Judgment and Decision Making, Vol. 13, No. 1, January 2018, pp. 137–149

Making good cider out of bad apples — Signaling expectations boosts cooperation among would-be free riders

Michiru Nagatsu* Karen Larsen† Mia Karabegovic‡ Marcell Székely§ Dan Mønster¶ John Michael‖

Abstract

The present study investigates how group-cooperation heuristics boost voluntary contributions in a repeated public goods game. We manipulate two separate factors in a two-person public goods game: i) group composition (Selfish Subjects vs. Conditional Cooperators) and ii) common knowledge about group composition (Information vs. No Information). In addition, we let the subjects signal expectations of the other's contributions in the experiment's second phase. Common knowledge of Selfish type alone slightly dampens contributions but dramatically increases contributions when signaling of expectations is allowed. The results suggest that group-cooperation heuristics are triggered when two factors are jointly salient to the agent: (i) that there is no one to free-ride on; and (ii) that the other wants to cooperate because of (i). We highlight the potential effectiveness of group-cooperation heuristics and propose *solution thinking* as the schema of reasoning underlying the heuristics. The high correlation between expectations and actual contributions is compatible with the existence of default preference to satisfy others' expectations (or to avoid disappointing them), but the stark end-game effect suggests that group-cooperation heuristics, at least among selfish players, function ultimately to benefit material self-interest rather than to just please others.

Keywords: group-cooperation heuristics, public goods, group composition, expectations, solution thinking

1 Introduction

In the literature on the voluntary provision of public goods in the public goods game (PG hereafter), experimental and econometric innovations have led to an increasing appreciation of the heterogeneity of subjects' attitudes and approaches toward cooperation (Fischbacher et al. 2001; Burland & Guala 2005; Bardsley and Moffatt 2007). As a result, it has become standard to distinguish among various "types" of subjects, such as "free-riders" (or "selfish players"), who behave uncooperatively regardless of what others do and "conditional cooperators" who cooperate as long as they believe that others will do the same. These developments have stimulated various studies focusing on the effect of group composition (i.e., consisting of similar or different types of subjects) on voluntary contributions in PG, giving rise to some interesting regularities. For example, Gächter & Thöni (2005) found that those with the highest cooperative tendency (identified as such in a prior ranking experiment) cooperated more when they were sorted with "like-minded" subjects and knew this, than in an unsorted treatment. This may not be surprising for conditional cooperators, given that they have more reason to expect others to cooperate (and therefore more reason to cooperate) in the sorted treatment than in the unsorted treatment, when the sorting mechanism is common knowledge. But it does underscore the importance of group composition, and also of expectations arising from group composition, in modulating voluntary contributions to public goods in PG (see also Fischbacher & Gächter 2010; de Oliveira et al. 2014).

More surprisingly, Gächter and Thöni (2005) also found that subjects with the lowest cooperative tendency also cooperated more in the sorted treatment than in the randomly

The study was made possible by a seed funding grant (nr. 26166) from the Interacting Minds Center at Aarhus University. John Michael and Marcell Székely were supported by a Starting Grant from the European Research Council (n 679092, SENSE OF COMMITMENT). Michiru Nagatsu was supported by the Academy of Finland Center of Excellence for the Philosophy of the Social Sciences (TINT) and Aalto Choice Tank (ACT). An earlier version of this paper was presented at the 3rd International Conference: Economic Philosophy at Aix-en-Provence, June 15–16 2016, and POS seminar, Helsinki, Dec 5 2016. We thank people who gave us useful comments in those events and elsewhere, in particular Cristina Bicchieri, Raul Hakli, Topi Miettinen, Cédric Paternotte, and Oana Stanciu. Finally, we thank the three anonymous reviewers, the action editor (Enrique Fatas) and the editor (Jon Baron) for invaluable comments.

*Department of Political and Economic Studies, University of Helsinki PL 24 (Unioninkatu 40 A) 00014, Finland E-mail: michiru.nagatsu@helsinki.fi.

†Department of Economics and Management, University of Helsinki. E-mail: karenwinnie77@hotmail.com.

‡Department of Cognitive Science, Central European University, Budapest, Hungary. E-mail: Karabegovic_Mia@phd.ceu.edu.

§Department of Cognitive Science, Central European University, Budapest, Hungary E-mail: szekelymarcell@gmail.com.

¶Department of Economics and Business Economics, Aarhus University, Interacting Minds Centre, Denmark. E-mail: danm@econ.au.dk.

‖Department of Philosophy, University of Warwick Coventry, UK. E-mail: J.Michael.2@warwick.ac.uk.

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

matched treatment. Gunnthorsdottir et al. (2007: 313) report a similar result, even when subjects did not know about the sorting rule. What is going on here? Social psychologists have long known about in-group favoritism. This is people's tendency to behave cooperatively with others who belong to the same group, even when the group affiliation in question is imposed exogenously and arbitrarily using the so-called minimal group paradigm (see e.g., Wit & Wilke 1992 for a case in PG; more generally Billig & Tajfel 1973; Brewer 1979). But what is puzzling about the case at hand is that these subjects are sorted precisely because of their uncooperative tendency and yet they nevertheless cooperate more than in a non-sorted treatment. Is this due to in-group favoritism of some sort, or to some other cause?

The present study addresses this question by investigating the effects of (a) information about group composition and (b) signaling of expectations about each other's contributions. To anticipate our results, we found a negative effect of common knowledge of each other's type on selfish subjects' contribution level, but a remarkably positive effect of common knowledge once the signaling of expectations about each other's contributions was allowed. This indicates that the higher cooperation among like-minded selfish players is not triggered by blind in-group favoritism, or some conformity bias, but rather by group-cooperation heuristics activated by beliefs that such cooperation is mutually profitable and sustainable (up to a certain point), and that the others see it this way too.

We proceed as follows: first we motivate our experimental design as a way forward towards systematically uncovering the mechanism of group-cooperation heuristics (Section 2). Next, we describe our experimental design and implementation (Section 3). We then report the results (Section 4), and conclude by discussing their implications and limitations, proposing *solution thinking* as a schematic explanation of how group-cooperation heuristics work in social dilemmas (Section 5).

2 Group-cooperation heuristics and their mechanisms

Gächter and Thöni (2005) identify two possible mechanisms, which are not necessarily mutually exclusive, to explain the puzzling finding that their least cooperative subjects cooperated more in a "like-minded" group than in a heterogeneous group. Taking this as our starting point, we develop our new experimental design to investigate the mechanisms underlying cooperation of a group of otherwise selfish subjects.

2.1 Group-cooperation heuristics

The first possible mechanism that Gächter and Thöni (2005) identify is that their "like-minded" group manipulation may have triggered a boundedly rational group-cooperation heuristics (Selten & Stoecker 1986) among these least cooperative subjects: "LOW contributors [i.e., those whose contribution level was the lowest one third in the one-shot PG called the Ranking experiment, which preceded the main experiment] have revealed to each other, that they chose the money-maximizing strategy in the Ranking experiment. They may therefore believe that there are no cooperators around to free ride on. Thus, they understand that they need to cooperate among themselves if they want to earn money." (Gächter & Thöni 2005, 310–311) In other words, the selfish subjects' common knowledge that they had been grouped together on the basis of their previous low contributions made it salient that it would not be possible to free-ride. If this is the case, then it should be possible to reduce contributions by depriving subjects of information about each others' types, since doing so would obscure the recognition that "there are no cooperators around to free ride on." However, as Jin & Yamagishi (1997) claim, an additional condition may need to be satisfied in order for such a group-cooperation heuristics to be triggered. For it is one thing to recognize (i) that mutual cooperation is preferred to mutual defection given that free-riding is not possible; it is quite another to believe (ii) that others expect mutual cooperation by recognizing (i). Without some confidence in (ii), cooperation may be a risky strategy.

One of the main contributions of our study is to systematically investigate the relation between conditions (i) and (ii) by introducing mutual signaling of expectations as an operationalization of (ii). We specify the following two scenarios under which the group-cooperation heuristics may be triggered among selfish subjects (SSs):

Scenario 1: If both (i) and (ii) are *interdependent* (viz. independently insufficient but jointly sufficient) conditions, then:

1.a we expect SSs to cooperate when both (i) and (ii) are satisfied

1.b we expect SSs to not cooperate when (i) is satisfied (with common knowledge) but (ii) is not satisfied (without signaling expectations)

1.c we expect SSs to not cooperate when (i) is not satisfied (without common knowledge) but (ii) is satisfied (with signaling expectations).

Alternatively,

Scenario 2: If (i) and (ii) are *independent* triggering conditions, the effects of common knowledge and signaling should be additive and then:

2.a we expect SSs to cooperate when both (i) and (ii) are satisfied,

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

2.b we expect SSs to cooperate (but less than in (2.a)) when (i) is satisfied (with common knowledge) but (ii) is not satisfied (without signaling expectations)

2.c we expect SSs to cooperate (but less than in (2.a)) when (i) is not satisfied (without common knowledge) but (ii) is satisfied (with signaling expectations)

The possibility that signaling of expectations alone may weakly but independently trigger group-cooperation heuristics (that is, scenario 2.c above) is motivated by the hypothesis, recently put forward by Heintz et al. (2015), that people have a default preference for fulfilling others' expectations (or for avoiding disappointing others' expectations). In support of this conjecture, Heintz and colleagues observed that a majority of dictators in a dictator game modulated their transfer to more closely match expectation that their respective receivers had indicated, when they learned of these expectations (provided they were not unreasonable). Heintz et al. also argue that this conjecture is supported by results from a study by Dana et al. (2006). In Experiment 1 of Dana and colleagues' study, the subject playing the role of dictator could pay $1 in order to exit from the game without the receiver knowing that the game had taken place. Many (about one-third) of the subjects did indeed choose this option. In Experiment 2, dictators were again offered a $1 exit option, but in this case it was clear that receivers would never know that a dictator game had taken place (i.e., any transfers would be surreptitiously added to a reward for a different task). In this setup, almost no dictator accepted the option or made any transfer. Thus, Dana and colleagues, like Heintz and colleagues, surmise that a default preference to fulfill others' expectations (or to avoid disappointing them) provides a compelling explanation of the finding that dictators transfer anything at all in typical dictator games (see Camerer, 2003; Ockenfels & Werner, 2014, for similar discussions). If Heintz and colleagues' conjecture is correct, then we should expect that signaling of expectations would increase selfish players' contributions in PG even when it is not common knowledge that they have been grouped together with other selfish players (i.e., even when condition (i) is not satisfied as in 2.c above); moreover, their contributions should be highly correlated with the amount that their partners expect them to contribute.

2.2 Strategic cooperation

Before describing our experimental design in detail, let us briefly consider the second possible mechanism of cooperation among selfish players identified by Gächter & Thöni (2005). Although our main focus in this study is the mechanism of group-cooperation heuristics, it is important to consider this second possible mechanism insofar as it could in principle present a relevant confounder. The mechanism in question concerns rational cooperation in a finitely repeated social dilemma (Kreps et al. 1982):

> LOW contributors actually believe that some other LOW contributors invested nothing in the Ranking experiment not because they are free riders, but because they are conditional cooperators with pessimistic beliefs. Then LOW contributors have an incentive to cooperate strategically until the ninth period to induce the conditional cooperators to contribute. They free ride in the final period, when cooperation is not in their rational self-interest anymore. Thus, if for whatever reason LOW contributors believe that some others are conditional cooperators, then rational cooperation is possible even in a finitely repeated cooperation game. (Gächter & Thöni, 2005, p. 311)

Note that this possibility crucially depends on the ambiguity about the exact types of other players. Gächter & Thöni's (2005) sorting rule leaves room for this ambiguity: subjects were sorted into HIGH, MIDDLE, and LOW contributors based on the level of contributions in their Ranking experiment. This way of grouping subjects cannot discriminate conditional cooperators with pessimistic beliefs from selfish players, because their behavior — low contributions — is equally compatible with both types.[1]

In order to control for this mechanism, and to focus on the mechanism of group-cooperation heuristics, we adopt a more fine-tuned sorting procedure, following de Oliveira et al. (2014), who removed type ambiguity by adopting, as a sorting procedure, the strategy-elicitation method (Fischbacher et al. 2001) in a separate online (incentivized) game. This procedure reveals the strategies underlying a subject's decisions, by asking him/her to state what he/she would contribute for each possible contribution level of the other(s), and thus makes it possible to identify selfish players while controlling for beliefs. So, when this more nuanced and precise information is common knowledge, the possibility of rational cooperation under uncertainty about others' types should be effectively removed. In contrast, without common knowledge, this mechanism should be activated. Our fine-tuned sorting procedure and systematic manipulation of common knowledge enable us to home in on group-cooperation heuristics as the source of selfish players' positive contributions under common knowledge, and thereby to rule out the potential confounder presented by Gächter & Thöni's (2005) second proposed mechanism.

2.3 Summary of our schema

Group-cooperation heuristics =

> Condition (i): realization of impossibility of free-riding and need to cooperate to make money. (Op-

[1] This is also the cases with a more recent study by Junikka et al. (2017).

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

erationalized as YES when it is common knowledge that both are typed as selfish; NO when common knowledge is removed)

+

Condition (ii): realization of others' expectation of mutual cooperation (Operationalized as YES when exchange of expectations is present; NO when it is absent)

Prediction: Cooperation will be boosted among selfish types only when both (i) and (ii) are satisfied (Scenario 1); alternatively, (i) or (ii) alone will boost cooperation to a weaker degree (Scenario 2).

3 Experimental design and implementation

We implemented a repeated, two-phase linear public goods game (voluntary contribution mechanism, or VCM), which was preceded by a sorting experiment (which we conducted online) in order to identify subjects' types (de Oliveira et al. 2014). We opted for a two-person design, unlike Gächter & Thöni (2005), who used a four-person design, or de Oliveira et al. (2014), who used a three-person design. Although there is no clear differences between these group sizes in terms of theoretical implications, our choice was motivated by the wish to focus on the effect of beliefs about others' expectations, which should be most straightforward in a two-player game. For similar reasons, both Jin & Yamagishi (1997) and Guala et al. (2013) also used two-person designs.

3.1 Online ranking experiment

The experiment was conducted at the Cognition and Behavior Lab at Aarhus University (Denmark) in March of 2015. All subjects gave their informed written consent.

We first recruited subjects from the subject database to participate in a study advertised as consisting of an online experiment and a lab experiment for a subset of subjects who completed the online experiment. They were informed that they could earn up to 35 DKK in the online experiment and up to 210 DKK in the lab experiment (if invited and participate). 227 subjects (125 females, mean age=24, range 18–59) completed the online public goods game on *Survey Monkey*, whereby we identified their types using the strategy elicitation procedure (Fischbacher et al. 2001).

After the standard comprehension questions were correctly answered, subjects were asked to make two decisions, one "unconditional" and the other "conditional". The first (i.e., unconditional) decision pertained to the amount they would contribute to the common pool out of the endowment (20 points = 20 Danish Kroner) in a one-shot public goods game (with two players, MPCR=.75). The second (i.e., conditional) decision consisted of making a schedule of one's own contribution corresponding to the full range of possible contributions (0–20 points) made by the other player. Subjects were told (i) they would be randomly matched with a partner; (ii) one of the two would be randomly picked as the "unconditional" player, whose "unconditional" decision was to be implemented, and the other's corresponding "conditional" decision would be implemented. The Nash/selfish strategy is to contribute zero in both decisions. However, if both partners decide to make the maximum contribution (mutual full contribution), each would earn 30 points whereas when neither contributes anything (mutual zero contribution) they each would earn 20 points (points were converted into DKK with 1 point=1 DKK). We identified three types with the following criteria:

- "Conditional cooperator" (CC): n=128, or 56% of the total number of subjects. These subjects increase conditional contribution weakly monotonically, that is, as their partner's contribution increases, they increase own contribution, or at least do not decrease it.[2]
- "Selfish player" (SS): n=49 (22%). These subjects' conditional contributions remain no more than five regardless of the other player's contribution.[3]
- "Other type": n=50 (22%). Those who are neither CC nor SS.

We checked the subjects for previous experience in the laboratory. There is no significant difference between CC and SS types with regards to experience.

Subjects were informed at this stage that they may be invited to the lab. In accordance with the protocol of the lab, subjects were asked to provide their (CPR) identification number, so that their payment could be transferred directly to their account. The average earning from the online experiment was 24,45 DKK, which was paid electronically after the experiment. Those who participated in the lab experi-

[2]Fischbacher et al. (2001, p. 401) include strategies that are not weakly monotonic in a strict sense as long as there is "a highly significant (at the 1% level) and positive Spearman rank correlation coefficient (between own and others' contribution)". Our criterion is not statistical but rather a mechanical application of the weak monotonicity. de Oliveira et al.'s (2014) exact criterion is not clear from their phrasing: "a strategy profile of a conditional cooperator involves higher contributions as expectations of others' contributions increase".

[3]Fischbacher et al. (2001, p.401) used the stricter criterion, according to which free-riders' conditional contribution is *always* zero. Since our overall design is closest to de Oliveira et al. (2014), we used their more permissible criterion, which classifies 'subjects who never give more than five [25% of endowment] (footnote 5)' as selfish. de Oliveira et al. (2014) report that "In the 21 decisions of the type elicitation task, over 90% of the decisions for our selfish subjects are zeroes, and over 98 % are either zero or one" (footnote 5). Our results are similar (94.1% and 98.5%, respectively). Of our 49 S-type subjects, 43 subjects always chose zero. 6 subjects' schedules (6*21=126) include positive contributions of 1 (46 times), 2 (11 times), and 3 (4 times).

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

TABLE 1: Summary of the experimental design.

		Type	
		Selfish players (SSs)	Conditional cooperators (CCs)
Information	Common knowledge (CK)	SS_CK (n=10)	CC_CK (n=12)
	No common knowledge (NCK)	SS_NCK (n=8)	CC_NCK (n=16)

ment received this amount, plus whatever they earned during the lab experiment, also electronically.

3.2 Lab Experiment

Following the online experiment a lab experiment was conducted with 21 people identified as SSs and 29 people identified as CCs.[4] The lab experiment used the same two-person VCM with MCPR=0.75, but the conversion rate was 10 points=3 DKK. Instructions for phase 1 can be found in Appendix 1. After reading the instructions all subjects had to pass comprehension questions to continue.

As can be seen in Table 1, the lab experiment implemented a 2x2 between-subject design with the following factors:

- Group Type: Conditional Cooperators (CCs) vs. Selfish Players(SSs);
- *Information* about the other player's type: Common Knowledge (CK) vs. No Common Knowledge (NCK).

We had 4 sessions, SSs with the CK condition (n=10), CCs with the CK condition (n=12), SSs with the NCK condition (n=8), CCs with the NCK condition (n=16).[5] In the CK condition, each subject was informed about (i) her own type and about (ii) the other player's type, and was also informed that (iii) her partner knew exactly as much as she did. In NCK, each subject was informed about (i) only.

Before the game started, each subject was asked three questions designed to reveal their beliefs and attitudes about social norms (Bicchieri, 2006): The questions were formulated as follows:

- 1) How much do you think each person should contribute to the group project in the first round?
- 2) What do you think the average answer of all the other subjects to the above question is?
- 3) You participated in an online experiment a couple of weeks ago. How do you think you contributed compared to the majority of subjects in the experiment? (1=less, 2=same, 3=more, 4=I don't remember)

Responses to these questions made it possible to corroborate the procedure by which we identified subjects' types: selfish types should report believing that they contributed less than the majority of the other subjects. This was indeed the case (more on this later).

The public goods game was played in two phases, each consisting of ten rounds. For the duration of each phase, subjects remained in stable pairs (partner design), but the pairs were shuffled and rematched prior to Phase 2. They were informed of this arrangement prior to the experiment. The two phases were identical except that in the second phase, but not in the first, each subject was prompted at the beginning of each round to indicate how much s/he expected her/his partner to contribute.[6] The number each subject gave was then communicated to their partner. Subjects were not informed of this addition to the game until the beginning of the first round in second phase. This manipulation has the same implementation advantage as the "cheap talk" that Cooper et al. (1990) and Clark et al. (2001) used in coordination games, namely, that in all sessions each subject has an identical role and can be given identical instructions. However, while Cooper et al. (1990), as well as Clark et al. (2001), asked their subjects to state their own *intention* in advance, we asked subjects to state how much they *expected* their partner to choose. Concretely, Cooper et al. (1990) asked subjects to complete the following sentence: "I INTEND TO CHOOSE___" (i.e., exchange of own intention), while we asked our subjects to respond to the following question: "How many points do you expect that the other subject in your group will contribute in this period?" (exchange of own expectation). This design reflects our focus on the role of expectations.[7] All subjects were prompted at the beginning of each round to make a non-binding, non-incentivized announcement to this question. Each player was then informed as to what her partner had announced prior to making a decision as to how much to contribute in that round.

[4]There was no significant difference between the people who showed up for the lab experiment and the people who opted out of participation in the lab experiment in terms of age, gender, education level, studies of economics/business/Maths, profit earned in online experiment or contribution in online experiment.

[5]We had the fifth, mixed session with a small group (3 SSs and 1 CC). We piloted skin conductance measurements with these subjects, and the data from this session are not included in the following analysis.

[6]This prediction was not incentivized, so it was "cheap talk" (see Crawford 1998 for a survey of "cheap talk" experiments).

[7]Isaac and Walker (1988), in their VCM experiment, allowed more field-like face-to-face communication (max 4 minutes) with some imposed rules on what can be communicated (such as no side-payments outside the experiment).

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

TABLE 2: Comparison of unconditional online contributions by selfish and conditional cooperator types, standard deviations in parentheses.

	SS	CC	
CK	5.100 (7.652)	15.417 (5.823)	
NCK	5.714 (6.701)	12.643 (6.890)	
Total	5.353 (7.185)	13.923 (6.449)	U = 365.000; p = .000
	U = 33.000; p = .837	U = 67.500; p = .376	

The lab public goods game was programmed and implemented using the software z-Tree (Fischbacher, 2007), and presented to subjects seated at desktop computers with 22 inch monitors (active display area: 474 mm × 296 mm). They gave their responses using mouse devices and keyboards. Each session lasted about 1 hour, and the average earning from the lab experiment was 160.92 DKK (about 22 euro). This includes a show-up fee of 40 DKK (about 5 euro).

4 Results

4.1 Online experiment and belief elicitation

We ran a pre-analysis on the online contributions to check that the subjects in the CK and NCK conditions did not initially differ in their online contributions in a way that could bias the further analyses. The descriptive data of the unconditional contributions in the online experiment, arranged by type, is shown in Table 2 (with Mann-Whitney U tests of differences). We found no apparent differences between the subjects relegated to the CK and NCK groups, for either the selfish or conditional cooperator types. Despite the small sample size, the conditional cooperators clearly contributing more to their partners than the selfish players.

We also found that the unconditional online contributions did not differ significantly from the contributions made in the first round of the lab-based experiment. Apparently, receiving the feedback about their "type" had no effect on the subjects' subsequent contribution.

Finally, the majority of selfish subjects (accurately) judged that they had given less than the average, and the majority of conditional cooperators judged they had given either the same or higher amounts than the average. However, the two types of subjects did not differ significantly in their beliefs about how much one should give in the first round and their predictions of what others think one should give did not differ across types.

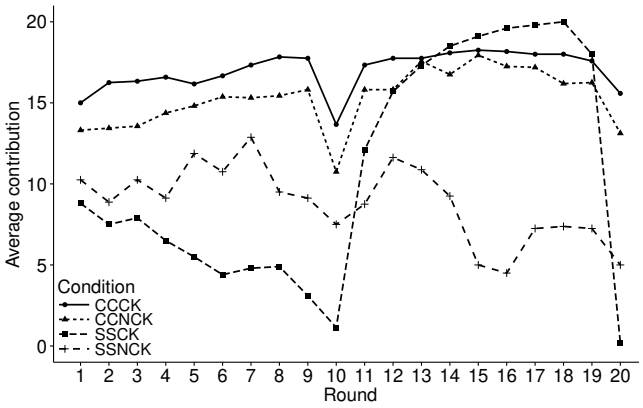


FIGURE 1: Average contributions by round and phase. In the second phase, rounds 11–20, subjects stated their expectations.

4.2 Impacts of common knowledge and signaling of expectations

Figure 1 shows the average contributions across rounds and conditions in the two experimental phases (for more detailed information about contributions across rounds, refer to Appendix 2). It is evident that the trajectories are quite different, especially for the subjects in the SS-CK condition. While in the first phase the trend follows a downward slope, in the second phase (with expectations) the SS-CK contributions rise with each time point to reach their peak between rounds 17 and 19, then drop drastically to zero in the final round.

The differences between the CK- and NCK-selfish types' contributions in Phase 1 (no communication of expectations) and Phase 2 (communication of expectations) are striking: while the NCK subjects contribute more in the later rounds of Phase 1 as the CK subjects' contributions drop, in Phase 2 this trend is altered by the fact that the selfish subjects in the CK condition contribute as much, and even more than the conditional cooperators, whereas the contributions in the NCK condition fall around round 14 and do not reach the same levels of efficiency. Looking at the same figure, the contributions between conditional cooperators in both the CK and NCK condition do not seem to differ to a significant degree and follow a similar pattern, though slightly higher in the case when the type is common knowledge.

Put simply, cooperation is increased in selfish types if and only if they communicate expectations and have common knowledge of their types. The communication of expectations by itself has no clear effect on any other group, and common knowledge alone does not improve the cooperation of the selfish types.

Given the small sample size, we report tests of only the results most relevant to the main hypotheses. For the means of rounds 1–9, SS-CK was non-significantly lower than

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

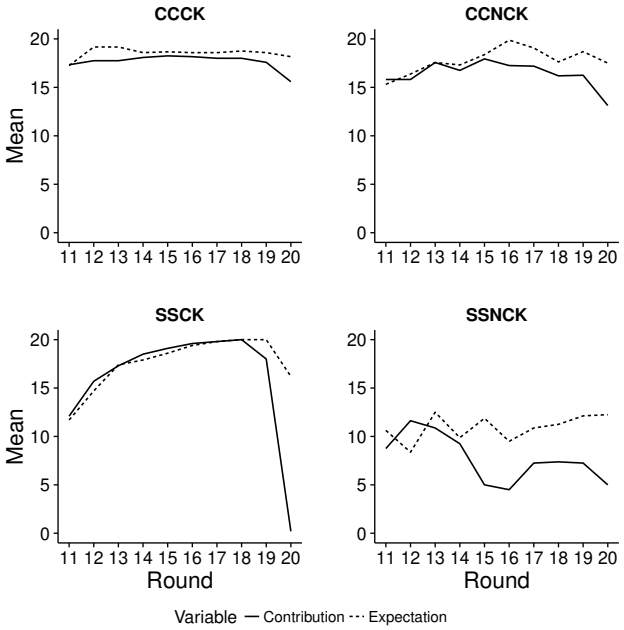


FIGURE 2

SS-NCK for Phase 1 but, importantly, higher in Phase 2 (p = .008, 1-tailed Wilcoxon test using pairs as the units of analysis).[8] For the CC subjects the difference between CC-CK and CC-NCK was not significant in either Phase. As should be expected, CC subjects contributed more than SS subjects in both CK and NCK conditions except for the CK condition in Phase 2, although this difference was significnt only for the CK condition in Phase 1 (p = .004 one tailed) and the NCK condition in Phase 2 (p = .025).

Partners' expectations had a significant effect on the contributions, namely that with every 1-point increase in the expectation, the contribution increased on average by .390 (see Figure 2 for the relationship between the average contribution and expectation in each condition). Finally, the end-game effect was observed to some extent in all conditions. The most drastic decrease from the mean of the first 9 rounds to the last round appeared in the SS-CK group (p = .010 in Phase 1, .000 in Phase 2, by one-tailed t test with pairs as the unit of analysis). The drop was significant for the CC-NCK condition (p = .008 in Phase 1, .028 in Phase 2) but was not significant for the CC-CK and NCK-SS conditions in either Phase.

[8]A less conservative analysis (based on the lmer() function in the lme4 package in R) treated pair as a random effect and subject as the unit of analysis, yielding p = .001. With a larger sample, this analysis would be preferred, but in this case the assumption of homoscedastic error was seriously violated. Here and elsewhere, we rely on the analysis by pairs. In no case was the random-effect analysis significant at $p < .05$ when the analysis by pairs was not.

4.3 Own and partner expectations

In an exploratory analysis of the relationships between own and partner's expectations and subsequent contributions, we found no correlation between own or partner's expectations on contribution, from one round to the next, in SS-NCK. Contributions in first round significantly correlated with one's own expectations for SS-CK and CC-NCK. For CC-CK contributions are correlated with own and other's expectations.

Specifically, Kendall's tau-b correlations between the three variables were calculated separately for each round of Phase 2 in the four conditions. The SS-NCK group differed from the rest in that neither the other's nor one's own expectations were significantly correlated with the actual contributions of the subjects across most of the rounds (the only exception being a significant, positive correlation between the subjects' own expectations and their contributions in R3: τ_b = .816; $p < .01$). Interestingly, in the SS-CK and the CC-NCK groups, the contributions in the first round were significantly correlated with one's own stated expectations (τ_b = .764; $p < .01$; τ_b = .678; $p < .01$, respectively), but not with that of the partner (τ_b = .422; p = .124; τ_b = .199; p = .383, respectively). In the CC-CK group, the correlation between one's own and the partner's expectation with the contribution was the same (τ_b = .761; $p < .01$), while they were also significantly inter-correlated (τ_b = .548; $p < .05$), which doesn't allow for a straightforward interpretation.

Finally, in the last round expectations are significantly correlated with contributions for SS-CK, CC-CK and CC-NCK. In the last round, the shadow of the future is removed. Here, the expectations of the SS-CK players are significantly *negatively* correlated with their contributions ($\tau_b = -.667$; $p < .05$), whereas in both the CC-CK and CC-NCK conditions, players' expectations are significantly *positively* correlated with their contributions (τ_b = .616, $p < .05$; τ_b = .681; $p < .01$, respectively).

5 Discussion

The present study was designed to reveal conditions under which group-cooperation heuristics among selfish players would be triggered. The findings clearly support Scenario 1, according to which the heuristics is triggered only when both conditions (i: realization of no possibility of free-riding and need for cooperation to make money) and (ii: realization of other's expectations of cooperation) are satisfied, but not by two conditions individually (Scenario 2).

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

Let us first consider what happened in Phase 1.[9] The common knowledge manipulation had opposite effects on CCs and SSs. While CCs in the CK condition (unsurprisingly) contributed more than in the NCK condition, this pattern was reversed for SSs. The slightly negative impact of common knowledge on SSs' contributions merits close attention. This indicates the ineffectiveness of condition (i) independently to trigger group-cooperation heuristics among selfish players (in line with Scenario 1.b not 2.b). Indeed, the result suggests that making common knowledge of selfish types explicit could be counterproductive. de Oliveira et al. (2014, 125), who used the same CK/NCK treatments, also report that the decay of the contributions among SSs in the CK was steeper, which is consistent with our results. How should we account for this possible negative impact of common knowledge? One plausible account is that the explicit common knowledge suppressed the prospect of strategic cooperation under ambiguity about the other's type, as we discussed in Section 2.2 above.

Let us now turn to Phase 2. Here, signaling of expectations dramatically increased contributions among SSs in the CK condition, as opposed to the NCK condition. This suggests that common knowledge of selfish types (condition (i)) and signaling of expectations (condition (ii)) jointly trigger group-cooperation heuristics. It is also important to note that condition (ii), like condition (i), appears to be insufficient on its own as even a weak trigger for group-cooperation heuristics (in line with Scenario 1.c). Contra the default conformity preference hypothesis, expectations alone do not seem to facilitate cooperation among selfish subjects. Although subjects' contributions are highly correlated with the amount that their partners expect them to contribute (Figure 2), the correlation between each subject's own behavior and her/his partner's expectations was apparently higher in the CK conditions than in NCK conditions (for both SSs and CCs). In addition, selfish subjects' expectations were higher in the CK than in the NCK condition. That is, selfish players, who did not know the other's type did not signal high expectations to begin with, and the other selfish players did not conform even to such moderate expectations. Thus, neither (i) nor (ii) is independently sufficient.

In view of the strong interaction that we observed between common knowledge of (selfish) types and the signaling of expectations, we conclude that condition (i) is dependent on (ii) as a trigger of group-cooperation heuristics, in line with Scenario 1. Presumably the selfish types in the CK condition without expectations were aware that there were no cooperators around to free-ride on, and that they would therefore have to cooperate in order to make money. But this was insufficient to make them cooperate; indeed, it had, if anything, a negative impact on contributions. This is not surprising given that selfish subjects in the CK condition must not only decide that mutual cooperation is the best available option but must also have some assurance that their partner also sees it that way – and also that their partner believes that they see it that way (because otherwise their partner might be reluctant to contribute despite being willing to). The opportunity to announce expectations, introduced in Phase 2 of the experiment, provided selfish players with a chance to signal such an assurance.

We see at least two distinct ways in which others' expectations and common knowledge of types jointly contribute to successful cooperation in line with Scenario 1.a. First, the normative force of expectations may be amplified by in-group favoritism (i.e., expectations matter more when they come from in-group members). Second, common knowledge of types may provide a rationale for each other's expectations, based on material self-interest, thus increasing their credibility. Although these mechanisms would work in the same direction, they are in fact distinct. We believe that our data is better explained by the latter. Let us explain.

5.1 Solution thinking

How and when is common knowledge of types conducive to selfish cooperation? In order to see this, we need to look somewhere other than to the orthodox best-reply reasoning in game theory. Morton (2003) proposes such an alternative, simulation-based model of reasoning, which he calls *solution thinking* (see also *mirror strategy*, Hurley 2005; *common reasoning*, Cubitt and Sugden 2014). Solution thinking proceeds in the following steps (cf. Guala 2016, ch. 7):

1. C is the obvious solution to the problem.
2. The other also thinks that C is the obvious solution to the problem.
3. To achieve C, I must do c_i and the other must do c_j.
4. The other also thinks that I must do c_i and she must do c_j.
5. Therefore, I do c_i.

Steps 1 and 3 correspond to condition (i) of group-cooperation heuristics: to realize that "since there is no one to free-ride on, the only way to make money is to cooperate." But in order to arrive at Step 5, one needs some confidence that the other is thinking in the same way (Step 2) and expecting the same way (Step 4). One plausible interpretation of the joint effectiveness (and disjoint ineffectiveness) of common knowledge of type and signaling of expectations is that the former activated Steps 1 and 3, and the latter Steps 2 and 4, thereby activating solution thinking. The announcement

[9] An anonymous reviewer pointed out the possible order effect of Phases 1 and 2. We agree that this possibility needs to be studied experimentally, but given practical limitations of the present study we opted for this order, mainly because it is the most natural way to see additional impact of signaling of expectations. Swapping the order of the two phases may create some carry-over effect of signaling to the next phase, which would be an interesting finding, but finding such an effect has not been a focus of our study here.

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

of (high) expectations in Phase 2 assured selfish subjects in the CK condition that their partner was also willing to try a new strategy (Step 2), and also that their partner believed that they did too (Step 4). This last step is crucial because, in its absence, there would be uncertainty about whether one's partner might do her part. In contrast, these steps were not activated in the NCK condition, because the pre-conditions for simulation is missing (one could not eliminate the doubt: "Is C the obvious solution to the other?"; nor could one make clear sense of the other's (high) expectation). We do not deny the possibility that the normative power of expectations to conform were at work, but the very strong end-game effect among selfish players in the common knowledge condition (basically no one conformed to others' expectations) suggests that the pressure to conform is not enough to motivate cooperation in social dilemma.

This rather unorthodox explanation becomes more plausible when we consider how the best-reply approach would accommodate our observations. Selfish cooperation is traditionally explained in terms of rational vs. boundedly rational behavior. The former is formulated by Kreps et al. (1982), and the latter by Selten and Stoecker (1986). But it would be rather *ad hoc* to explain the shifts of behavior of selfish subjects in the CK condition from Phase 1 to Phase 2 by saying that selfish subjects were rational in Phase 1, but they became less rational or more boundedly rational in Phase 2, where the only change introduced was the exchange of expectations at the beginning of each round. This is not coherent as an explanatory strategy. We therefore have good reason to favor the solution thinking model.

The solution thinking model also generates unique and testable predictions, which would be important for future research to address. For example, the solution thinking model predicts that targeting second-order beliefs (i.e., beliefs about the other's expectation), as in the present study, is more effective than targeting first-order beliefs (i.e., belief about the other's behavior), e.g., by allowing subjects to announce their own intentions. This is because the signaling of expectations more directly targets step 4 of the solution thinking model. The solution thinking model also predicts that it makes no difference whether the expectation is framed in normative or descriptive terms, e.g., "I believe that the other ought to contribute X" vs. "I believe that the other will contribute X". This is because what triggers solution thinking is not a preference for conformity per se but the confidence that the other is thinking in the same way. The conformity-based explanation, in contrast, would predict a discrepancy between the two conditions.

On a more general note, our results also highlight the need of further investigation of subject types. There is much uncertainty as to the characterization and stability of these "types" (Moffatt, 2016, p. 10). Since conditional cooperators are characterized by the dependence of their behavior on beliefs about others' behavior, this type cannot be identified with a simple other-regarding utility function. Similarly, our main results give us reason to doubt that selfish players can be identified with a fixed utility function (i.e., with selfish preferences): even the behavior of a relatively well-established category, namely selfish types (or free-riders), turned out to be sensitive to expectations. This is not to say that there is no such thing as types at all. Indeed, we observed systematic differences in the ways in which different types responded to our manipulations. Rather, what our results suggest is the need to take different ways of reasoning or thinking into account in addition to preferences when categorizing subjects into types in the context of social dilemmas.

In particular, we might be able to categorize our selfish subjects (thus categorized based on our one-shot strategy method) further into three types[10]: (1) unconditional free-riders, who always contribute zero, regardless of beliefs about others' types or behavior; (2) solution thinkers, who contribute positive amounts if they believe there are no cooperators to free-ride on, and this amount is positively correlated with others' expectation of their contribution; and (3) strategic cooperators, who contribute if they believe they are playing against a conditional cooperator. A fully structural finite mixture modelling approach (Moffatt, 2016, ch. 8) should be used to precisely specify these types and systematically investigate their distribution. The small sample size of the present study precludes such an approach, or making conclusive inferences about the dynamics of selfish cooperation.[11] Nevertheless, our findings are potentially significant, and point to promising new avenues of research on the hitherto under-investigated psychological mechanism (group-cooperation heuristics as solution thinking), which could be exploitable to induce cooperation among those who are conventionally categorized as selfish or free-riders.

References

Bardsley, N. & Moffatt, P. G. (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, *62*(2), 161–193.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.

Billig, M. & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology, 3*(1), 27–52.

[10]We thank an anonymous reviewer for suggesting these types.

[11]This is due mainly to two factors: first, only a very small proportion of the Danish student population seem to be Selfish-type in the first place; and second, many of these Selfish-type subjects did not come to the lab experiment. In contrast, conditional co-operators were more abundant. We aim to secure resources to recruit a sufficient number of selfish-type subjects in the future. In the meantime, we would like to invite replications of our results concerning the effect of expectation-signaling on those categorized as the selfish-type.

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive- motivational analysis. *Psychological bulletin, 86*(2), 307–324.

Burlando, R. M. & Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics, 8*, 35–54.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Clark, K., Kay, S. & Sefton, M. (2001).When are Nash equilibria self-enforcing? An experimental analysis. *International Journal of Game Theory, 29*, 495–515.

Cooper, R., DeJong, D. V., Forsythe, R. & Ross, T. W. (1990). Selection Criteria in Coordination Games: Some Experimental Results. *The American Economic Review, 80*(1), 218–233.

Crawford, V. (1998). A survey on experiments on communication via cheap talk. *Journal of Economic Theory, 78*, 286–298.

Cubitt, R. P. & Sugden, R. (2014). Common reasoning in games: A Lewisian analysis of common knowledge of rationality. *Economics and Philosophy, 30*(3), 285–329.

Dana, J., Cain, D. & Dawes, R. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes, 100*, 193–201.

de Oliveira, A. C., Croson, R. T. & Eckel, C. (2014). One bad apple? Heterogeneity and information in public good provision. *Experimental Economics*, 1–20.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics, 10*(2), 171–178.

Fischbacher, U. & Gächter, S. (2010). Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Goods. *American Economic Review, 100*(1), 541–56.

Fischbacher, U., Gächter, S. & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397–404.

Guala, F. (2016). *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton: Princeton University Press.

Guala, F., Mittone, L. & Ploner, M. (2013). Group membership, team preferences, and expectations. *Journal of Economic Behavior & Organization*, *86*, 183–190.

Gunnthorsdottir, A., Houser, D. & Kevin McCabe, K. (2007). Disposition, history and contributions in public goods experiments, *Journal of Economic Behavior & Organization, 62*(2), 304–315.

Gächter, S. & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association, 3*(2–3), 303–314.

Heintz C., Celse J., Giardini F. & Max S. (2015) Facing expectations: Those that we prefer to fulfil and those that we disregard. *Judgment and Decision Making, 10*(5), 442–55.

Hurley, S. (2005). Social heuristics that make us smarter. *Philosophical Psychology, 18*(5), 585–612.

Isaac, R. M. & Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic inquiry, 26*(4), 585–608.

Jin, N. & Yamagishi, T. (1997). Group heuristics in social dilemma. *Shakai Shinrigaku Kenkyu, 12*(3), 190–198. [in Japanese]

Junikka J., Molleman, L., van den Berg P., Weissing, F.J. & Puurtinen, M. (2017). Assortment, but not knowledge of assortment, affects cooperation and individual success in human groups. *PLoS ONE 12*(10), e0185859.

Kreps, D., Milgrom, P., Roberts, J. & Wilson, R. (1982). Rational Cooperation in the Finitely Repeated Prisoner's Dilemma, *Journal of Economic Theory, 17*, 245–52.

Moffatt, P. G. (2016). *Experimetrics: Econometrics for experimental economics*. New York: Palgrave Macmillan.

Morton, A. (2003). *The importance of being understood: Folk psychology as ethics*. New York: Routledge.

Ockenfels, A. & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior & Organization, 108*, 453–462.

Selten, R. & Stoecker, R. (1986). End behavior in sequences of finite prisoner's dilemma supergames. A learning theory approach. *Journal of Economic Behavior & Organization, 7*, 47–70.

Wit, A.P. & Wilke, H.A.M. (1992) The effect of social categorization on cooperation in three types of social dilemmas, *Journal of Economic Psychology, 13*(1), 135–151.

Appendix 1: Instructions for phase 1.

Welcome to the Cognitive and Behavioural Lab at Aarhus University. You will be participating in an experiment financed by Aarhus University.

All participants in this experiment participated in a similar game online last week. The online experiment was the same for all participants.

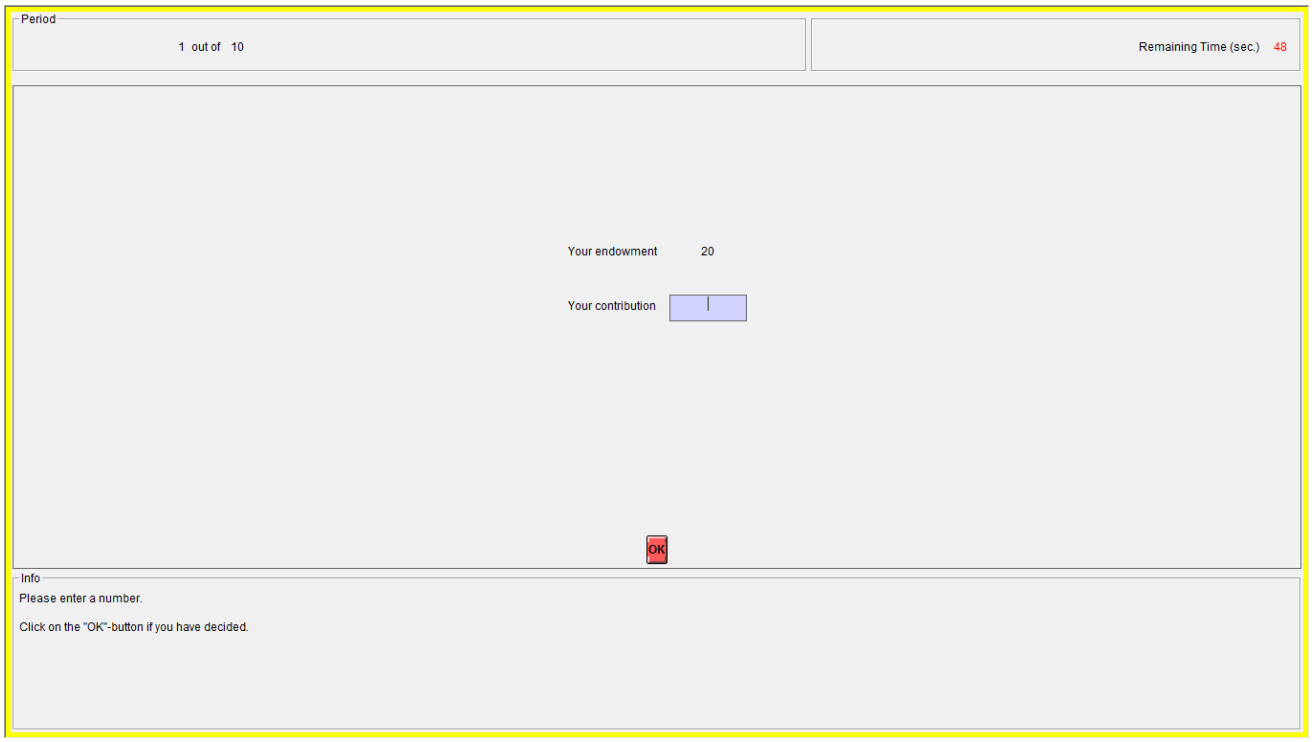
This experiment has two phases. These are the instructions to Phase 1 the instructions for Phase 2 will come later. All participants have the same instructions. Please read them carefully.

It is not allowed to communicate with other participants during the experiment. If you have any questions please raise your hand.

You can earn money in this experiment. Your income will be calculated by the computer during the experiment. You will be told how much you earned and the amount will be transferred to your bank account. During the experiment we do not calculate in kroner but in experimental points. At the end the points you earn will be converted into kroner. The points are converted into kroner with following conversion rate:

10 points = 3 kroner.

All participants are randomly selected into groups of two participants. Thus you are not necessarily seated next to the other person in your group. You will remain in the same group throughout the first phase of the experiment but you will never receive any information about who these people are and they will not be informed of your identity.



These are the rules:

Phase 1 has 10 periods. In the beginning of each period each participant receives 20 points. We call this his/her endowment. You now have to choose how you want to invest your endowment. You can put some or all in a project or keep them to yourself. The consequences of your choice are shown below. In the beginning of each period you will see the following screen on your computer:

Number of periods in phase 1 and which period you are in now is shown in the top left corner. Top right corner show how many seconds you have left to make your decision. You should make your decision before the clock goes to zero. The program is in English.

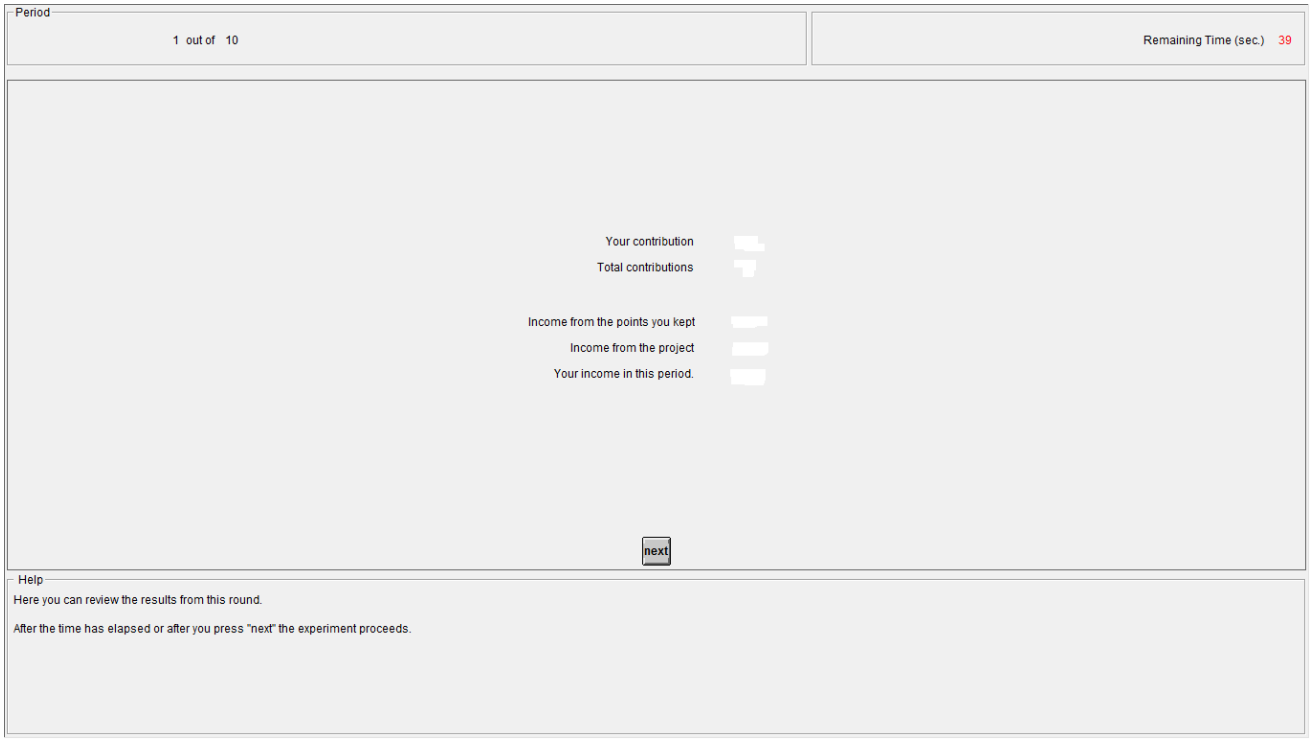
Your Endowment in the beginning of each period is 20 points. You decide how many points you wish tom contribute to the project by typing a number from 0–20. When you choose how many points to contribute you automatically choose how many points to keep to yourself. This is your Endowment (20 points) minus your contribution to the project.

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

Your income for the period = (20 – your contribution) + (0,75*(sum of contributions in your group))

When you have typed your contribution please press "OK". You choice have now been registered and can no longer be changed.

When the other participant in your group has decided on his/her contribution and pressed “OK” you will see following screen:



We have blanked out the numbers but to get an idea you can see what information the screen will reveal. The screen shows first Your contribution to the project, then the Total contributions. Below it states the Income from the points you kept in this period, then your Income from the project and below that Your income in this period.

As explained above your income consists of two parts:

1. The points you kept to yourself = 20 point – the points you contributed to the project
2. Income from the project, which is calculated in the following way: Your income from the project = 0,75 * (”sum of all contributions to the project in your group”) Therefore the income from the project is the same for all participants in your group. Assume for example that the sum of contributions to the project is 20 points. Then you and the other participants in your group will receive 0.75*20=15 points from the project.

So your income for the period = The points you kept to yourself + Income from the project.

If you have any questions please raise your hand and we will come to you.

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press

Appendix 2: Mean individual contributions across rounds 1–10, Phase 1, and mean contributions and expectations in rounds 11–20, Phase 2, across conditions. Standard deviations in parentheses.

	Phase 1									
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
SS-CK (n = 10)										
Contribution	8.800	7.500	7.900	6.500	5.500	4.400	4.800	4.900	3.100	1.140
	(8.203)	(7.649)	(7.978)	(8.182)	(6.916)	(5.211)	(5.903)	(6.488)	(4.977)	(2.271)
SS-NCK (n = 8)										
Contribution	10.250	8.875	10.250	9.125	11.875	10.750	12.875	9.500	9.125	7.500
	(7.459)	(7.846)	(7.402)	(7.736)	(5.939)	(6.735)	(7.120)	(8.452)	(9.230)	(10.351)
CC-CK (n = 12)										
Contribution	15.000	16.250	16.333	16.583	16.167	16.667	17.333	17.833	17.750	13.667
	(6.368)	(4.901)	(4.292)	(4.033)	(4.629)	(4.677)	(3.798)	(3.738)	(4.634)	(8.038)
CC-NCK (n = 16)										
Contribution	13.313	13.438	13.563	14.375	14.813	15.375	15.313	15.438	15.813	10.750
	(6.172)	(5.750)	(6.491)	(6.386)	(5.648)	(5.476)	(5.338)	(6.450)	(5.180)	(8.505)
	Phase 2									
	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20
SS-CK (n=10)										
Contribution	12.100	15.700	17.300	18.500	19.100	19.600	19.800	20.000	18.000	.200
	(5.626)	(5.579)	(4.373)	(3.171)	(1.912)	(.843)	(.632)	(.000)	(6.325)	(.632)
Expectation	11.700	14.700	17.400	17.900	18.600	19.400	19.800	20.000	20.000	16.200
	(6.750)	(5.774)	(4.222)	(3.479)	(2.271)	(1.350)	(.632)	(.000)	(.000)	(8.011)
SS-NCK (n=8)										
Contribution	8.750	11.625	10.875	9.250	5.000	4.500	7.250	7.375	7.250	5.000
	(8.763)	(6.739)	(8.132)	(8.155)	(7.071)	(6.633)	(6.316)	(8.245)	(10.025)	(9.258)
Expectation	10.625	8.375	12.500	9.875	11.875	9.500	10.875	11.250	12.125	12.250
	(6.781)	(6.610)	(6.949)	(8.576)	(7.529)	(7.131)	(6.642)	(7.723)	(8.097)	(10.167)
CC-CK (n=12)										
Contribution	17.333	17.750	17.750	18.083	18.250	18.167	18.000	18.000	17.583	15.583
	(4.355)	(4.351)	(4.351)	(4.274)	(3.957)	(4.303)	(4.671)	(4.671)	(4.699)	(6.612)
Expectation	17.250	19.167	19.167	18.583	18.667	18.583	18.583	18.750	18.583	18.167
	(3.671)	(1.946)	(1.946)	(3.630)	(3.085)	(3.630)	(3.630)	(3.108)	(3.630)	(3.738)
CC-NCK (n=16)										
Contribution	15.813	15.813	17.563	16.750	17.938	17.250	17.188	16.188	16.250	13.125
	(4.385)	(5.913)	(3.983)	(6.846)	(4.389)	(6.768)	(6.316)	(7.064)	(7.188)	(9.465)
Expectation	15.313	16.375	17.563	17.313	18.375	19.875	19.063	17.625	18.687	17.500
	(4.270)	(4.856)	(3.577)	(5.474)	(3.442)	(.500)	(2.720)	(5.058)	(2.983)	(5.477)

https://doi.org/10.1017/S1930297500008883 Published online by Cambridge University Press