Supplementary Materials for Norm Shifts Under the Strategy Method

Simon Columbus* Robert Böhm†

Statistical Details

Simple effects at different levels of belief (Study 1)

We compared cooperation under the strategy method and under the simultaneous method at each level of belief using paired t-tests and Wilcoxon signed rank tests. Tests were significant for beliefs $B_{m,i} = \{0, 2, 3, 4, 5\}$ (as well as $B_{m,i} = 8$ using the non-parametric test; Table S1). These comparisons are robust for adjustment for multiple comparisons using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995).

Norms at each level of belief (Study 2)

We elicited perceptions of normativity for all behaviours $T_{sim.,i} = [0, 10]$ under the simultaneous method. Table S2 shows that equitable transfers $T_{sim.,i} = 5$ were perceived as most normative by participants with relatively low expectations of cooperation, i.e., $B_{sim.,i} = [2, 5]$. In addition, fully cooperative transfers $T_{sim.,i} = 10$ were perceived as most normative by participants with very positive expectations, $B_{sim.,i} \geq 8$. However, contrary to our expectations, such transfers did not form a second mode across all beliefs. Under the strategy method, we only elicited beliefs for behaviours $T_{strat.,i} = [0, 5]$. Table S3 shows that equitable transfers $T_{strat.,i,j} = B_{strat.,i}$ were perceived as most normative at all levels of belief.

*Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 2300 København, Denmark. Email: simon@simoncolumbus.com. ORCID: 0000-0003-1546-955X

†Department of Psychology, Department of Economics, and Copenhagen Center for Social Data Science (SODAS), University of Copenhagen. ORCID: 0000-0001-6806-0374

TABLE S1: Mean differences in cooperation between the simultaneous and the strategy method at east level of belief. p-values are not adjusted for multiple comparisons.

$B_{m,i}$	$T_{sim.,i} - T_{strat.,i}$	p (t-test)	p (Wilcoxon)
0	1.909	.020	.027
1	1.700	.321	.552
2	2.079	< .001	< .001
3	2.429	.006	.015
4	1.795	< .001	< .001
5	2.071	< .001	< .001
6	.022	.963	.838
7	.444	.665	.670
8	-.506	.154	.023
9	-.667	.545	.850
10	.060	.862	.865

TABLE S2: Mean ratings of normativity of behaviours $T_{sim.,i}$ at different levels of belief $B_{sim.,i}$ under the simultaneous method. The mode for each level of belief is highlighted in bold. Equitable transfers $T_{sim.,i} = 5$ are rated most normative by participants with beliefs $B_{sim.,i} = [2, 5]$.

$B_{sim.}$	n	Behaviour $T_{sim.,i}$										
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	9	.04(.82)	**.56(.58)**	**.56(.58)**	.41(.52)	.33(.58)	.41(.78)	.26(.78)	.33(.82)	.26(.91)	.19(.99)	.33(.88)
1	6	−.44(.89)	−.44(.66)	−.11(.34)	−.22(.27)	.22(.27)	.22(.27)	.00(.56)	**.33(.84)**	**.33(1.03)**	.22(.98)	.11(.91)
2	38	−.44(.74)	−.19(.64)	.00(.62)	.19(.54)	.30(.49)	**.54(.58)**	.26(.58)	.16(.61)	.04(.71)	−.11(.81)	−.09(.84)
3	14	−.33(.78)	.00(.57)	.24(.44)	.52(.41)	.57(.42)	**.62(.43)**	.10(.33)	−.19(.39)	−.24(.51)	−.38(.71)	−.38(.80)
4	47	−.79(.40)	−.45(.54)	−.11(.51)	.11(.54)	.33(.50)	**.67(.41)**	.56(.51)	.42(.57)	.25(.60)	.15(.66)	.16(.69)
5	34	−.78(.43)	−.61(.44)	−.37(.43)	−.08(.49)	.25(.56)	**.59(.43)**	.47(.46)	.41(.46)	.35(.67)	.22(.74)	.22(.83)
6	28	−.88(.32)	−.62(.34)	−.43(.35)	−.19(.42)	.02(.50)	.50(.47)	.50(.43)	**.60(.38)**	**.60(.46)**	.55(.55)	.45(.70)
7	4	−.17(1.00)	.00(.86)	−.17(.64)	.00(.38)	−.17(.33)	.17(.64)	.00(.38)	.17(.33)	.00(.67)	.00(.67)	**.33(.94)**
8	30	−.64(.57)	−.56(.53)	−.36(.57)	−.18(.49)	.04(.45)	.38(.49)	.47(.41)	.49(.49)	.60(.54)	.62(.57)	**.73(.60)**
9	1											
10	5	−1.00(.00)	−.87(.30)	−.73(.37)	−.60(.37)	−.33(.00)	.20(.30)	.33(.00)	.47(.30)	**.73(.37)**	**.73(.60)**	**.73(.60)**
All	222	−.62(.62)	−.39(.6)	−.18(.56)	.02(.54)	.21(.52)	**.51(.51)**	.39(.5)	.35(.56)	.3(.67)	.21(.75)	.23(.8)

TABLE S3: Mean ratings of normativity of behaviours $T_{strat.,i,j}$ at different levels of belief $B_{strat.,i}$ under the strategy method. The mode for each level of belief is highlighted in bold. Reciprocal transfers $T_{strat.,i,j} = B_{strat.,i}$ are rated most normative at any level of belief.

$B_{strat.,i}$	**Behaviour** $T_{strat.,i,j}$					
	0	**1**	**2**	**3**	**4**	**5**
1	**.70(.66)**	.08(.70)	−.06(.64)	−.24(.66)	−.41(.68)	−.46(.73)
2	−.24(.68)	**.80(.41)**	.23(.56)	−.12(.59)	−.38(.62)	−.53(.64)
3	−.57(.60)	−.08(.58)	**.83(.39)**	.24(.57)	−.17(.62)	−.38(.68)
4	−.69(.55)	−.48(.54)	−.06(.57)	**.83(.39)**	.23(.56)	−.11(.65)
5	−.75(.53)	−.65(.52)	−.40(.55)	.04(.56)	**.83(.42)**	.32(.58)
6	−.80(.54)	−.72(.50)	−.57(.53)	−.30(.56)	.16(.56)	**.89(.36)**

Additional Robustness Checks

Controls for experimental manipulations (Study 1)

Study 1 re-analysed previously published data from an experiment which employed two orthogonal frames (the NAME of the game, 'community game' vs. 'stock exchange game' and an explicit statement of CONFLICT of interests in the game, low conflict vs. no statement vs. high conflict). Both frames were randomised and balanced. Here we show that the key findings from Study 1 are not materially affect by these manipulations. In addition, we provide results with further demographic controls for age, gender, educational attainment, and naivety to the game (either prior experience with or prior knowledge of Prisoner's Dilemma-like games).

First, we found that overall, players were more cooperative under the simultaneous method than under the strategy method (at the stated level of belief). As a baseline model, we fitted linear mixed model with decisions under either method as the dependent variable, the method as an independent variable, and a random factor for participants. In subsequent models, we included the three-way interaction of method, NAME, and CONFLICT, as well as controls for age, gender, education, and naivety. Table S4 shows that the effect of method was robust across all models.

Second, we also examined the correlation between transfers under the simultaneous method and under the strategy method (at the stated level of belief). As a baseline model, we regressed transfers under the simultaneous method on transfers under the strategy method. In subsequent models, we included the three-way interaction of method, NAME, and CONFLICT, as well as controls for age, gender, education, and naivety. Table S5 shows that the relationship between transfers under both methods was robust across all models, although it was qualified by a method × NAME × CONFLICT interaction. Plotting the pattern of transfer under both methods faceted by experimental conditions shows, however,

that the cubic relationship held under all conditions (Figure S1).

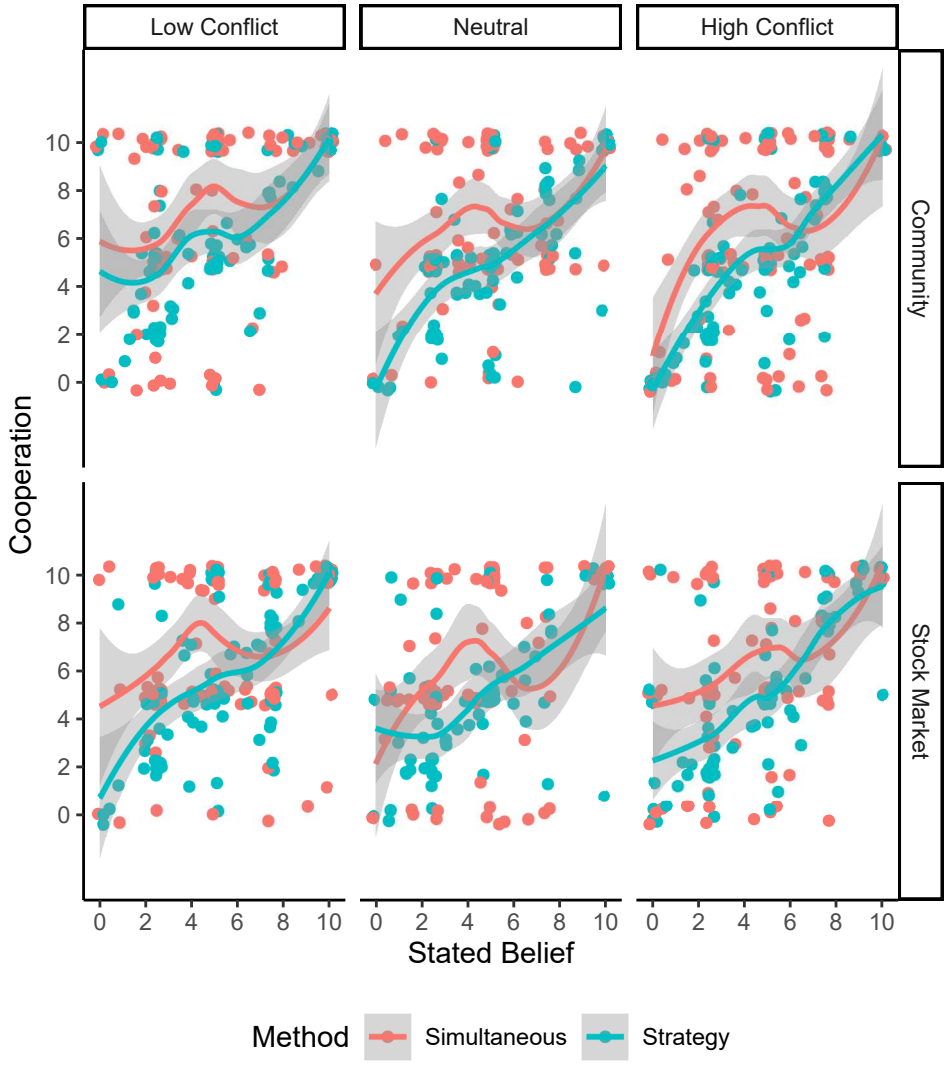

FIGURE S1: Relationships between stated beliefs and cooperation of conditional cooperators under the simultaneous and the strategy method, by experimental condition. Loess regression shows a cubic relationship under the simultaneous protocol, and a largely linear relationship under the strategy method. $N_{Comm.,LC} = 226$, $N_{Comm.,N} = 204$, $N_{Comm.,HC} = 238$, $N_{Stock.,LC} = 216$, $N_{Stock.,N} = 166$, $N_{Stock.,HC} = 182$.

Robustness to the Location of Elicited Beliefs (Study 1)

Though it is conventional to use the midpoint of the interval elicited using the MLI method as the location of the belief, the underlying theory does not specify a point estimate (Schlag & van der Weele, 2015). If the underlying belief is symmetrically distributed, then the midpoint aligned with the median/mean of the interval. If the distribution is skewed,

TABLE S4: Mean difference between transfers under the simultaneous method and the strategy method. Model 1 shows the baseline model. Model 2 adds controls for NAME, CONFLICT, as well as the three-way interaction between Method, NAME, and CONFLICT, and all lower-order interactions. Model 3 additionally adds demographic controls. Results show a type III analysis of variance with Satterthwaite's method.

	Model 1				**Model 2**				**Model 3**			
	F	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***	***F***	***df1***	***df2***	***p***
Method	123.62	1	1027	< .001	122.20	1	1021	< .001	121.59	1	1018	< .001
Name					1.02	1	1071	.314	1.28	1	1055	.258
Conflict					5.43	2	1072	.004	4.93	2	1056	.007
Age									3.85	1	1033	.050
Gender									3.17	1	1053	.076
Education									2.50	10	1054	.006
Naive									.02	1	1055	.887
Method×Name					.06	1	1021	.808	.06	1	1018	.807
Method×Conflict					.35	2	1021	.702	.37	2	1018	.694
Name×Conflict					.30	2	1072	.743	.31	2	1055	.734
Method×Name×Conflict					.16	2	1021	.854	.16	2	1018	.848

however, it could be argued that a different point on the interval provides a better location of the belief (though importantly the interval will always contain the median). We address this concern in two ways. First, we show that there is a significant proportion of individuals for whose transfers under the simultaneous method fall outside the range of transfer made under the strategy for any of the beliefs inside the elicited interval. Second, we show that our key results are robust for many locations of the belief below and above the midpoint of the elicited interval.

First, we examined whether there are individuals whose transfers under the simultaneous method deviate from their transfers under the strategy method even if we consider the location of their belief to encompass the entire interval. To do so, we first computed the range of transfers under the strategy method at any level of belief contained in the interval $[L, U]$. Second, we tabulated whether the transfer under the simultaneous method $T_{sim,i}$ fell within this range. For 25.6% of participants, their transfer under the simultaneous method fell outside the range of transfers under the strategy method at any level of belief within the elicited interval. This proportion was weakly, but negatively correlated with the level of belief as indicated by the midpoint of the elicited interval, $r = .09$, $t(983) = 2.97$, $p = .003$, indicating that the proportion of individuals whose transfers fell outside the expected interval was higher among those with lower beliefs, in line with our finding that the association between beliefs and decisions under both methods diverged specifically

TABLE S5: Linear relationship between transfers under the simultaneous method and under the strategy method. Model 1 shows the baseline model; Transfer indicates the strategy method transfer. Model 2 adds controls for NAME, CONFLICT, as well as the three-way interaction between Transfer, NAME, and CONFLICT, and all lower-order interactions. Model 3 additionally adds demographic controls. Results show the analysis of variance table.

	Model 1			**Model 2**			**Model 3**		
	F	***df***	***p***	***F***	***df***	***p***	***F***	***df***	***p***
Transfer	277.61	1	< .001	280.56	1	< .001	286.64	1	< .001
Name				.47	1	.491	.47	1	.495
Conflict				.35	2	.708	.37	2	.694
Age							4.07	1	.044
Gender							.56	1	.452
Education							3.10	10	.001
Naive							.58	1	.445
Transfer×Name				3.02	1	.082	2.83	1	.093
Transfer×Conflict				1.99	2	.137	2.67	2	.070
Name×Conflict				.25	2	.781	.19	2	.830
Transfer×Name×Conflict				5.89	2	.003	5.18	2	.006
Residuals		984			974			959	

among individuals with stated beliefs $B_{sim,i} < 5$ (see Figure S2).

Second, we tested to what degree our key results were robust to choosing different locations of the belief on the elicited intervals. To do so, we re-ran the model regressing transfers on the interactions between method and belief, the square of belief, and the cube of belief, setting the location of the belief at eleven points between the minimum and the maximum of the elicited interval (i.e., increasing the location from the minimum to the maximum by deciles). The significant interaction between the method and the linear effect of belief was robust between the third and the seventh decile (Table S6); the significant interaction between the method and the cubic effect of belief between the third and the sixth decile.

We also re-ran the follow-up models separately for each method (Table S7). Under the strategy method, we expected a linear effect of belief, which held for all but the second and ninth decile, and no cubic effect, which held for the fifth, sixth, and tenth decile. Under the simultaneous method, we expected a cubic effect of belief, which held for all but the tenth decile, and no linear effect, which held for all but the first and third decile. In sum, our key results were thus robust under a range of choices for the location of the belief on the elicited interval. Importantly, we found different results if we set the location of the belief towards the edges of the elicited intervals. This is in line with our theoretical rationale:

if the true location of the belief is located towards the midpoint of the elicited interval, locations further towards the edges of the interval should provide biased estimates of belief. In turn, for the true location of beliefs to be located at these edges, we would have to assume rather implausible distributions.

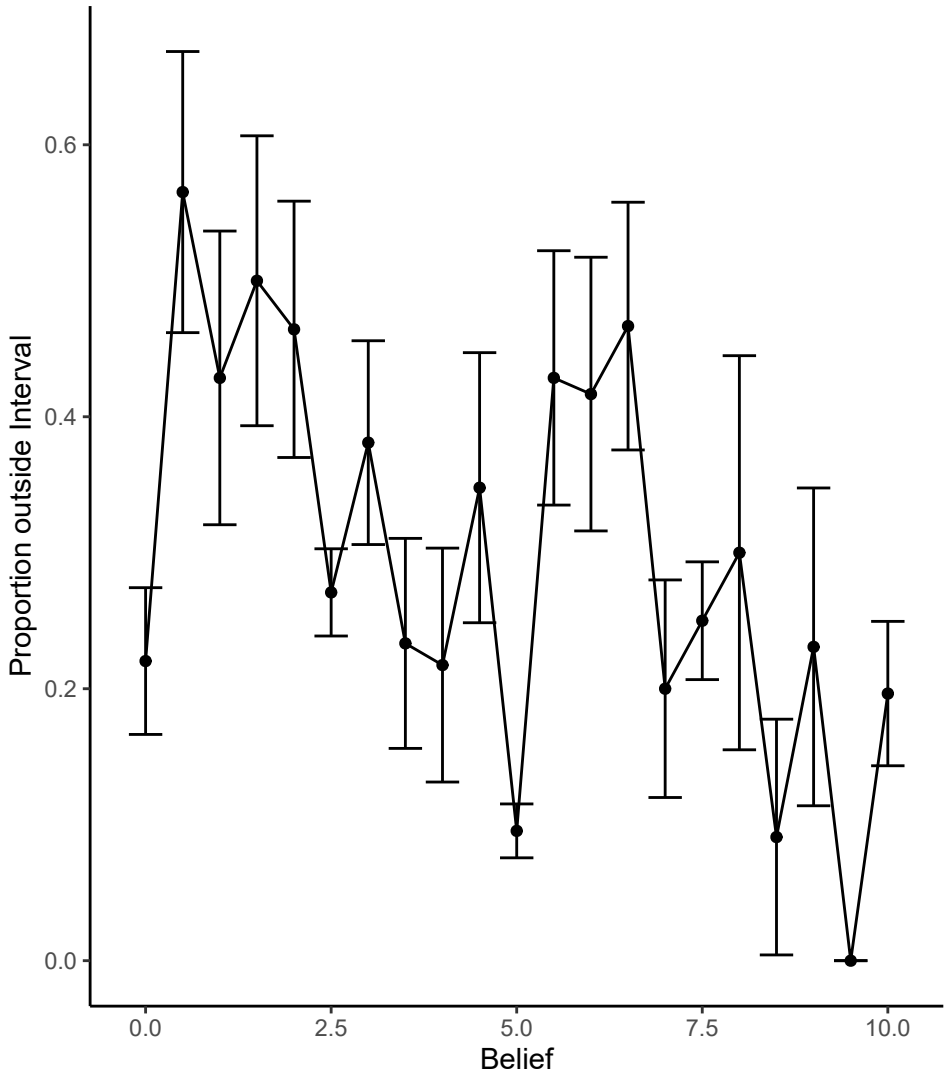

FIGURE S2: Proportion of individuals whose transfer under the simultaneous method fell outside the range of transfers under the strategy method at any level of belief within the elicited interval, by level of belief as indicated by the midpoint of the elicited interval.

Main effects and lower-level interactions (Study 2)

In the main text, we report the effects of regressions of normativity ratings on fairness and reciprocity from models including two-way interactions with method and three-way interactions with method and belief. Because higher-order interactions may obscure lower-

TABLE S6: Estimates for the interaction between method and linear belief (left) as well as cubic belief (right) on transfers, varying the location of belief in deciles from the minimum to the maximum of the elicited interval.

	Linear			Cubic		
Decile	***B***	***SE***	***p***	***B***	***SE***	***p***
1	-.16	.22	.478	.01	.01	.099
2	.36	.18	.042	-.00	.01	.554
3	.65	.14	< .001	-.02	.01	.007
4	.73	.12	< .001	-.02	.01	< .001
5	.68	.12	< .001	-.02	.01	< .001
6	.52	.11	< .001	-.01	.01	.016
7	.32	.11	.004	-.00	.01	.405
8	.10	.12	.408	.00	.01	.441
9	-.10	.15	.504	.01	.01	.117
10	.18	.19	.350	-.00	.01	.656

TABLE S7: Estimates for the linear and cubic effects of belief on transfers under the strategy method and the simultaneous method, varying the location of belief in deciles from the minimum to the maximum of the elicited interval.

	Strategy Method						Simultaneous Method					
	Linear			Cubic			Linear			Cubic		
Decile	***B***	***SE***	***p***	***B***	***SE***	***p***	***B***	***SE***	***p***	***B***	***SE***	***p***
1	.93	.17	< .001	-.02	.01	.021	1.08	.21	< .001	-.03	.01	< .001
2	.21	.13	.107	.02	.01	.004	-.15	.17	.392	.02	.01	.005
3	.34	.10	.001	.02	.00	.001	-.31	.13	.016	.04	.01	< .001
4	.52	.09	< .001	.01	.00	.016	-.22	.12	.059	.04	.01	< .001
5	.61	.08	< .001	.01	.00	.098	-.07	.11	.495	.03	.01	< .001
6	.60	.08	< .001	.01	.00	.077	.08	.10	.445	.02	.01	< .001
7	.50	.08	< .001	.01	.00	.004	.18	.10	.073	.02	.01	.002
8	.28	.08	.001	.02	.00	< .001	.18	.11	.092	.02	.01	.007
9	.00	.11	.969	.03	.01	< .001	.11	.14	.452	.02	.01	.018
10	.47	.14	.001	.00	.01	.751	.29	.18	.104	.01	.01	.459

order effects, we ran additional models including only main effects and only two-way interactions S8. As in the main text, we did not centre the belief variable (i.e., a value of 0 corresponded to a reported belief of 0). Reciprocity and fairness were dummies that were 1 when the norm was present and 0 otherwise. Method was a dummy that was 0 for the

simultaneous method and 1 for the strategy method.

The hypothesised two-way interactions between method and fairness and method and reciprocity were robust in the model only including two-way interactions (Model 3). These interactions accounted for the main effects that are present when no interactions are included in the model (Models 1 and 2).

TABLE S8: Predictors of elicited normativity ratings for different rates of cooperation and at varying levels of beliefs. Model 1 shows only main effects of fairness and reciprocity; Model 2 includes controls for beliefs and method, and Model 3 shows main effects and second-order interactions.

	Model 1			**Model 2**			**Model 3**		
	B	***SE***	***p***	***B***	***SE***	***p***	***B***	***SE***	***p***
Intercept	2.16	.04	< .001	2.60	.08	< .001	2.91	.10	< .001
Fairness	.61	.06	< .001	.61	.06	< .001	.11	.15	.479
Reciprocity	1.05	.06	< .001	1.05	.06	< .001	-.17	.15	.258
Method				-.17	.04	< .001	-.30	.10	.005
Belief				-.11	.02	< .001	-.18	.03	< .001
Method × Belief							.01	.03	.657
Method × Fairness							-.77	.11	< .001
Belief × Fairness							.25	.04	< .001
Method × Reciprocity							1.30	.11	< .001
Belief × Reciprocity							.14	.04	.001

References

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Schlag, K. H. & van der Weele, J. J. (2015). A method to elicit beliefs as most likely intervals. *Judgment and Decision Making*, *10*(5), 456–468.