1	Donation
Andersen et al. 2018 (*)	AGKM	208	TP	2 (high / low stakes)	DG
Capraro et al. 2017 (*)	CCEHG	460	TP	2 (India / USA)	DG
Chen and Krajbich 2018 (*)	CK	102	TP	1	DG
Gärtner (2018) (*)	G	1501	TP	3 (no / selfish / prosocial status quo)	DG
Grolleau et al. 2018	GSEHJ	223	TP	2 (short / long time delay)	DG
Jarke and Lohse 2016 (*)	JL	190	TP	2 (Heidelberg / Hamburg)	DG
Krawczyk and Sylwestrzak 2018 (*)	KS	68	TP	1	DG
Mrkva 2017 (*)	M	357	TP	3 (study 1 / study2 low stakes / study2 high stakes)	Donation
Merkel and Lohse 2018 (*)	ML	176	TP	2 (weak / strong time pressure)	DG
Rand et al 2016 (*)	RBECB	1,025	TP	5 (study 1 A / B / C / I / J)	DG
Strømland and Torsvik 2019 (*)	ST	1,368	TP	1	DG
Tinghög et al. 2016 (*)	TABJKKW	1,102	TP	3 (Austria / Sweden / USA)	Donation

Note: In a few cases, the same study reports data from interventions using different manipulations of cognitive resources (e.g. time pressure and priming, Rand et al., 2016). In those cases, we report the same study more than once in the table. N = number of subjects involved in the experiment; CL = cognitive load; PR = priming; DPL = ego depletion; TP = time pressure. DG = dictator game, Donation = donation experiment. "Acronym" refers to the acronym used in Figure 1 in the main paper. (*) indicates that we obtained the raw data of the study or additional statistics not included in the original paper (e.g. results disaggregated by gender).

TABLE A.2
Descriptive statistics of variables used in mediator analysis

Variable	Mean	N of experiments for which data is available	N of subjects
Stakes of the experiment (in 2017 USD)	4.34	60	12,574
Sample		60	12,574
1 if students	0.55		
1 if AMT workers	0.33		
Location of experiment		60	12,574
1 if US	0.52		
1 if Europe	0.38		
Type of game		60	12,574
1 if donation game	0.28		
1 if dictator game	0.72		
Frame of the game		60	12,574
1 if give	0.87		
1 if take	0.13		
Type of intervention		60	12,574
1 if cognitive load	0.26		
1 if time pressure	0.44		
1 if ego depletion	0.14		
1 if priming	0.16		
Gender of participants		50	10,728
1 if female	0.50		

Note: The means are computed at the experiment level, except for gender where we report the fraction of female participants in the 50 experiments for which we have data. Stakes are computed as the maximum nominal payoff available to the decision-maker multiplied by the probability a subject is actually paid for that decision, converted to 2017 USD using PPP.

We refer to the unit of observation in the meta-analysis as an "experiment". Many studies contribute more than one experiment to our meta-study. This is because some of the included studies investigate the interaction between the cognitive resource manipulation and some other treatment variable (e.g., the frame of the game, Banker et al., 2017; the size of the stakes, Andersen et al., 2018; etc.), and so report results from more than one experiment. Other studies report several independent experiments conducted at different points in time, with different subject pools, or in different countries. Finally, in a few cases, the same study reports different experiments using different types of manipulations of cognitive resources. In a few cases, the studies are based on multiple rounds or on different versions of an underlying game (e.g. games with different payoff configurations). In these cases we aggregate across the rounds of the

experiment (e.g., Achtziger, et al., 2015) or across the different versions of the game (e.g., Schulz et al., 2014) and use only one observation per study in the analysis.[3]

The meta-study focuses on the effect of manipulating cognitive resources on altruism. We measure altruism as the amount (or fraction of endowment) that the decision-maker gives to the passive player. In games where decision-makers face binary choices (give / not give) we measure altruism as the fraction of decision-makers sacrificing own payoff to increase the passive player's payoff. In order for a study to be included in the meta-analysis we must be able to compute this measure of altruism, either from the statistics, tables and graphs reported in the paper, or from its raw data. We exclude three studies because we could not retrieve this information.[4]

For each study, we derive two measurements of altruism, one for the treatment condition where cognitive resources were manipulated to promote intuitive responses (e.g., treatments where decisions had to be taken quickly in time pressure studies), and one for the condition that promoted deliberation. To quantify the effect of promoting intuition vs. deliberation on altruism, we calculate the standardized mean difference (Cohen's d) in altruism between these two conditions. To account for small sample bias, we apply Hedges' correction to the computed Cohen's d and associated standard error.[5, 6] In all cases, a positive value of Cohen's d indicates

[3] In one case (Merkel and Lohse, 2018) we only use one of the four games reported in the paper (the "Medium" game). This is because Merkel and Lohse argue that in the other three games the "fairness is intuitive" hypothesis is confounded with another possible explanation (perceived difficulty in making a choice).

[4] The three studies are: Small et al. (2007), Strombach et al. (2016), and Balafoutas et al. (2018).

[5] In most cases, the intuitive/deliberative conditions were administered between subjects and we use the formulas provided in Fritz et al. (2012) to compute $d = \left[\frac{M_I - M_D}{\sqrt{\frac{(N_I-1)*SD_I^2+(N_D-1)*SD_D^2}{N_I+N_D-2}}}\right] * \left[1 - \frac{3}{4*(N_I+N_D-2)-1}\right]$ and its standard error $se(d) = \sqrt{\frac{N_I+N_D}{N_I*N_D} + \frac{d^2}{2*(N_I+N_D)}}$, where M_I and M_D is the mean altruism in the intuitive and deliberative conditions respectively, and SD_I SD_D, N_I and N_D are the associated standard deviations and samples sizes. Two studies (Chen and Krajbich 2018; Krawczyk and Sylwestrzak 2018) use within-subject designs. In these cases, we use the method discussed by Bonett (2015) and compute $d_{within} = \left[\frac{M_I - M_D}{SD_D}\right] * \left[1 - \frac{3}{4*(N-1)-1}\right]$ and its standard error $se(d_{within}) = \sqrt{\frac{d_{within}^2}{2*(N-1)} + \frac{2*(1-\rho)}{N}}$, where N is the sample size and ρ is the sample Pearson correlation between the paired scores.

[6] In nearly all cases, the effect sizes are computed using pairs of treatment/control groups that are independent from one another. In two cases (Grolleau et al., 2018; Merkel and Lohse, 2018), the same control group is compared against two treatment groups. In another case (Hauge et al., 2016, study 2) we have two effect sizes computed using the same group of participants who played in two different game frames (give and take). In these cases, to avoid the unit-of-analysis bias that would arise from "double-counting" participants from shared groups, we follow the recommended practice and split the observations from the shared groups evenly among the effect sizes (see Higgins and Green, 2011 for details).

that promoting intuition triggered more altruistic behavior relative to the deliberative condition, while a negative value indicates the opposite.

In some of the studies using time pressure manipulations (Andersen et al., 2018, Gärtner, 2018 and Capraro et al., 2017), the reported analysis is based on sub-samples that only include participants who complied with the time pressure/delay requirements (e.g., under time pressure, subjects who took longer than the allotted time to make decisions are excluded from the analysis). This procedure implies that the analysis relies on self-selected samples, which has been criticized (see, e.g., Bouwmeester et al., 2017). In our meta-analysis, in all cases except one, we compute effect sizes using the full samples (i.e., including subjects who failed to comply with the time pressure/delay requirements).[7] The exception is Andersen et al. (2018): in their time-delay condition, subjects were given an extra day to make a decision, but 13 subjects did not show-up to the experiment on the second day and therefore, due to missing data, they are excluded from their (and our) analysis.[8] There are four further cases of data exclusions in studies using other types of manipulations of cognitive resources:

1) In the cognitive load/depletion study by Tinghög et al. (2016), 2 out of 311 subjects (0.6%) have missing dictator decisions and so are excluded from our analysis.
2) In the cognitive load study by Schulz et al. (2014), subjects had 20 seconds to make a choice in each of the 20 dictator games that they played; after that the observation was recorded as missing. In their data, 2.57% of choices are missing and so excluded from our analysis (0.88% in the low-load treatment and 4.26% in the high-load treatment – including 1 subject who never made choices in any of the 20 games).
3) In the ego depletion study of Halali et al. (2013), subjects in the ego depletion condition had to write a story avoiding the use of certain letters of the alphabet. They exclude from the analysis 8 subjects who did not follow the instructions and used forbidden letters and 1 subject who did not write any text. Moreover, they exclude 5 further subjects because they reported not to believe the game involved a real person in the role of recipient. In total, they exclude 20.6% of their observations. However, Halali et al.

[7] Capraro et al. (2017) and Gärtner (2018) provided their datasets, including subjects who are not compliant with the time pressure/delay requirements, and so our analysis is based on their full samples.
[8] There are two other cases of missing data that are *not* related to the time pressure/delay conditions: Mrkva (2017) reports that 10 subjects dropped out from her MTurk experiment before being assigned to a treatment condition. Strømland and Torsvik (2019) also report missing data for 22 subjects who dropped out from the study.

(2013) provided us with a dataset that comprises the observations they had excluded, and so our meta-analysis is based on their full sample.

4) Finally, Banker et al. (2017) exclude 3 out of 462 subjects (0.6%) from their priming manipulation study (involving text reading and comprehension) because they were non-native English speakers. Moreover, they exclude 520 out of 1315 subjects (39.5%) from their ego depletion study due to either attrition during the study (the experiment was run on MTurk; 107 subjects), failure to pass an attention check performed prior to treatment assignment (408 subjects), technical problems (4 subjects), or missing data (1 subject). We did not manage to recover the full data for this study, and so we are forced to base our meta-study on the reduced sample in this case.

A.2 Additional results

In Table A.3 we report additional results of the mediator analysis performed in the main paper. We report meta-regressions that are analogous to those reported in columns (1) to (6) of Table 1 of the main paper, but without including study fixed effects. The conclusions from this additional analysis are the same as those reported in the paper, except for the frame of the game variable (column 5). As already discussed in the paper in relation to the regressions reported in columns (7) and (8) of Table 1, we find a significantly positive effect of the take frame ($p = 0.035$) when we do not control for study fixed effects.

TABLE A.3

Mediator analysis: random-effects meta-regressions

Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Stakes	-0.000 (0.004)					
1 if students		-0.017 (0.090)				
1 if AMT workers		-0.029 (0.090)				
1 if exp. run in USA			-0.123 (0.092)			
1 if exp. run in Europe			-0.120 (0.093)			
1 if donation game				-0.003 (0.068)		
1 if take frame					0.177** (0.082)	
1 if cognitive load						0.037 (0.081)
1 if ego depletion						0.000 (0.088)
1 if priming						0.038 (0.082)
Study FE?	No	No	No	No	No	No
N. effect sizes	60	60	60	60	60	60
N. experiments	60	60	60	60	60	60
N. subjects	12,574	12,574	12,574	12,574	12,574	12,574

Note: Dependent variable is the effect size associated to an experiment.

Appendix B: Experimental Procedures – Further Details

In this appendix we report details of the procedures used to run the experiment described in Section 3 of the paper.

At the beginning of each session, we randomly allocated subjects to either the Conflict treatment or the NoConflict treatment. Whether a subject was allocated to the Conflict or NoConflict treatment depended on the computer terminal he or she was randomly assigned to at the beginning of the session. All instructions and procedures during the session were kept identical across treatments. The only difference between treatments was in the payoffs of the 16 dictator games, which were shown to subjects privately on their computer screens.

At the start of the session, the experimenter distributed and read aloud preliminary instructions explaining the general two-part structure of the experiment (see Appendix C for a copy of the instructions). Participants were then provided with detailed instructions for part one, which were again read aloud. These instructions included a series of control questions aimed at testing subjects' understanding of the task. Part one was started once all subjects had answered all questions correctly.

At the beginning of part one, subjects were asked to type their first name on their computer screen. This was then shown to the person they were paired with throughout part one. Subjects were then randomly assigned either the role of dictator or recipient. Dictators made their choices in the 16 games of part one. Recipients did not have any choices to make but were informed about the dictator's decision in each of the games, and had to confirm that they had seen each decision before a new game was presented to the dictator.

Once everyone had completed part one, instructions for part two appeared on subjects' computer screens. These instructions explained the Stroop task and illustrated how to submit answers on the computer. After performing the Stroop task, which lasted for five minutes, subjects completed a post-experimental questionnaire collecting standard socio-demographic measurements (such as gender, age, nationality) and the level of perceived motivational conflict in making the dictator game choices.

After completing the questionnaire subjects were shown their payoffs from part one and two of the experiment, and were randomly paid according to one of the two parts.

Appendix C: Experimental Instructions and Informed Consent

PRELIMINARY INSTRUCTIONS

Welcome! You are about to take part in a decision-making experiment. This experiment is run by the "Centre for Decision Research and Experimental Economics" and has been financed by various research foundations.

There are other people in this room, who are also participating in the experiment. Everyone is participating for the first time, and all participants are reading the same instructions. It is important that you **do not communicate with any of the other participants** during the experiment. If you have a question at any time, raise your hand and an experimenter will come to your desk to answer it.

The experiment consists of two parts: **Part 1 and Part 2**.

In each part of the experiment you will be asked to make decisions, and will have a chance to earn money. Decisions that will be made in one part of the experiment will not affect decisions or earnings in the other part of the experiment.

You will be informed of your earnings from Part 1 and Part 2 of the experiment once everyone in the room has completed Part 2. Therefore everyone will make their decisions in Part 2 without knowing their earnings from Part 1.

Only **one part of the experiment will be taken into account in determining your final earnings** from today's experiment. At the end, we will toss a fair coin. If the coin lands heads, all participants in today's experiment will be paid according to their earnings from Part 1. If the coin lands tails, all participants will be paid according to their earnings from Part 2. Your earnings will be paid out to you in private and in cash.

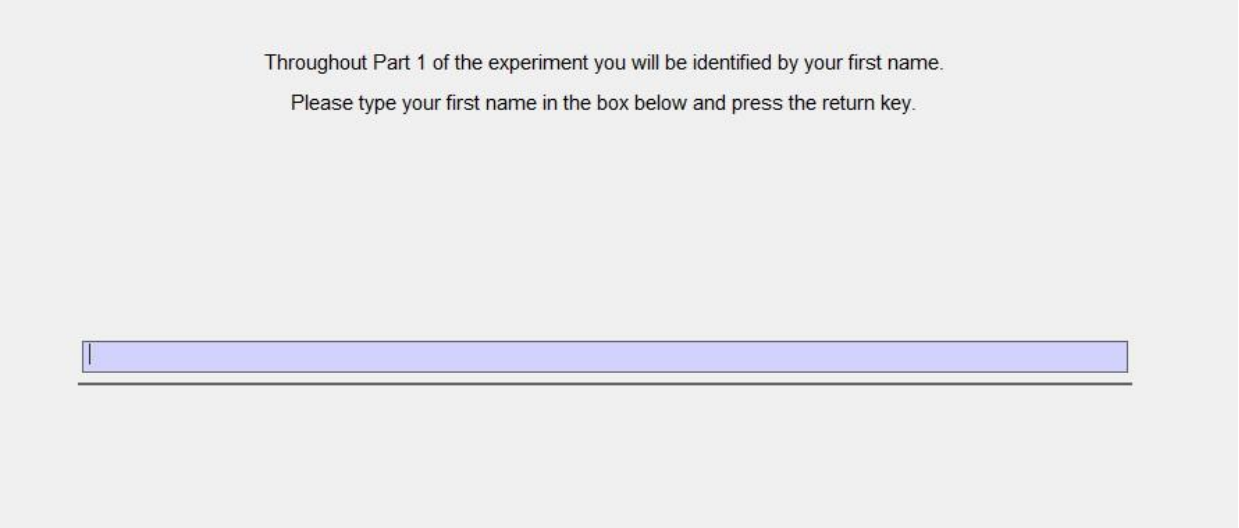
Shortly, you will receive detailed instructions about Part 1 of the experiment. You will receive detailed instructions about Part 2 directly on your computer screen once you have completed Part 1.

If you have a question now, please raise your hand and an experimenter will come to your desk to answer it.

PART 1 INSTRUCTIONS

At the beginning of PART 1 you will be **randomly paired with another person** in this room. You will remain paired with this person for the whole duration of PART 1. At the end of PART 1 the pair will be dissolved, and you will not be matched with this person again during today's experiment.

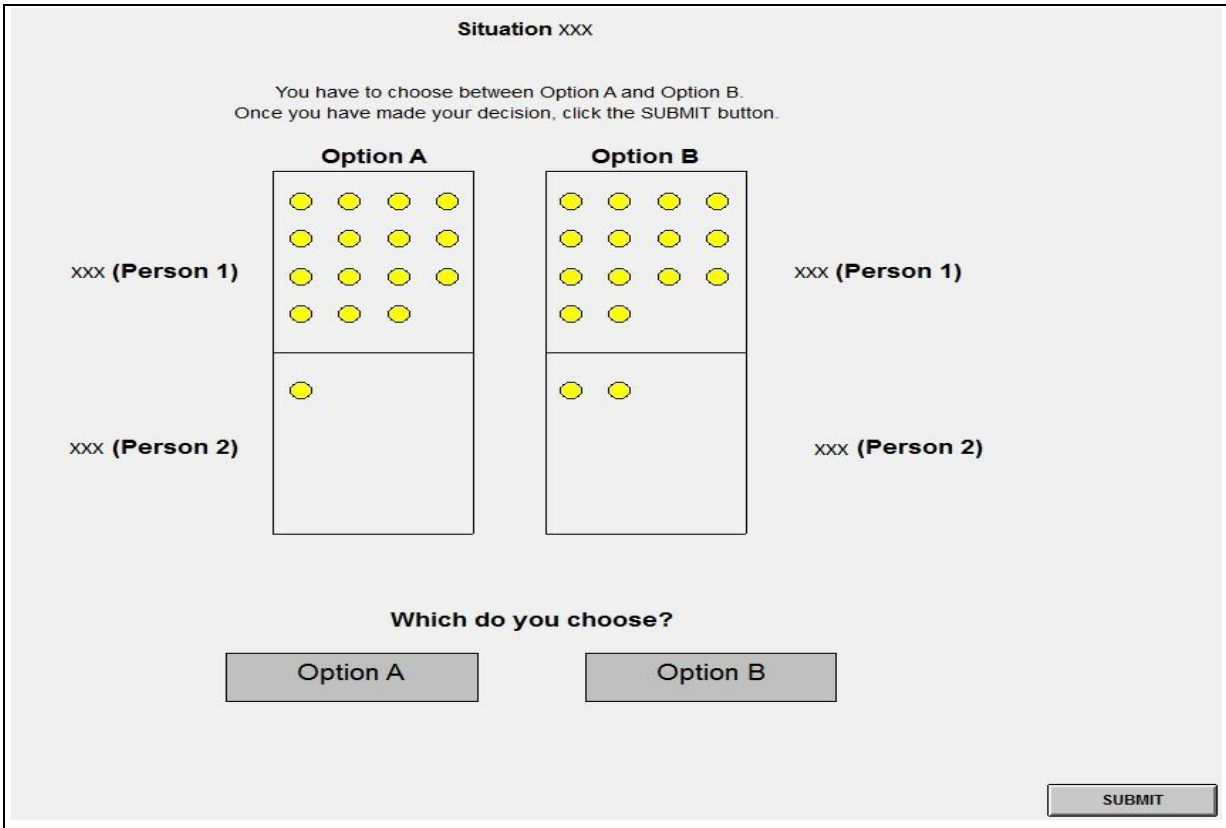
Throughout PART 1 you and the person you are paired with will be **identified by your first names**. At the beginning of PART 1 you will be asked to type your first name in a screen like the one below. Please enter your first name exactly as shown on your ID card and then press the return/enter key on your keyboard. On the following screen you will be shown the name of the person you are paired with in PART 1 of the experiment.



Each person in the pair will then be **randomly assigned a role**, either 'Person 1' or 'Person 2'. The computer will inform you of your role, which will stay the same throughout PART 1 of the experiment.

The participant in the role of **Person 1** will then be asked to make choices in a series of **16 situations**. For each situation he/she will have to choose between two options: Option A or Option B. Each option specifies an amount of money that Person 1 will receive and an amount of money that Person 2 will receive (all amounts are in British pounds).

For example, a possible situation may look as follows:



In this example situation, if Person 1 chooses Option A, Person 1 receives £15 and Person 2 receives £1. If Person 1 chooses Option B, Person 1 receives £14 and Person 2 receives £2.

The participant in the role of **Person 2** will have no choices to make in PART 1 of the experiment. However, Person 2 will be informed in real time of the choices made by Person 1 in each of the 16 situations, and will have to confirm that he/she has seen each choice before a new situation will be presented to Person 1.

At the end of the experiment **one of the 16 situations will be selected at random** by the computer. Your final earnings in PART 1 of the experiment will be based on this randomly selected situation. Each situation has an equal chance of being selected, so please consider each situation carefully. If PART 1 is selected for payment you will be paid this amount in private and in cash.

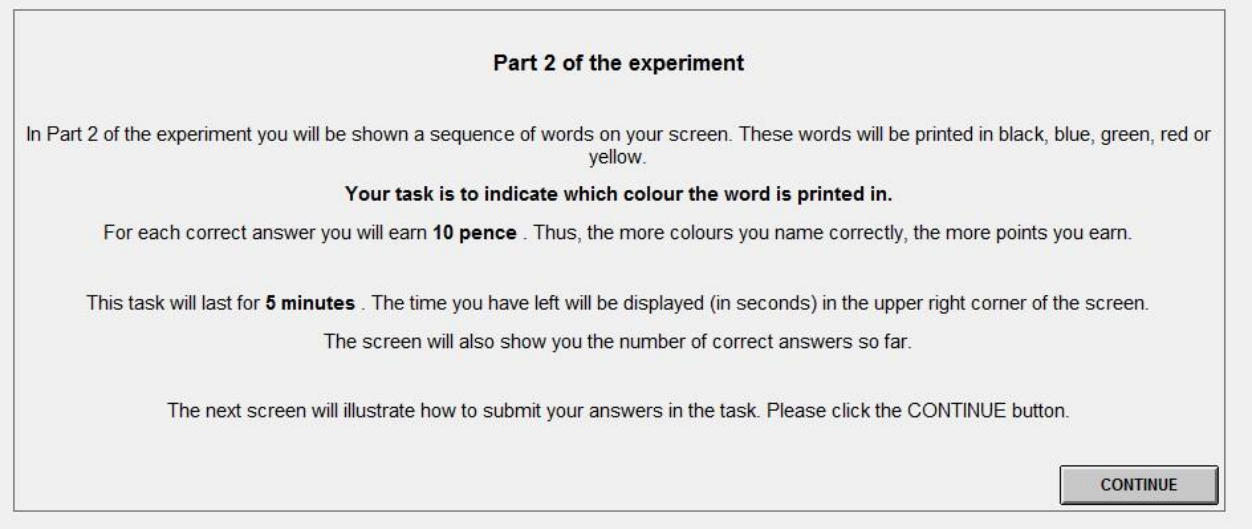
Please raise your hand if you have any questions.

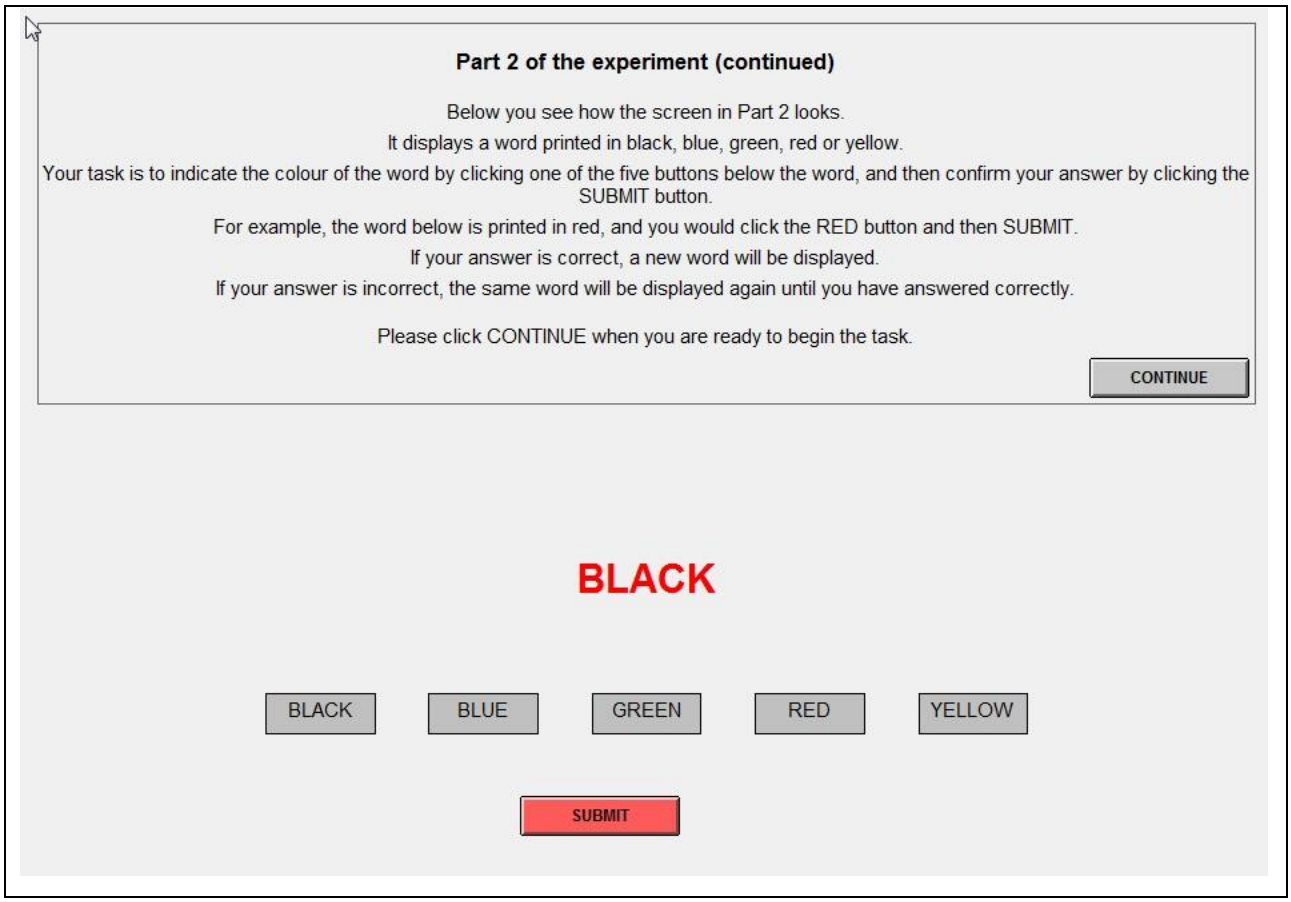
To make sure that everyone understands the instructions, please complete the questions below. In a couple of minutes an experimenter will come to your desk to check the answers.

Questions about PART 1:

1. How many choices will Person 1 make in total in PART 1? _____
2. Is the following statement true or false: you will be paired with the same person throughout PART 1 of the experiment _____
3. Suppose you are randomly assigned the role of Person 1. What will be the role of the other person in your pair? _____
4. Is the following statement true or false: if you are Person 2, you will not learn the choices made by Person 1 during PART 1 _____
5. Is the following statement true or false: your final earnings in PART 1 will be based on one of the 16 situations, which will be randomly selected at the end of the experiment _____

PART 2 INSTRUCTIONS





Form of consent

In this experiment participants' first name will be revealed to one other participant in the same experiment. In order to take part, we need your consent to the above mentioned procedure. Your identity will not be revealed to any other party in any other way.

If you do not agree to the above, you may not participate in this experiment and we kindly ask you to leave the experiment now.

If you understand and agree to the above described procedures please sign below.

Date: _____ **Signature:** _____

Print name: _____

Appendix D: Additional Analysis

In this appendix we report additional analysis of the data from the new experiment. We start by reporting, in Table D.1, the choices of dictators in the 16 dictator games of the experiment. The table reports the fraction of dictators who chose the own payoff maximizing option in the Conflict and NoConflict treatments. In the NoConflict treatment nearly all dictators chose the option that maximized the own (and the recipient's) payoff in all of the games. In the Conflict treatment the proportion of own payoff maximizing choices is lower and varies between 18.2 and 67.7 percent across games, showing that a sizeable fraction of dictators took both the recipient's and their own interests into consideration when making their choices.[9]

TABLE D.1
Percentage of subjects choosing the own payoff maximizing option

Game	Conflict		NoConflict	
	Payoffs from A vs. B	% own payoff maximizing choices	Payoffs from A vs. B	% own payoff maximizing choices
1	(8; 8) vs. (12; 0)	60.6	(8; 0) vs. (12; 8)	99.0
2	(11; 1) vs. (8; 8)	55.6	(11; 8) vs. (8; 1)	99.0
3	(9; 1) vs. (7; 5)	46.5	(9; 5) vs. (7; 1)	98.0
4	(7; 5) vs. (8; 2)	44.4	(7; 2) vs. (8; 5)	97.0
5	(10; 6) vs. (12; 0)	60.6	(10; 0) vs. (12; 6)	100
6	(8; 8) vs. (16; 0)	37.4	(8; 0) vs. (16; 8)	97.0
7	(6; 6) vs. (8; 4)	41.4	(6; 4) vs. (8; 6)	98.0
8	(10; 2) vs. (9; 7)	27.3	(10; 7) vs. (9; 2)	96.0
9	(12; 0) vs. (6; 6)	61.6	(12; 6) vs. (6; 0)	99.0
10	(11; 5) vs. (15; 1)	67.7	(11; 1) vs. (15; 5)	98.0
11	(8; 4) vs. (10; 2)	53.5	(8; 2) vs. (10; 4)	99.0
12	(9; 3) vs. (8; 4)	52.5	(9; 4) vs. (8; 3)	100
13	(10; 6) vs. (8; 8)	40.4	(10; 8) vs. (8; 6)	97.0
14	(13; 3) vs. (10; 6)	40.4	(13; 6) vs. (10; 3)	99.0
15	(6; 6) vs. (8; 2)	29.3	(6; 2) vs. (8; 6)	99.0
16	(10; 0) vs. (6; 4)	18.2	(10; 4) vs. (6; 0)	99.0

Note: in each cell of the column "Payoffs from A vs. B", the first number indicates the dictator's payoff and the second number the recipient's payoff (both displayed in GBP).

We next report regression analysis to support the results presented in Section 3.3 of the main paper. Table D.2 reports OLS regressions of dictators' performance in the Stroop task.

[9] It is interesting to note how the percentage of own-payoff maximizing choices is particularly low in the last game of the Conflict treatment (18.2%). This may be due to an order-effect (recall that the games were played in the order shown in the table). For instance, dictators may have been more prone to make an altruistic choice knowing that this was their final decision (perhaps as a way to improve their self-image).

TABLE D.2
Performance in the Stroop task

Sample	**All**	**Men**	**Women**
Model	**(1)**	**(2)**	**(3)**
Conflict	-3.679	-3.354	-4.074
	(2.339)	(4.364)	(2.981)
Male	0.407		
	(2.535)		
Age	-1.290***	-1.630***	-1.175
	(0.427)	(0.564)	(0.564)
Economics	1.533	-1.607	6.532
	(3.528)	(5.056)	(5.362)
Native	0.689	7.247	-3.388
	(3.350)	(5.503)	(4.123)
Time elapsed (in seconds)	-0.093	-0.003	-0.177
	(0.096)	(0.148)	(0.096)
Constant	249.161***	181.594	313.413***
	(72.362)	(110.926)	(72.171)
Session dummies	Yes	Yes	Yes
R^2	0.14	0.26	0.19
N	198	78	120

Note: OLS regressions, robust standard errors in parentheses. In all models the dependent variable is the number of correct answers in the Stroop task. All models include 13 session dummies, not reported in the table. *** 1% significance level.

In Model (1) we use the whole sample and regress dictators' performance on a treatment dummy ("*Conflict*") taking value 1 for observations from the Conflict treatment and 0 for observations from the NoConflict treatment. We also control for observable characteristics of the subjects by including the following variables: *Male* is a dummy variable taking value 1 for male subjects and 0 for female subjects; *Age* is the subjects' age; *Economics* is a dummy assuming value 1 if the subject is an economics student and 0 otherwise (the experiment was run in the School of Economics); *Native* is a dummy variable taking value 1 if a subject is a native English speaker and 0 otherwise[10]; *Time elapsed* measures the amount of time (in seconds) subjects spent in the lab between the beginning of part one and the beginning of the Stroop task to control for the impact on performance of any time-related factors such as boredom or opportunity to rest. The model also includes session dummies to control for session-specific effects, although these

[10] We identify this by checking subjects' nationality and considering as "native" only those who are citizens of a country with English as its official language. We include this control to assess whether language proficiency may have affected performance in the task, which contains an element of language specificity since it uses incongruence between word meaning and font color as its manipulation.

are not reported in Table D.2. In light of the debate in the literature on the role of gender as a mediator of the effect of intuition, Models (2) and (3) replicate the regression of Model (1), but for the subsample of male and female dictators, respectively.

In all models, the coefficient of the Conflict dummy is not significantly different from zero at any conventional level, indicating that performance in the Conflict treatment is indistinguishable from performance in the NoConflict treatment. This is also true when we focus on the subsample of male participants or the subsample of only female participants. Among the control variables, age has a negative and significant effect on performance, with older participants performing worse than younger ones.[11] None of the other variables significantly affect performance.

Next, we report an analysis of the effect of treatment on Stroop task performance distinguishing between subjects who reported to have experienced a more or less strong motivational conflict in the dictator games of part one. To distinguish between these two subgroups, we use subjects' responses in the questionnaire to the two questions about the extent to which they found the dictator game choices "hard" and "uncomfortable".[12] We take the average of these two questions as our measure of the strength of the motivational conflict experienced by the subject, and perform a median split of the sample in the Conflict treatment. Subjects with an average response below the median are classified as "low conflict" (49 subjects), while those with a response above the median are classified as "high conflict" (50 subjects).[13]

Figure D.1 shows the Stroop task performance of subjects in each subgroup, as well as those in NoConflict. High-conflict subjects gave on average 147.78 correct answers in the Stroop task (s.d. 16.69), while low-conflict subjects gave 144.61 correct answers (s.d. 18.65). Using Wilcoxon-Mann-Whitney tests, we do not find any significant difference in Stroop task performance between these two subgroups ($p = 0.763$), or between each subgroup and the subjects in the NoConflict treatment ($p > 0.172$).

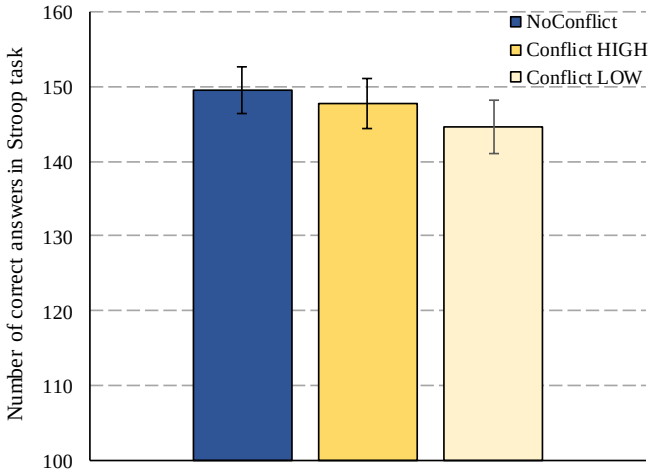
---

[11] The mean and median age in our sample are 21 and 20, respectively. The variable, which is self-reported, ranges from 18 to 42 years.

[12] Recall that these items were: "Overall, how hard was it to choose between option A and option B in the 16 situations of Part 1?", and "Overall, to what extent did you experience discomfort in making your choices in the 16 situations of Part 1?". In both cases, responses were collected on a scale from 0 ("not at all") to 10 ("very much").

[13] One caveat is in order: since the classification is based on self-reported measures of perceived conflict, it may be prone to bias. For instance, dictators who make many selfish choices may strategically report they found their choices hard and uncomfortable in order to appear less selfish in the eyes of the experimenter.

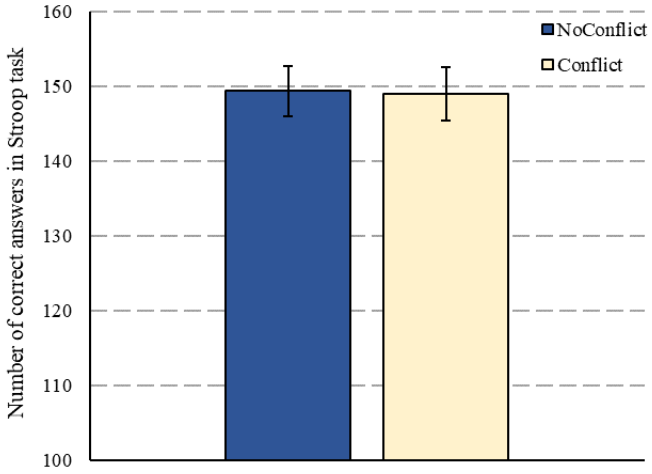
FIGURE D.1
Performance in the Stroop task, by treatment and strength of conflict



Note: Bars indicate 95% confidence intervals computed as $\mu \pm 1.96 * SE^{\mu}$ where μ is the estimated mean, SE^{μ} is its standard error, and 1.96 is the z-score for the 97.5 percentile point of the normal distribution.

Finally, we report an analysis of the effect of treatment on Stroop task performance for the recipients. Figure D.2 reports average performance of the recipients in the two treatments. In each treatment, recipients' performance is similar to that of dictators (Conflict, recipients: mean 149.06, s.d. 18.24; dictators: mean 146.21, s.d. 17.67; Wilcoxon-Mann-Whitney test p = 0.099; NoConflict, recipients: mean 149.38, s.d. 16.81; dictators: mean 149.44, s.d. 15.82; Wilcoxon-Mann-Whitney test p = 0.904). Moreover, recipients' performance is very similar and not statistically different between the two treatments according to a Wilcoxon-Mann-Whitney test (p = 0.842). OLS regressions, reported in Table D.3, confirm the result.

FIGURE D.2
Performance in the Stroop task, recipients



Note: Bars indicate 95% confidence intervals computed as $\mu \pm 1.96 * SE^{\mu}$ where μ is the estimated mean, SE^{μ} is its standard error, and 1.96 is the z-score for the 97.5 percentile point of the normal distribution.

TABLE D.3
Performance in the Stroop task, recipients

Sample	**All**	**Men**	**Women**
Model	**(1)**	**(2)**	**(3)**
Conflict	-1.273	-2.224	-0.564
	(2.546)	(4.517)	(3.019)
Male	3.600		
	(2.784)		
Age	-0.472	2.398	-2.296**
	(0.923)	(1.345)	(0.933)
Economics	-2.365	-2.323	-3.394
	(3.680)	(5.910)	(5.224)
Native	-2.768	-4.746	-3.397
	(3.118)	(4.619)	(4.250)
Time elapsed (in seconds)	-0.316***	-0.418***	-0.274**
	(0.092)	(0.127)	(0.137)
Constant	389.294***	414.509***	394.751***
	(70.239)	(97.982)	(106.029)
Session dummies	Yes	Yes	Yes
R^2	0.14	0.24	0.26
N	198	88	110

Note: OLS regressions, robust standard errors in parentheses. In all models the dependent variable is the number of correct answers in the Stroop task. All models include 13 session dummies, not reported in the table. *** 1% significance level; ** 5% significance level.