University of Cyprus
Department of Economics

Working Paper 04-2019

The Virtuous Cycle of Agreement

Philippos Louis, Matias Núñez and Dimitrios Xefteris

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

The Virtuous Cycle of Agreement

Philippos Louis* Matias Núñez† Dimitrios Xefteris‡

University of Cyprus Université Paris-Dauphine University of Cyprus

March 2019

Abstract

Collective choice mechanisms are used by groups to reach decisions in the presence of diverging preferences. But can the employed mechanism affect the degree of post-decision actual agreement (i.e. preference homogeneity) within a group? And if yes, which are the features of the choice mechanisms that matter? Since it is difficult to address these questions in natural settings, we employ a theory-driven experiment where, after the group collectively decides on an issue, individual preferences can be properly elicited. We find that the use of procedures that promote apparent consensus with an outcome (i.e. agreement in manifest behaviors) generate substantially higher levels of actual agreement compared to outcome-wise identical mechanisms that push subjects to exaggerate their differences.

JEL classification codes: D71, D72.

Keywords: implementation; mechanism design; consensus; agreement; congruence; experiment; endorsements.

1 Introduction

Collective choice procedures are used to allow groups to make decisions in the presence of disagreement. Not everyone in a group is expected to have the same preferences on any given issue, but high levels of disagreement can be detrimental to the functioning of collective entities. Preference misalignment may undermine the effective implementation of a group's decisions and hinder a team's performance

* *Department of Economics, University of Cyprus, P.O. Box 20537, Nicosia 1678, Cyprus. email: louis.philippos@ucy.ac.cy*

† *Université Paris-Dauphine, PSL Research University, CNRS, Lamsade, 75016 Paris, France. email: matias.nunez@dauphine.fr*

‡ *Department of Economics, University of Cyprus, P.O. Box 20537, Nicosia 1678, Cyprus. email: xefteris.dimitrios@ucy.ac.cy*

in general. Achieving higher levels of agreement within a group is therefore seen as desirable.[1] The economics literature typically assumes that individual preferences are stable over time and do not depend on past behavior. While convenient for analytical purposes, this assumption is unlikely to always hold.[2] In fact, contrary to this assumption, scholars that study group decision making in organizations seem to suggest that the process through which decisions are made, and the way individuals behave during such process, have an important influence on the group members' preferences, and hence on the degree to which they agree with each other. Ultimately, whether agreements can be fostered is an empirical question, the answer to which we seek in this paper.

It is critical for our research to have a clear definition of agreement in the context of collective choice. In particular, we need to differentiate between group members' *actual agreement* with each other (i.e. the similarity of actual individual preferences) and *apparent consensus with an outcome* (i.e. individuals' agreement with an outcome in manifest behaviors).[3] During decision making, group members are given the opportunity to express their degree of agreement with possible outcomes. The exact way this is done varies: they can make proposals, accept or reject them, cast a veto against them; vote in favor or against different alternatives, rank them, approve a number of them, abstain, etc. A high level of *apparent consensus with an outcome* may be observed (e.g. a unanimous vote in favor of a proposal) without precluding substantial levels of *actual disagreement* remaining. This distinction is especially important as the main hypothesis we want to test in this paper is whether high levels of *apparent consensus with an outcome* lead to high levels of *actual agreement* within a group.

The incentives for an individual to express her agreement with a specific outcome within the framework of a collective choice mechanism will depend on the details of said mechanism. In some cases, these may be such that it is in one's best interest to exaggerate her disagreement with a specific proposal in order to move the final outcome closer to what she deems optimal. In other cases, individuals can reap higher benefits by being more accommodating. The closer a mechanism is to the latter type, the more *congruent* we say it is. If our hypothesis is true, then designing and using *congruent* mechanisms can help groups enter a "virtuous circle of agreement": by promoting apparent consensus with an outcome, we push, eventually, towards higher levels of actual agreement.

A major problem with trying to answer this question in natural settings is the endogeneity of out-

[1]See De Dreu and Weingart (2003) and De Wit, Greer, and Jehn (2012).

[2]See Becker (1993), Aaron (1994) and Bowles (1998) for important discourses related to this issue.

[3]See Kelman (1958), Priem, Harrison, and Muir (1995) and Kellermanns, Walter, Lechner, and Floyd (2005) for detailed discussions of this distinction.

comes: different mechanisms often produce different outcomes.[4] If levels of apparent consensus with an outcome change together with it, identification of an effect of consensus on actual agreement becomes impossible. Even if one finds a change in actual agreement, this may be because the outcome is "better" in some dimensions. Furthermore, in real life the contexts in which different rules apply are very disparate and the mechanisms themselves differ in a number of aspects apart from the agreement incentives that they provide (e.g. formal versus informal deliberation, time constraints, and revision procedures). Therefore, we need to turn to more controlled environments to effectively isolate the value of agreement in group decisions.

Overcoming the endogeneity issue described above, even in the laboratory, is not trivial. To achieve this we use classic (Moulin, 1980) and more recent (Yamamura and Kawasaki, 2013; Núñez and Xefteris, 2017) findings regarding mechanism design in the single-peaked context. In particular, we compare subjects' behavior under two decision mechanisms: the Median Approval and the Simple Mean. These two rules are *outcome-wise identical*: For any given preference profile, they result in the same unique Nash equilibrium outcome. At the same time, they largely differ in the incentives for apparent consensus. Considering that the outcome space is the unit interval, the Median Approval rule allows each individual to support any subset of alternatives (i.e. give one vote to as many alternatives as she wants) and the outcome coincides with the median of the distribution of the votes cast by all voters. Since players have single-peaked preferences (i.e. each voter is characterized by an ideal policy and prefers that the outcome is as close as possible to her ideal policy), incentives are such that in equilibrium the implemented policy is included in all individuals' sets of approved alternatives: everyone appears to be in consensus with the outcome. According to the Simple Mean rule, each individual reports a number, and the outcome coincides with the mean of the reports. The incentives lead individuals to vote for extreme alternatives. In equilibrium, this leads to an exaggerated disagreement between individual votes and the implemented policy. Thus, the Median Approval is a congruent mechanism and the Simple Mean an incongruent one. However, they both apply to the same class of problems, and, more importantly, they produce identical outcomes.

After subjects make a collective decision under one of these two mechanisms (in Part A), they move to a random dictatorship phase (in Part B), where each of them is allowed to propose a revision of the original decision, and each of the proposals gets implemented equiprobably. In this manner, we elicit the

[4]For example, all decision rules in Schweiger, Sandberg, and Ragan (1986) and similar studies typically lead to different resolutions.

individuals' preferences after the collective choice is made. This allows us to gauge the level of agreement within a group. Any differences that we observe between the Median Approval and the Simple Mean treatment should be attributed to the mechanisms and the way individuals behave differently in each one of them. That is, since both mechanisms deliver similar outcomes, a potential difference in post-decision levels of actual agreement should be due to factors that are not related to the outcome itself or to initial preferences.

In Part A we do not find important differences between the outcomes of the Median Approval and the Simple Mean treatment: Conditional on the group's preference profile, the two mechanisms implement very similar outcomes, just as theory predicts. We do find, though, that the apparent consensus for the winning alternative is much larger under the Median Approval treatment than under the Simple Mean treatment. For instance, in about 80% of the cases a subject endorsed the implemented alternative in the Median Approval treatment, while in the Simple Mean treatment individual votes were substantially far from the outcome: on average one-third of the total measure of the alternatives' space. In Part B, we compare the level of ex-post agreement in the two treatments by measuring the dispersion of within-group proposals. For both treatments we find that dispersion is smaller than that of the exogenously given payoff-maximizing points; a proxy for ex-ante disagreement. But *the reduction in dispersion is double in size in the Median Approval treatment compared to the Simple Mean one*. In other words, we find that following a decision made using a more congruent mechanism, groups exhibit higher levels of actual agreement. Importantly, this large difference is statistically significant at any conventional level.

After finding this empirical support to our main hypothesis we take a closer look at the individual behavior to understand what drives the result. We observe that under both the Median Approval and the Simple Mean treatment the proposal of a subject is, essentially, a convex combination of her ideal policy and the original group decision. That is, we find that individuals incorporate the outcome of a collective decision process into their own utility function. It then operates as a common attractor for all group members preferences, bringing them closer together. Importantly, this effect has different magnitudes in the two mechanisms: *the weight assigned to the original decision is about 30% in the Median Approval treatment and about 10% in the Simple Mean treatment.*

We construct a measure of the apparent consensus of a voter with the outcome (expressed through her voting behavior in Part A) and we find that it has important across- and within-treatment explanatory power. In particular, we find that the distance between the implemented outcome and the closest vote of an individual (i.e. the voted alternative that is closer to the outcome than any other voted alternative) has large across- and within-treatment variation and strongly relates to the degree of preference shift in the second stage of the experiment. Hence, we provide evidence that not only supports the idea that congruent mechanisms move a group members' preferences closer together, but additionally, that one's apparent acceptance of an outcome –even in the context of the same mechanism– plays a part in explaining the change in their own preferences.[5] We extend this analysis to the group-level by defining a compatible measure of apparent consensus of all voters with the outcome, and we find that, unlike its individual-level counterpart, this measure cannot explain the treatment effect. That is, we find that when an individual declares an outcome acceptable, the individual's ideal policy moves closer to it. At the same time, the strategies of the rest of the players do not play a significant role as far as her ex-post preferences are concerned.

This last finding is arguably of independent interest as it points towards a potential novel route through which procedures affect individuals. It has been suggested (Sen, 1997; Frey, Benz, and Stutzer, 2004) that procedures affect individuals' utility: a) directly, through their attitudes about specific procedural features (e.g. whether they are fair, just, democratic, etc.) and b) indirectly, through the way they are treated by others within a procedure. In our study we detect a factor that acts orthogonally to these suggested factors: The incentives provided by the procedure shape individuals' within-procedure behavior, and the within-procedure behavior of an individual is found to have a significant effect on her attitude towards the procedure's outcome.

In what follows we discuss the relevant literature (section 2), provide a discussion regarding the two decision rules (section 3), detail our experimental design (section 4), present our results (section 5) and conclude (section 6).

2 Relevant Literature

There has been a growing interest in the use of laboratory experiments to measure the effects of different collective decision processes on the effectiveness or acceptance of decision outcomes. Walker, Gardner, Herr, and Ostrom (2000) show that voting can increase efficiency through coordination in a common pool resource game. Dal Bó, Foster, and Putterman (2010) illustrate that the effects of a policy on

[5]Our results are robust to alternative measures of apparent consensus between a voter's strategy and the implemented outcome –such as the average distance between one's voted alternatives and the implemented outcome.

cooperation are stronger when it is chosen democratically. A similar effect is found for the performance of sanctioning institutions in public good games that are voted on instead of imposed exogenously (Tyran and Feld, 2006; Sutter, Haigner, and Kocher, 2010; Markussen, Putterman, and Tyran, 2013; Kamei, Putterman, and Tyran, 2015; Kamei, 2016) or chosen by an elected –versus imposed– leader (Grossman and Baldassarri, 2012). In contrast to the above, Markussen, Reuben, and Tyran (2014) find that the effectiveness of a scheme of intragroup competition is not affected by whether or not it is chosen democratically. Beyond social dilemmas, Mellizo, Carpenter, and Matthews (2014) find that democratic processes can lead to higher effort in the workplace when compensation schemes are chosen by voting. As we measure the effect of the group choice process directly on group members' preferences concerning the outcome, our results can help explain this typically positive effect of democratic institutions.

With the exception of Walker et al. (2000), these papers compare exogenous to endogenous choice. By contrast, since we focus on the apparent consensus on an outcome, the processes we compare are all endogenous. Furthermore, our design allows us to experimentally control for differences in the outcomes produced by different mechanisms, giving a clean identification of the effect of a given mechanism on individuals' preferences. Previous work typically controls for such effects only econometrically.

Scholars in both management and psychology have long been interested in the effect of different decision processes on group decisions (see, for instance, Mason and Mitroff, 1981; Schweiger et al., 1986; Schweiger, Sandberg, and Rechner, 1989; Priem et al., 1995; and Hartnett, 2011). Another long stream of literature in organizational studies examines the role of conflict in teams and groups. Two extensive meta-studies (De Dreu and Weingart, 2003; De Wit et al., 2012) find that conflict is in general negatively correlated with group performance, as measured by different metrics. While there is some support for the idea that conflict can be beneficial in specific contexts (Jehn, 1994; Jehn, Chadwick, and Thatcher, 1997; Jehn and Mannix, 2001), this is not universally true for the relationship between conflict and group satisfaction. Our work complements this literature by looking at conflict that is created –or mitigated– by the decision processes used. We do not measure the effect of disagreement on the outcome; in fact we use a design that minimizes any possibility for such an effect. This allows us to obtain estimates of the causal effects of apparent consensus with an outcome on the ex-post levels of actual agreement in a group. To our knowledge, such an incentivized elicitation of ex-post preferences has not been applied in this literature.

Our results can be interpreted as evidence for procedural utility: Processes matter above and beyond their explicitly associated outcomes. The idea has its origins in social psychology (Thibaut and Walker, 1975; Lind and Tyler, 1988), but it has also been advanced by economists (Sen, 1997; Frey et al., 2004; Frey and Stutzer, 2005). Research in this area focuses mainly on moral characterizations of processes, such as whether participants are treated equally or fairly, and the effect the moral characterization of a process may have on outcomes and their acceptance. In our case, we look at processes that, while resulting in divergent levels of apparent consensus, can arguably not be ranked in terms of how fair they are. Hence, the effect of procedures on preferences in our case is independent of procedural justice. Our finding that an individual's agreement with an outcome, as expressed through her vote, moves her ideal policy closer to said outcome seems to be parallel with what Corazzini, Kube, Sebastian, Maréchal, André, and Nicolo (2014) find. Their paper looks at how the process of electing leaders may incentivize them to keep their promises. This of course relates to the literature on lying aversion (Gneezy, 2005; Fischbacher and Föllmi-Heusi, 2013). In our experiment subjects are not aware when voting that they will have an opportunity to propose a new outcome. Hence, their post-outcome behavior seems to reveal a desire for self-consistency or self-concept maintenance (Mazar, Amir, and Ariely, 2008).

To properly choose the decision context and the employed mechanisms, one has to turn to the theoretical literature. We have chosen to focus on the single-peaked domain since: a) it is quite intuitive and easy to explain in the lab, and b) it is the only one, to our knowledge, for which outcome-wise identical congruent and incongruent mechanisms exist. Renault and Trannoy (2005) and Yamamura and Kawasaki (2013) analyze the properties of the Simple Mean mechanism and show that the unique Nash equilibrium outcome under the average voting rule must be equivalent to the median of $(h_1, h_2, \ldots, h_n, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n})$, where h_i is player i's ideal policy. They also prove that in equilibrium most players select an extreme announcement (hence, the Simple Mean mechanism is incongruent). Núñez and Xefteris (2017) show how to implement the same outcome using the Median Approval mechanism that leads players to endorse the implemented alternative (hence, the Median Approval mechanism is congruent). Finally, Gershkov, Moldovanu, and Shi (2017) show how to make the same decision through sequential quota procedures. A study of the comparative effects of simultaneous versus sequential mechanisms is beyond the scope of this analysis, but it presents itself as an interesting avenue of research for the future.

3 Theoretical Framework

A group of N members chooses a point x in a compact policy space $\mathcal{P} \subset \mathbb{R}$. When considering voter preferences we assume that they are single-peaked and we focus on each voter's preferred policy $h_i^t \in \mathcal{P}$. The time superscript $t \in \{0, 1\}$ indicates that preferences are allowed to be different before a collective choice is made ($t = 0$) and after ($t = 1$). What we are interested in is the group's actual agreement level, which can be captured by a measure of dispersion of voters' preferred policies. Let $S : \mathcal{P}^N \to \mathbb{R}$ be such a measure. The lower the dispersion, the higher the agreement. Our main hypothesis is that the post-decision levels of agreement will be different depending on the mechanism used to reach the decision. In particular, mechanisms with higher levels of apparent consensus with an outcome will lead to higher levels of actual agreement. We now formalize the mechanics that lead to this hypothesis.

Before any decision is taken by the group, preferred policies will depend on voters' own material (and single-peaked) payoff function, $u_i : \mathcal{P} \to \mathbb{R}$, and, potentially, on the payoffs of the other voters in the group, u_{-i}. Hence, $h_i^0 = g^0(u_i, u_{-i})$ or, simply, $h_i^0 = g^0(u)$. This formulation is general enough to allow for other-regarding preferences such as welfare maximization or inequity aversion.[6]

The group makes a decision using some mechanism M that asks voters to signal their preferences and produces an outcome based on these messages. We denote a voter's submitted message by b_i, but the exact form of this message depends on the mechanism used. Nevertheless, we assume it is possible to define a individual measure of "apparent consensus with a policy",

$$c_i = c(x, b_i), \tag{1}$$

which captures the compatibility between a voter's message, b_i, and the group outcome, x.

We want to formalize the intuition that post-outcome preferences will depend not only on material payoffs but also on the outcome and voters' behavior in the choice process. A general way to capture this is by writing

$$h_i^1 = g^1(u, x, b). \tag{2}$$

Actually, we propose something more specific. We postulate that what matters is not the outcome and the behaviors in general, but the level of the individuals' apparent consensus with the outcome, captured by the c_is: the higher the level of apparent consensus with the outcome, the closer will the

[6] See Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002) for examples of such preferences.

voters' preferred policy move towards the outcome. Hence we write:

$$h_i^1 = g^1(u, c_i, c_{-i}) \tag{3}$$

and

$$\frac{\partial |x - h_i^1|}{\partial c_j} \leq 0, \ \forall \, j. \tag{4}$$

If the above holds, then our main hypothesis follows. The implication of (4) is that the outcome x becomes an attractor for post-outcome preferences. All voters' ideal policies move closer to x and, hence, closer to each other:

$$\frac{\partial S(h^1)}{\partial c_j} \leq 0, \ \forall \, j. \tag{5}$$

Our experiment tries to test the above theory by allowing groups to make decisions using different mechanisms and eliciting post-outcome preferences. There is, of course, an obvious endogeneity problem we need to overcome. As seen by (1) and (2) or (3), preferences depend on material payoffs u, the outcome x and behavior b. It is straightforward to control for u in the lab and we do so. Nevertheless, in principle, different mechanisms affect x and b simultaneously, not allowing us to identify the effect postulated in (4). One way to overcome this is by finding specific mechanisms that, given u, lead to the same outcome x but with different behavior b. In particular, behavior should differ in such a way that we observe higher levels of apparent consensus with the outcome in one mechanism compared to another. More formally, the two mechanisms M and M' need to satisfy:

$$x(u, M) = x(u, M') \tag{6}$$

and

$$c_i(x(u, M), b_M^*) \geq c_i(x(u, M'), b_{M'}^*) \, , \ \forall \, i. \tag{7}$$

In the next subsection we present two mechanisms that conform with these desiderata and form the basis of our experimental design.

3.1 Two Alternative Mechanisms

With respect to the group decision in the first stage of the game, we focus on two mechanisms: the *Simple Mean* mechanism and the *Median Approval* mechanism. This section reviews their definitions and the equilibrium prediction for the situation tested experimentally. For a formal derivation of these

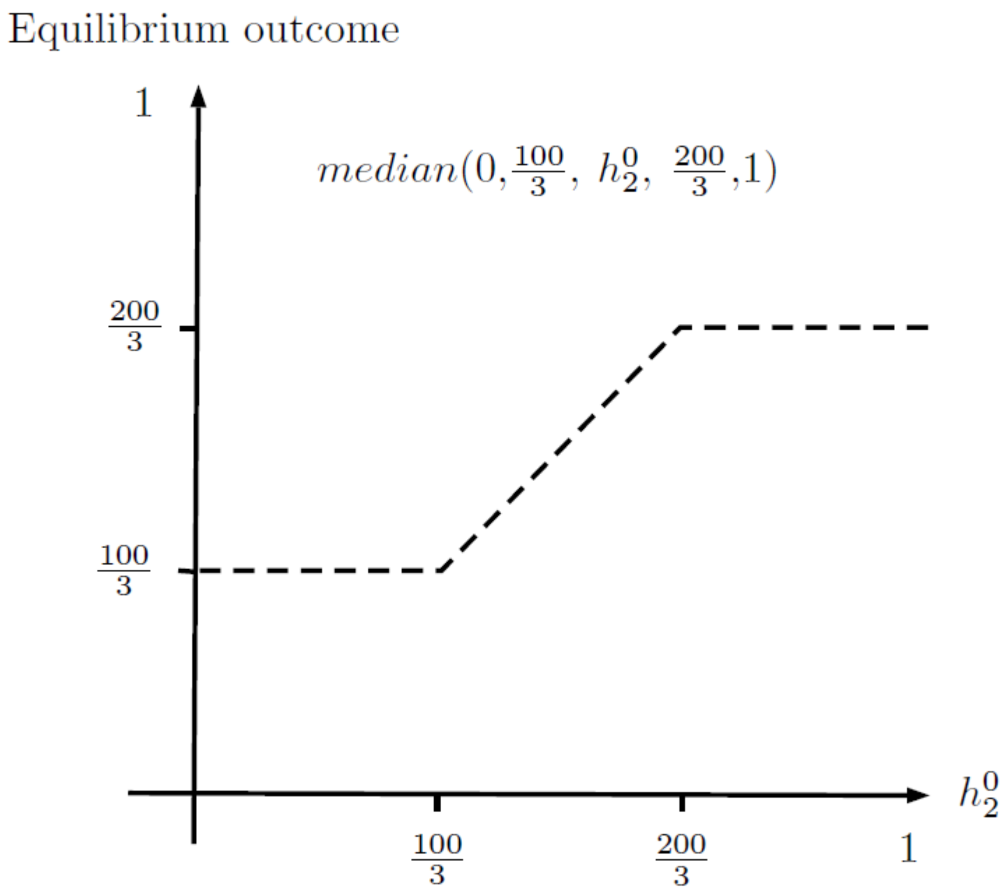


Figure 1: Equilibrium outcome as a function of h_2^0 with $h_1^0 \leq \frac{200}{3}$ and $h_3^0 \geq \frac{100}{3}$.

results, we refer the reader to Renault and Trannoy (2005) and Yamamura and Kawasaki (2013) for the Simple Mean mechanism, and to Núñez and Xefteris (2017) for the Median Approval mechanism.

We consider a committee with three individuals (i.e. $n = 3$); a decision $x \in [0, 100]$ needs to be made. Individual preferences over outcomes are summarized by the utility function $-|x - h_i^0|$. Individual utility is maximized at $x = h_i^0$ (i.e. at the individual's preferred decision). The larger the difference between the decision x and h_i^0, the smaller the individual utility. Our arguments do not depend on the precise shape of the utility functions and extend as long as individual preferences have a unique preferred decision (i.e. single-peaked preferences).

In order to ease the comparison, we focus on the case with $h_1^0 < h_2^0 < h_3^0$ with $h_1^0 \leq \frac{200}{3}$ and $h_3^0 \geq \frac{100}{3}$. In this case, both mechanisms under consideration admit a unique equilibrium outcome and a simple derivation of the equilibrium strategies. This equilibrium outcome, as depicted by Figure 1, is the median of the peaks h_1^0, h_2^0, h_3^0 jointly with $\frac{100}{3}$ and $\frac{200}{3}$.

Simple Mean mechanism: Each player $i \in N$ simultaneously submits a value $b_i \in [0, 100]$. For each vector of announcements $b = (b_1, b_2, b_3)$, the outcome $\theta_{\mathbf{SM}}(b)$ equals:

$$\theta_{\mathbf{SM}}(b) = \frac{b_1 + b_2 + b_3}{3}.$$

Equilibrium Behavior: In equilibrium, each player casts a strategy that minimizes the distance between the outcome $\theta_{\mathbf{SM}}(b)$ and her own peak. Player 1, the player with the lowest peak, always announces 0 since she anticipates that the outcome is higher than her peak and wants to shift the outcome as much as possible to the left. Similarly, Player 3, the player with the highest peak, always announces 100 since she wants to shift the outcome as much as possible to the right.

The strategy of the median player, Player 2, depends on the value of her type h_2^0. If the median type is low ($h_2^0 \leq \frac{100}{3}$), then Player 2 announces 0 and, by symmetry, if the median type is high (i.e. if $h_2^0 \geq \frac{200}{3}$), then Player 2 announces 100. Finally, if the median peak is centered ($\frac{100}{3} \leq h_2^0 \leq \frac{200}{3}$), the median player plays a strategy that allows to obtain h_2^0 as an outcome: This strategy equals $3h_2^0 - 100$.

In any equilibrium, the outcome is equal to $f(h_1^0, h_2^0, h_3^0) = median(h_1^0, \frac{100}{3}, h_2^0, \frac{200}{3}, h_3^0)$. The equilibrium is unique since slightly altering one's announcement affects the final outcome independent of the announcement of the rest of the players (see Proposition 3 in Yamamura and Kawasaki, 2013 for a precise statement of the conditions that lead to a unique equilibrium).

Median Approval mechanism: Each player $i \in N$ simultaneously submits an interval $b_i = [b_i^-, b_i^+]$ with $b_i^- \leq b_i^+$. Player $i \in N$ casts one vote for each alternative included in her chosen interval. Let $\mu(b_i) = b_i^+ - b_i^-$ denote the measure of b_i and, for each set of intervals $b = (b_1, b_2, b_3)$, $\mu(b) = \mu(b_1) + \mu(b_2) + \mu(b_3)$ the measure of b. For each $x \in [0, 100]$ and each set b, $s_x(b) = \#\{i \in N \mid x \in b_i\}$ denotes the score of x at b. Note that if $\mu(b) = 0$, each announcement is a singleton. If $\mu(b) > 0$, the distribution of votes $\phi : \mathcal{B}^n \times [0, 100]$ is denoted by $\phi(b, z) = \frac{1}{\mu(b)} \int_0^z s_x(b)dx$.

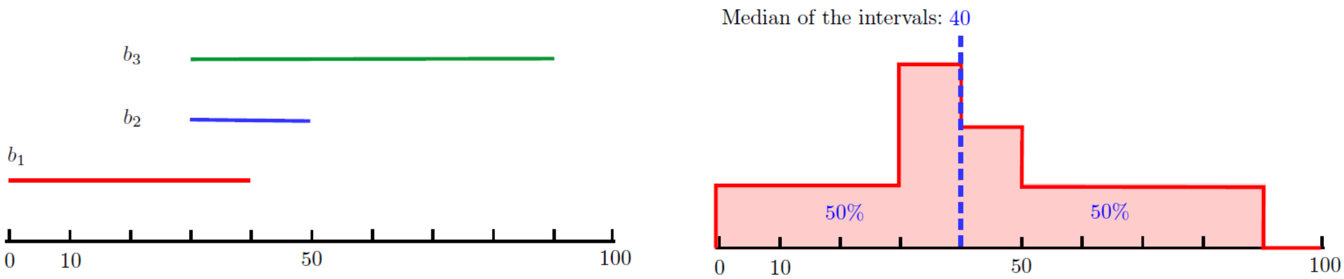
For each vector of announcements $b = (b_1, b_2, b_3)$, the outcome $\theta_{\mathbf{MA}}(b)$ equals:

$$\theta_{\mathbf{MA}}(b) = \begin{cases} median(b_1, b_2, b_3), & \text{if } \mu(b) = 0 \\ \min\{z^* \in [0, 100] \mid \phi(b, z^*) = \frac{1}{2}\}, & \text{otherwise} \end{cases}$$

Figure 2 depicts the computation of the median of the announced intervals. After plotting the intervals (Figure 2b), we plot the vote distribution (Figure 2c) –that is, the number of votes that each alternative gets by the players. The median of the intervals coincides with the point that divides the area below the vote distribution into two equal parts.

Announcements	b_i^-	b_i^+
Individual 1	0	40
Individual 2	30	50
Individual 3	30	90

(a) Individuals report intervals.



(b) Graphic representation of the intervals.

(c) Distribution generated by the intervals.

Figure 2: Computing the median of the intervals.

Equilibrium Behavior. In a similar fashion to the Simple Mean mechanism, each player chooses a strategy that minimizes the distance between the outcome and her own peak. The unique equilibrium outcome is also equal to $f(h_1^0, h_2^0, h_3^0) = median(h_1^0, \frac{100}{3}, h_2^0, \frac{200}{3}, h_3^0)$.

Player 1 announces an interval b_1 that ranges from 0 to $f(h_1^0, h_2^0, h_3^0)$. Namely, she votes for the outcome $f(h_1^0, h_2^0, h_3^0)$ and for any alternative located to its left. By symmetry, Player 3 approves the interval b_3 that goes from $f(h_1^0, h_2^0, h_3^0)$ until 100, voting for that outcome and all the alternatives located to its right.

The median player, Player 2, plays a strategy that depends on the value of h_2^0. When $h_2^0 < \frac{100}{3}$ ($h_2^0 > \frac{100}{3}$) then she votes for $b_2 = [0, \frac{100}{3}]$ ($b_2 = [\frac{200}{3}, 0]$), and the outcome is equal to $\theta_{\mathbf{MA}} = \frac{100}{3}$ ($\theta_{\mathbf{MA}} = \frac{200}{3}$). When $\frac{100}{3} \leq h_2^0 \leq 50$ ($50 \leq h_2^0 \leq \frac{200}{3}$), she can vote any alternative from 0 to $4h_2^0 - 100$ (from $4h_2^0 - 200$ to 100) –that is, also h_2^0– inducing the implementation of her ideal policy.

To better understand the equilibrium behavior, consider an example with $h_1^0 < h_2^0 = 40 < h_3^0$ and the strategy profile $b_1 = [0, 40]$, $b_2 = [0, 60]$ and $b_3 = [40, 100]$. These strategies lead to the implementation of alternative 40 –which is voted by all players– since in total 40+60+60=160 units of votes are cast, and half of them are given to alternatives to the left (right) of 40. To see why this is an equilibrium consider deviations of the first player. If, for example, she expands her interval to $b_1' = [0, 46]$, then the total votes cast will be 46+60+60=166, and the implemented alternative will thus move from 40 to 41. Since $h_1^0 < h_2^0 = 40$, this is not a profitable change for player 1. If she shrinks her interval to $b_1'' = [0, 20]$, then the total votes cast will be 20+60+60=140, and the implemented alternative will thus move from 40 to 45. Since $h_1^0 < h_2^0 = 40$ again this is not a profitable deviation. Of course, these deviations are just indicative of what may happen: Players have a variety of different options to choose from. These few cases though are sufficient to show that both by voting for alternatives to the right of the implemented outcome and by not voting for alternatives to the left of the implemented alternative, a player can shift the outcome to the right. Hence, if a players's ideal policy is to the left, her only best response is to vote for the outcome and all alternatives to its left.

3.2 A Mechanism for Second-Stage Decisions

For second-stage decision making we employ the random dictatorship mechanism. According to this mechanism each individual proposes an outcome and each of these proposals is selected as the second-stage group decision with probability one third. Notice that this mechanism is strategy-proof and admits a unique equilibrium. In this unique equilibrium each individual i reports truthfully her preferred decision h_i^1.

4 Experimental Design and hypothesis

4.1 Design

The experimental design is geared towards answering our main question on the influence of apparent consensus with an outcome on actual agreement. For this, we use a between-subject design with two treatments, each of which has two parts. In the first part, subjects make collective decisions using one of the two mechanisms described in the previous section, a different one in each treatment. The decision rules are such that theory predicts the same outcome with different levels of apparent consensus. Payoffs in the first part depend on this outcome. In the second part subjects make individual proposals, one of which is chosen randomly for each group. Payoffs in this part are determined by the randomly selected proposal in each group. Given no difference between treatments in the outcomes of the first part, a treatment effect in the second part can be interpreted as an effect of apparent consensus with the outcome on post-outcome individual preferences. We now explain the details of the experiment and our design choices.

The experiment took place at the University of Cyprus Lab of Experimental Economics (UCY LExEcon). A total of 120 subjects, all students of the University of Cyprus, participated in eight equally sized sessions, with four sessions per treatment.[7] Recruitment was done using ORSEE (Greiner, 2015). The experiment was computerized, and the software was programmed and run using zTree (Fischbacher, 2007). An outline of the design is presented in Table 1.

TABLE 1: THE TWO EXPERIMENTAL TREATMENTS

Treatment	Part A (20 periods)	Part B (20 periods)	N	# of Sessions	Subjects per session	Group size
MA	Interval voting with median as outcome	Random dictator game	60	4	15	3
SM	Single vote with mean as outcome	Random dictator game	60	4	15	3

Timing For both treatments, subjects received written instructions for part A after entering the lab.[8] These were also read aloud to establish common knowledge. After part A finished, instructions for part B were distributed and read aloud. At the end of part B, subjects were informed about their profits and paid privately before leaving the lab. Thus, while in part A subjects knew there will be a second part to the experiment, they were entirely unaware of the experimental task in part B or if and how it was connected to the task in part A.

Collective choice - Part A In each round of part A, subjects are placed in groups of three. Each group needs to choose collectively an integer between 1 and 100 as the group's *destination*. Each group member has an individual *starting point*, that is, a different integer between 1 and 100. The payoff in each period is then 100 points minus the distance between the *destination* and the subject's *starting*

[7] Two pilot sessions were completed before the main experiment to finalize the design, fine-tune some of the parameters, and receive feedback on the instructions. Data from these pilot sessions are not included in any of our analysis. A first version of this paper was circulated with data from the first six sessions. Following suggestions by referees we conducted two additional sessions using the same parameters to enlarge our dataset. There was no qualitative change in our results.

[8] A translation of the instructions, originally in Greek, can be found in the Appendix.

point. *Starting points* are common knowledge and are different for every subject in each period.[9] Groups are reshuffled in each period, and subjects do not know the identity of the other group members.

The only parameters that differed between subjects and rounds were the subjects' starting points, which determine their payoffs. Nevertheless, the exact same set of parameters was used across all eight sessions. That is, for any combination of starting points used for a group in a specific round of a session, there was another group in all other sessions with the same starting points in the same round. Furthermore, the exact same sequence of parameters was assigned to subjects in all sessions.

Treatment - Median Approval (*MA*) In treatment *MA*, to choose the *destination*, each group member chooses an interval of integers between 1 and 100, and each location in the interval receives a single vote. The collective choice is the maximum median of the distribution of votes. Subjects choose the interval by moving specific bars on their screen that mark the lower and higher limits of the interval of votes.

Treatment - Simple Mean (*SM*) In treatment *SM* subjects can vote for a single location by choosing an integer between 1 and 100. The collective choice is the mean of all three votes. Voting takes place by moving a bar to the specific location that the subjects wishes to vote for.

Voting, information and time limit Voting in both treatments lasts for $60+x$ seconds,[10] where x is a number between 1 and 10, chosen randomly in each round and not known to the subjects. During this time, each subject is informed about her and others' starting points and can enter her votes (interval or single vote, depending on the treatment). She can also observe the votes entered by other group members in real time. At any given point in time the software calculates the *destination* and the group members' payoffs. These are shown on the screen as a clock counts down from one minute. At 10 seconds, a text starts blinking indicating that time is almost up, after which it turns red and indicates that voting may finish at any moment. The *destination* for the period is determined by the votes entered when the $60 + x$ seconds finish. After that, a screen appears informing subjects about the results of the voting: each subject's votes, the final *destination*, and subjects' payoffs.

[9] We chose starting points in a way that maximized power for the experiment in terms of detecting a treatment effect in the second part. To that end, we ran simulations using many different sets of starting points and hypothesizing a treatment effect of magnitude and variance similar to what we found in the pilot session. We then chose the set of starting points where the effect was stronger in a linear regression similar to the one corresponding to the first column of Table 4. More details on the exact process are available upon request.

[10] For the first two periods this is extended to $90 + x$ to allow subjects to get familiarized with the voting environment.

Random Dictator Game - Part B Part B is identical for both treatments. In each round of part B, subjects are placed in the same groups as in the corresponding round of part A. Again, the group needs to choose a collective *destination* from the same *starting points* as in the respective round of part A. Unlike part A, the choice is now made by a random dictator: Each group member proposes a location by choosing an integer between 1 and 100. One of the three proposals is chosen randomly as the group's destination for the round, and payoffs are determined in the same manner as in part A. Before making a proposal, subjects are reminded of all *starting points*, all votes, and the chosen *destination* in the corresponding round of part A. They make a proposal by clicking on a location on the screen and then on a 'submit' button. They can revise their proposal as many times as they wish before clicking 'submit'. For each location they click, the software calculates and shows all players' payoffs if that proposal is selected. They cannot see the proposals of the other group members, and they are not informed about the others' proposals and the final outcome until the end of all rounds.

Matching protocol As mentioned, subjects are put in a new group in each of the 20 rounds in Part A and this matching is repeated in Part B. We reshuffle the groups to avoid repeated game effects. While subjects may meet another subject again, they cannot know when this happens. Furthermore, given that reshuffling happens among all session participants, the probability of playing with the same players in consecutive rounds is small. Despite this one could still worry that behavior across groups within each session might not be independent.[11] We deal with the issue in two ways: through design features that we expect to minimize such possibility, and through appropriate care in the analysis of results. We explain the former here and differ discussion of the latter for the results section.

In random matching protocols the behavior of players in one group can affect their behavior in the other groups they participate in at later rounds. The effect is typically expected to be through the belief formation process. For example, in a public good experiment, if one's peers did not contribute in one group, then the subject might expect his peers in the next group to do the same. The real time voting process used in Part A helps avoiding this problem. Subjects do not need to rely on their experience in previous rounds to guess what their peers may vote for, as they can see it happening and react to it. In Part B, subjects do not receive any feedback after making their proposals until the end of the experiment. Thus, while of course others' voting behavior in part A may affect one's proposals in part B, there cannot be any direct dependency between different subjects' proposals.

[11] Logistic considerations did not allow the use of smaller within-session cohorts that would mitigate the problem.

Payments After part B is completed a message on the screen informs subjects about the outcomes of all rounds in both parts. One round from each part is chosen randomly, and payoffs in that round are used to determine the subjects' payment for the experiment. Subjects receive €1 for every 15 points earned in the selected round of each part, plus an additional €3 as a participation fee. Subjects earned €13.21 on average across all sessions.

4.2 Hypotheses

The experiment is designed in a way that allows us to test the hypotheses that can be derived from our theory. We state these hypotheses here as a guide to aid the reader through the result section.

In Part A groups make choices using a different mechanism in each treatment. As explained in section 3, the *MA* and *SM* mechanisms were chosen specifically to overcome the endogeneity of outcomes. Of course, the arguments given in support of choosing the particular mechanisms were purely theoretical and solely based on equilibrium predictions. It remains an empirical question whether subjects' behavior in the two mechanisms satisfies the desiderata stipulated in (6) and (7). We therefore formulate the following two hypotheses, which are auxiliary to our main research question, but nevertheless of interest in their own right.

Auxiliary Hypothesis A. The *MA* and *SM* mechanisms produce the same group outcomes:

$$x(u, MA) = x(u, SM).$$

Auxiliary Hypothesis B. Individuals' apparent consensus with the outcome is higher in the *MA* treatment compared to *SM* (or *MA* is more congruent than *SM*):

$$c(x(u, MA), b_{MA}) > c(x(u, SM), b_{SM}).$$

If our theoretical predictions about behavior in Part A play out, we can turn to subjects' behavior in Part B. Using the data from the second part it is possible to test our main hypothesis, which compares levels of actual agreement following decisions made by each mechanism:

Hypothesis 1. Post-outcome levels of actual agreement are higher in the *MA* treatment compared to *SM*:

$$S_{MA} < S_{SM}.$$

Assuming our main hypothesis holds true, we wish to understand the mechanism that drives it and whether it conforms to our theory concerning the role of apparent consensus and the function of the outcome as an attractor for post-outcome preferences. This "attractor effect" and its potential difference across treatments is captured by the following hypotheses to be tested:

Hypothesis 2. Subjects' proposals in Part B are affected by the group outcome:

$$\frac{\partial g^1(u, x, b_{MA})}{\partial x} \neq 0.$$

Hypothesis 3. Subjects proposals in Part B are affected by the group outcome more in the *MA* treatment than in *SM*:

$$\left|\frac{\partial g^1(u, x, b_{MA})}{\partial x}\right| \geq \left|\frac{\partial g^1(u, x, b_{SM})}{\partial x}\right|.$$

Finally, the role of apparent consensus in generating the "attractor effect" is formalized as follows:

Hypothesis 4. Subjects proposals in Part B move closer to the group outcome when individual apparent consensus is high:

$$\frac{\partial |x - h_i|}{\partial c_j} \leq 0, \ \forall\, j.$$

5 Results

5.1 Part A

5.1.1 Voting process

We use a voting process with real-time feedback to allow for fast within-round learning.[12] A random ending point is used to discourage extreme "sniping" behavior, which was observed in a pilot session with fixed ending points.[13] One difference between the voting mechanisms that is worth noting concerns the

[12]Moreover, it has been shown that feedback exchange among players prior to the group decision point helps diminish outcome-related institutional differences (see, for instance, Goeree and Yariv, 2011 and Gerardi and Yariv, 2007), which is desirable in our case.

[13]With a fixed end point, many subjects would significantly change their votes in the last seconds of voting in an effort to achieve a more favorable outcome. The term "sniping" has been used in online auctions to describe bidders that only

degree to which a single voter can affect the outcome, given the others' votes. In MA the outcome can, theoretically, move up to 98 points by a single individual's change in votes, but only for very particular choices of the rest of the players.[14] In most scenarios, a single voter's power over the outcome is quite limited: When two voters approve of many alternatives, the median of the induced vote distribution is moderately responsive to a change in the strategy of the third voter. On the other hand, in SM, it is always possible for a single voter to move the outcome to any point within a range of 33 points. Moreover, in SM, it is also practically easier to move the outcome since an individual's vote can change to any other with one direct move, while in MA a change to a different strategy involves a two-step process: A subject needs to change one end of the interval before making a change to the other.

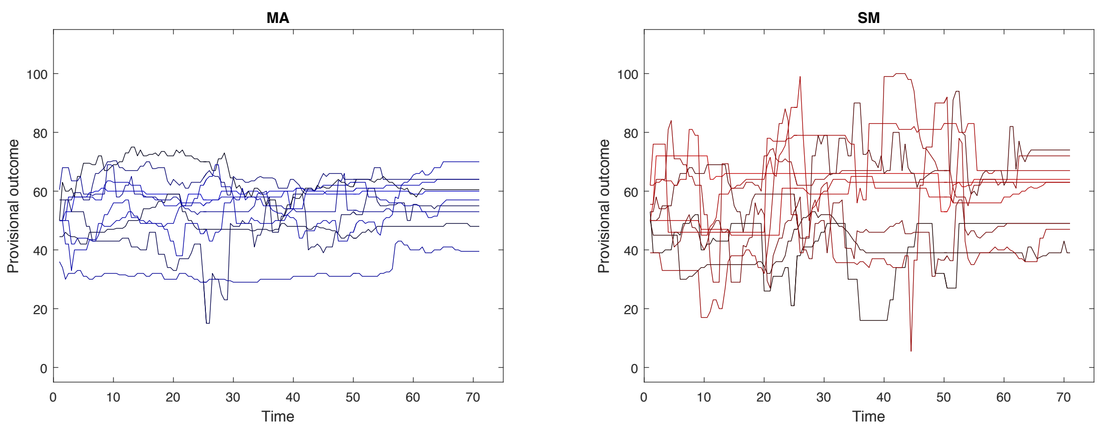


Figure 3: Volatility of collective choices. The left panel shows 10 randomly chosen groups and how the provisional outcome changes across time during the voting process in the MA treatment. The right panel shows the same for 10 groups with the same starting points in the SM treatment.

In the two panels of Figure 3, we show the provisional outcomes across time for 10 randomly chosen groups in each treatment. A common pattern emerges for all groups. In MA, movements are more gradual. Substantial movements happen mostly in the first 30 seconds. In SM, there is more volatility throughout the round. There are often substantial moves of the provisional outcome between the 40th and 60th second. After that movements are rare and small in magnitude. The difference in volatility reflects the preceding discussion. The increase in volatility towards the end of the round in SM could reflect some residual "sniping" attempts.

submit a bid in the last moment to avoid driving up the price through a bidding war. See, for example, Ockenfels and Roth (2006).

[14] This happens in the extreme scenario where all voters cast a single vote on location 1. Then one of them can switch and vote the interval [99,100], moving the outcome from 1 to 99.

5.1.2 Votes and Outcomes

Figure 4 gives an impression of individual voting behavior in each treatment, as captured by subjects' votes in the end of each voting process. We provide these here for the sake of completeness. A detailed analysis of voting behavior in Part A is beyond the scope of the paper. Nevertheless, we do note a qualitative conformity of these results with the Nash equilibrium predictions about voters' behavior discussed in subsection 3.1.

What is more important for our analysis is to verify whether our Auxiliary Hypothesis A holds and the two mechanisms produce the same outcomes, as predicted by Nash equilibrium. We do this here.

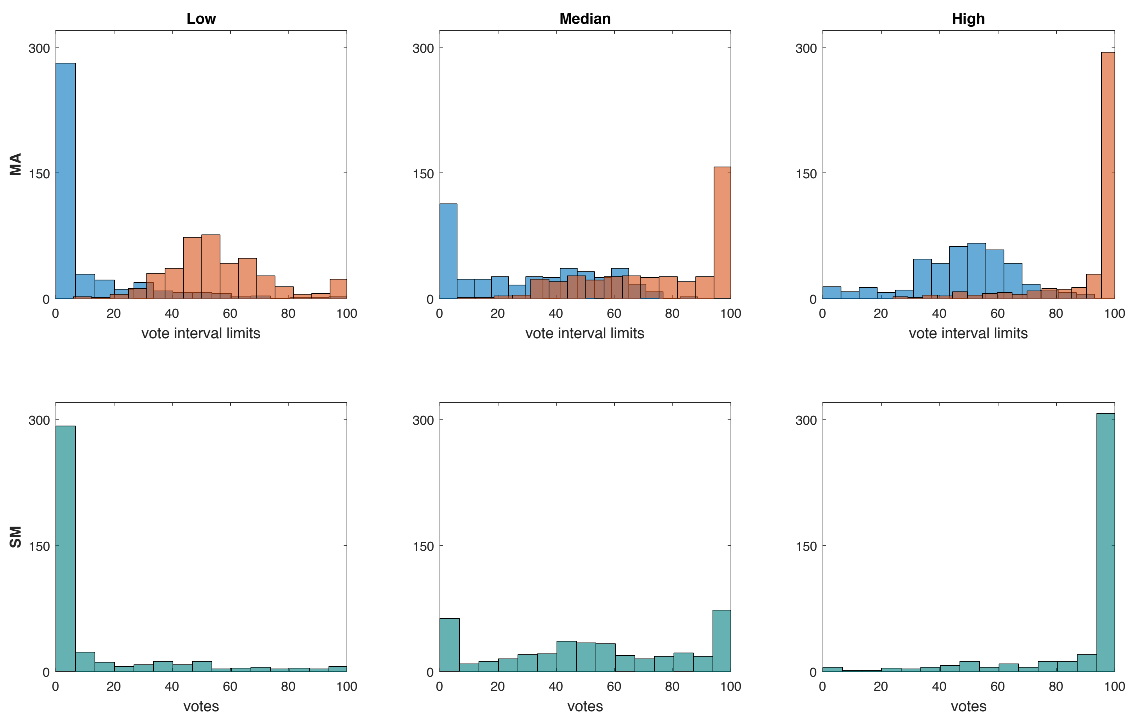


Figure 4: Voting according to type. The first row of graphs shows the distribution of vote interval limits for voters with the lowest, median and highest starting point within each group in the *MA* treatment. The blue and orange histrograms correspond to the left and right limits respectively. The graphs in the second row show the distribution of votes in the *SM* treatment according to each type of voter.

The left panel of Figure 5 shows the collective choices of all groups, in all rounds, for both treatments. As can be seen from the graph the Nash equilibrium does a relatively good job of predicting the outcomes in both treatments. Starting points in all groups are chosen so that the equilibrium outcome is constant and equal to 33 ($\approx \frac{100}{3}$) when the median starting point in a group is below 33. For values higher than that, but lower than 67 ($\approx \frac{200}{3}$), the equilibrium outcome coincides with the median starting point. The equilibrium is again constant and equal to 67 when the median exceeds that value. As we observe, the

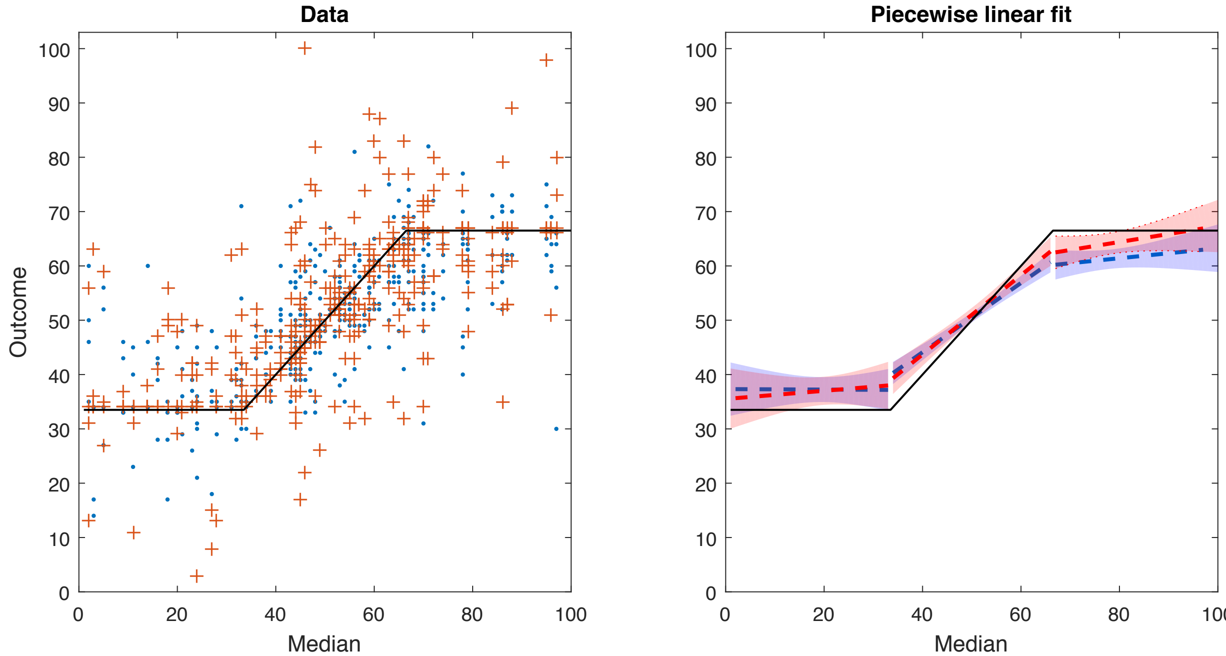


Figure 5: Collective choices for all groups in all rounds. Median refers to the group members' median starting point. The solid line in both panels corresponds to the Nash equilibrium outcome. For our choice of parameters this equilibrium depends entirely on the position of the median. In the left panel, data from the *MA* and *SM* treatments are indicated by dots and crosses respectively. In the right panel, dashed lines correspond to linear fits to the data with respect to the three Nash equilibrium regions, for the *MA* and *SM* treatments in blue and red, respectively. The corresponding shaded areas indicate 95% confidence intervals.

collective choice tends to be very close to the median when it lies between 33 and 67. When the median is below or above this interval the dependency disappears, and the collective choice hovers around 33 and 67.

The above is supported by the piecewise linear fits shown in the right panel of Figure 5. From this graph we also note that the outcomes in both treatments tend to be closer to the center of the range compared to the predicted Nash outcome. Most importantly, though, for our research question, we do not observe any systematic differences in the outcomes across the two treatments.

We further explore this issue by comparing the outcomes across treatments in three dimensions: their location, their efficiency, and their degree of inequality. Any of these dimensions could affect how an individual evaluates the collective choice. Table 2 summarizes the outcomes in each treatment across these dimensions.

TABLE 2: SUMMARY STATISTICS OF COLLECTIVE CHOICES IN PART A FOR BOTH TREATMENTS.

Treatment	average absolute deviation from Nash[a]	average deviation from Nash to center[b]	average efficiency (% of max)[c]	average min to max ratio[d]
MA (N = 400)	6.17 (std 6.18)	3.45 (std 8.03)	95.4 (std 4.90)	67.7% (std 13.8)
SM (N = 400)	6.43 (std 8.05)	2.41 (std 10.03)	95.2 (std 5.67)	65.3% (std 14.7)
statistical difference (p-value)[e]	0.580	0.118	0.544	0.003

There are 400 observations in each treatment. Each observation k refers to a group making a collective choice in a specific round.

[a] The absolute deviation from Nash is calculated as: $|outcome_k - Nash_k|$.

[b] The deviation from Nash to center is calculated as: $(outcome_k - Nash_k) \times sign(50.5 - Nash_k)$.

[c] Efficiency is calculated as: $\sum_i pay_{ik}(outcome_k)/\sum_i pay_{ik}(med_k)$, where $pay_{ik}(x) = 100 - |start_{ik} - x|$ and $med_k = median(start_{ik})$.

[d] The min to max ratio is calculated as: $\frac{\min\{pay_{ik}\}}{\max\{pay_{ik}\}}$.

[e] The p-values are obtained by regressing each variable on a treatment dummy and calculating statistical significance using the wild cluster bootstrap with errors clustered on the *subject* and *session* level (see Cameron, Gelbach, and Miller, 2008, MacKinnon and Webb, 2018 and Roodman, MacKinnon, Nielsen, and Webb, 2018).

Location As one can see from Figure 5, outcomes in the two treatments seem to lie close to each other in a statistical sense. This is further supported by comparing the distribution of absolute deviations from Nash and the distribution of deviations from Nash to the center (see Table 2, columns 2 and 3). In both cases the means are very close and their differences are not significantly different.[15] We conclude that the outcomes in the two treatments do not differ substantially in terms of location.

Efficiency In our setup, maximum efficiency is achieved when the collective choice coincides with the median starting point. Since collective choices are close to the Nash equilibrium outcome, which in turn coincides with the median for a broad range of observations, it is not surprising that high levels of efficiency are achieved in both treatments. Outcomes are slightly more efficient in the *MA* treatment,

[15] Throughout the text we refer to differences as being statistically significant at the 1% level. We also report the p-value for the corresponding test.

but not significantly so.

Inequality We measure inequality in the payoffs associated with the outcome of a collective choice by taking the ratio of the lowest to highest payoff within each group. The lower the ratio, the highest the inequality in payments. In our setup, this measure is maximized when the collective choice coincides with the midpoint between the two more extreme group members. This point is also the maxmin choice: It maximizes the lowest payoff achieved by any group member. Outcomes in the *MA* treatment are slightly less unequal than in *SM*. The difference is small in magnitude but statistically significant. It seems to be the consequence of outcomes being noisier in the *SM* treatment. This finding is robust to measuring inequality in different ways, such as calculating the Gini index or simply looking at the minimum payoff. We conclude that there are some small differences between treatments in terms of inequality. Nevertheless, the differences are small in magnitude and, as we show in the next section, they are not able to account for the large treatment effect we find in part B.

We summarize these findings in the following result:

Result A: *We find substantial support for our Auxiliary Hypothesis A. The outcomes do not differ significantly across treatments in terms of location or efficiency. We only find minor differences in terms of inequality.*

5.1.3 Agreement and Mechanism Congruity

Our premise for the choice of experimental design is that the *MA* mechanism is more congruent than the *SM* mechanism. We now quantify this claim and show how this difference is reflected in the data.

In section 2 we assumed that it is possible to measure the *apparent consensus with an outcome* of an individual, given her behavior in the collective choice mechanism (see equation (1)). Here we construct such a measure of apparent consensus. At this point we simply want to verify whether our Auxiliary Hypothesis B holds. Nevertheless, we also use this measure at a later stage to test some of our other hypotheses.

Given the differences of the two mechanisms, we need to select a measure of consensus that is applicable both to interval and single votes and that can provide a smooth estimate of outcome endorsement by each individual. We do that by taking the distance between the subject's vote that is closest to the outcome and the group's outcome and subtracting it from 100. Formally, let $V(b_i)$ be the set of all

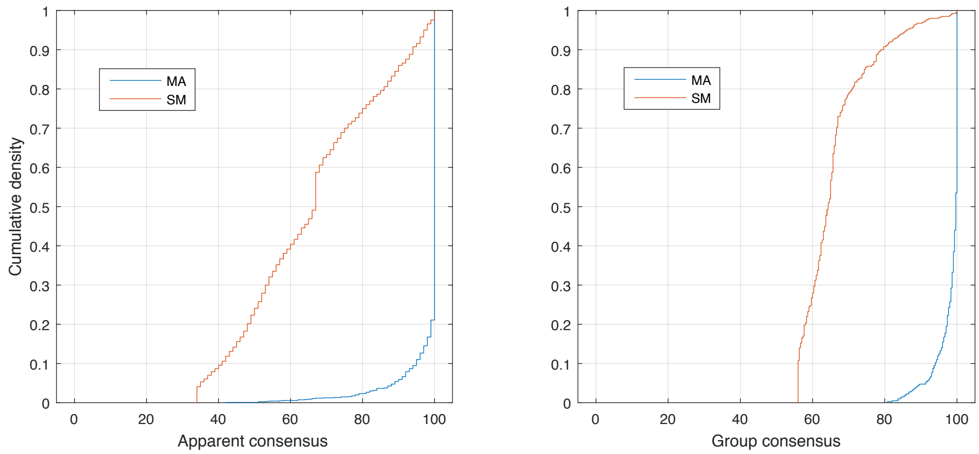


Figure 6: Measures of apparent consensus. The left panel shows the empirical cdf for *individual agreement* in each treatment. The right panel shows the emprical CDF for *group agreement* in each treatment.

locations that subject i votes for (in *SM* this set is a singleton). Then,

$$apparent\ consensus\ of\ i = c(x, b_i) = 100 - \min\{|v - outcome|\ s.t.\ v \in V(b_i)\}$$

This gives an individual measure of apparent consensus with a range from 0 (complete disagreement with the outcome) to 100 (the outcome is a location the individual voted for). Averaging this measure across members of the same group gives the level of *group apparent consensus*, which is also of interest for our analysis. Figure 6 shows the empirical CDFs for these measures for each treatment. It is clear that treatment *MA* displays substantially higher levels of both individual and group apparent consensus.[16] In particular, about 80% of individual strategies in *MA* exhibit the highest degree of outcome endorsement, while in nearly half of the cases, the outcome is unanimously approved by all group members. Finally, there is a stochastic dominance relationship in the apparent consensus levels between treatments –both at the individual and the group level– which indicates that subjects vote for alternatives closer to the outcome in *MA* compared to *SM*, even if one focuses on cases where the outcome is not fully endorsed.

All of the above observations support the following result:

Result B: *We find substantial support for our Auxiliary Hypothesis B. There are higher levels of apparent consensus at both the individual and group levels in treatment MA compared to SM. We conclude that MA is a substantially more congruent mechanism than SM.*

[16] Alternative measures can be constructed using, for instance, the average instead of minimum distance of votes from the outcome. Our results do not change qualitatively. Still, measures based on the minimum distance are stronger predictors of proposals in part B both across and within treatments (see section 5.2).

5.2 Part B

Recall from the previous section that in each round in part B, subjects are put in the same group as in the corresponding round in part A. They are shown the group's votes and outcomes in part A and are assigned the same starting points. They are asked to propose a new destination for part B. Proposals are not restricted in any way and can be any point between 1 and 100. For each group, one of the proposals is selected randomly to determine payoffs. Still, others' proposals, the randomly selected proposal, and the corresponding payoffs are not shown to subjects until the end of the experiment. Hence, for each individual, we have 20 proposals from different groups, which gives 1200 observations per treatment. We interpret individual proposals as a subject's post-outcome ideal policy.

5.2.1 Actual Agreement

We first look at the levels of actual agreement achieved in each group in the two treatments. Our main hypothesis is that groups in the *MA* treatment achieve higher levels of post-outcome actual agreement. In Table 3 we present three different measures of within-group post-outcome dispersion of ideal policies. High levels of dispersion reflect high levels of actual disagreement within a group. We find that for all three measures, post-outcome disagreement is substantially lower in the *MA* treatment and the difference is highly significant.

While we do not elicit pre-outcome preferences, one can use subjects' starting points as a proxy to gain some insight regarding how far agreement has moved due to the collective choice process.[17] Figure 7 shows the empirical cumulative distribution of a normalized measure of disagreement, where we use standard deviation as a measure of dispersion and divide it with the standard deviation of within-group starting points. Normalized disagreement in *MA* is clearly lower. The average normalized disagreement is 0.80 for *MA* and 0.98 for *SM*. We also run a regression of post-decision disagreement $S(h^1)$ (within-group dispersion) on the proxy for pre-outcome disagreement $S(h^0)$ (starting point dispersion) and the interaction of the latter with a treatment dummy (SM=1). We find the following estimates when we

[17] Actual pre-outcome ideal policies may not coincide with each subject's starting point, as social preferences such as inequity aversion may play a role. Nevertheless, as becomes clear in the subsequent analysis, subjects' ideal policy is determined to a large degree by their starting points.

TABLE 3: SUMMARY STATISTICS OF DISAGREEMENT IN PART B FOR BOTH TREATMENTS.

Treatment	Standard deviation within group	Mean abs. deviation within group	Range within group
MA (N = 400)	18.84 (std 9.00)	16.99 (std 8.26)	43.84 (std 20.63)
SM (N = 400)	23.35 (std 10.06)	21.03 (std 9.18)	54.37 (std 23.30)
treatment effect (p-value)*	0.002	0.002	0.002

* The p-values are obtained by regressing each variable on a treatment dummy and calculating statistical significance using the wild cluster bootstrap with robust errors clustered at the *session* level. Very similar and slightly lower p-values are obtained for the treatment effect if one controls for within-group dispersion of Starting points.

use the standard deviation measure:[18]

$$S(h^1) = 2.17 + 0.67 \cdot S(h^0) + 0.19 \cdot (treatment \times S(h^0))$$

These estimates indicate that disagreement in *SM* is determined by 86% from the disagreement in starting points. In the *MA* treatment this effect of pre-outcome disagreement falls down to 67%.

We summarize all the above in the following result:

Result 1: *We find substantial support for Hypothesis 1. Post-outcome levels of actual agreement are significantly higher in the* MA *treatment compared to* SM.

Having established the existence of a treatment effect in the direction we hypothesized, we turn our attention to individual behavior in order to understand the mechanics driving this effect.

[18]Coefficients are all highly significant. We calculate p-values using the wild cluster bootstrap with robust errors clustered on the *session* level. The coefficients we estimate if we use mean absolute deviation, or range, as measures of dispersion are almost identical with the ones for standard deviation.

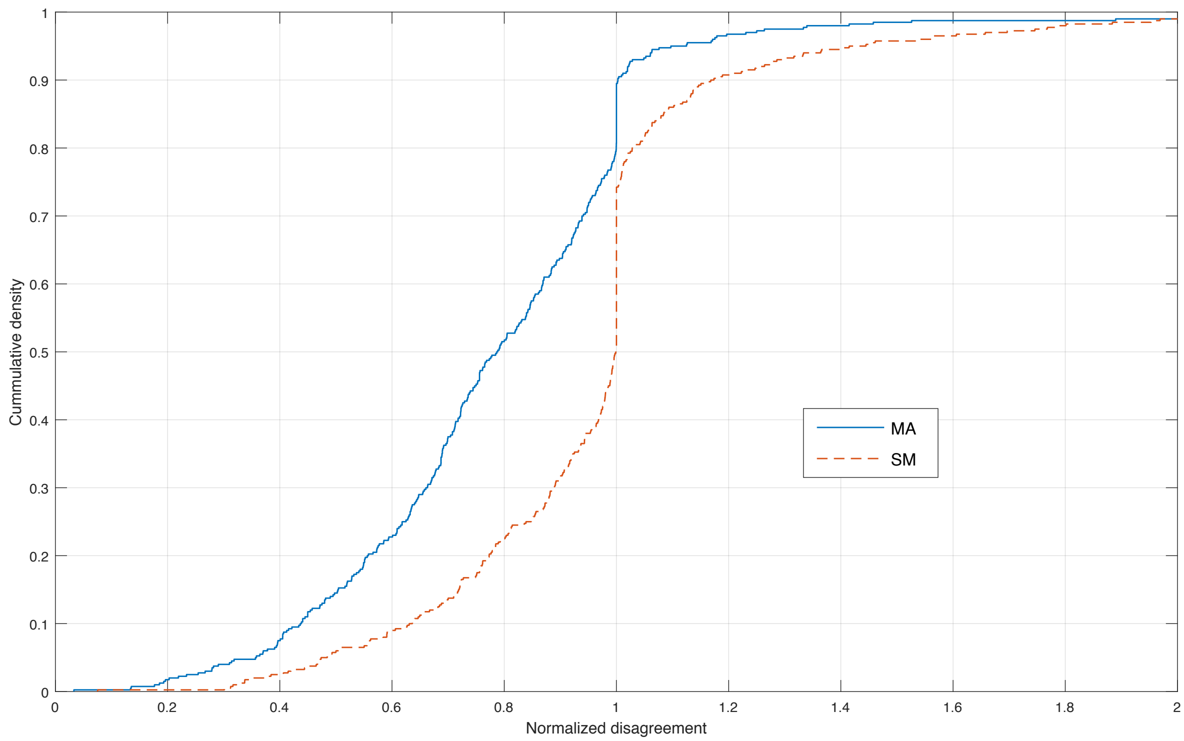


Figure 7: Normalized disagreement. The graph shows the empirical cumulative density functions for the normalized disagreement levels in Part B of each treatment. The measure is calculated by computing the standard deviation of group members' proposals and normalizing by dividing with the standard deviation of the group's starting points. For expositional purposes the graph is truncated on the right, excluding some extreme values (the maximum values are 7.2 for MA and 6 for SM).

5.2.2 The outcome as an attractor

To summarize the data on individual post-outcome ideal policies, we compute the deviation of the proposal from the corresponding part A result for each subject and each round. This is normalized to 0 when these coincide and 1 when the proposal coincides with the subject's corresponding starting point, which is also the payoff maximizing choice (formally, our deviation measure is given by $\frac{proposal - part\ A\ result}{starting\ point - part\ A\ result}$). These deviations are presented for each treatment in the left panel of Figure 8. Values larger than 1 indicate deviations away from the part A result to the direction of one's starting point, which lie even farther away from the part A result compared to the subject's starting point; and values smaller than zero indicate deviations away from the part A result to the opposite direction of the subject's starting point. Naturally, the majority of proposals (83.8% in across both treatments) take values between 0 and 1.

The large jumps in the empirical cumulative density functions at 1 indicate that a large fraction of proposals in both treatments (52.7% in total) coincide with the individuals' starting points. Nevertheless, a substantial number of proposals do not. In particular, 31.1% proposals have a deviation smaller than 1 but greater than or equal to 0. This suggests that the outcome in part A affects subjects' proposals in part B. In fact, while small, there is a noticeable jump in the cumulative distribution of proposal deviations at 0 in both treatments. It is also worth noting that the above observations cannot be attributed to a subset of subjects consistently proposing their own starting points. Heterogeneity in behavior is, of course, present. Still, only 1 out of 90 subjects proposed his/her starting point in all rounds of part B, while all subjects proposed their own starting point at least once.

The above lends support to the following result:

Result 2: *We find substantial support for Hypothesis 2. The group outcome in Part A functions as an attractor for individual post-outcome ideal policies*

We also observe differences in the distribution of proposals across treatments. In particular, 466 (38.83%) proposals equal the individuals' starting points in treatment *MA*, while this number goes up to 689 (57.42%) in treatment *SM*.

Outcomes in Part A are most of the time close to each group's median starting point and farther away from the starting points of non-median voters. It therefore makes sense to take a closer look at potential differences across types of voters within and across treatments. The middle and right panels of Figure 8 break down the data with respect to voters' types (median vs. non-median). A visual

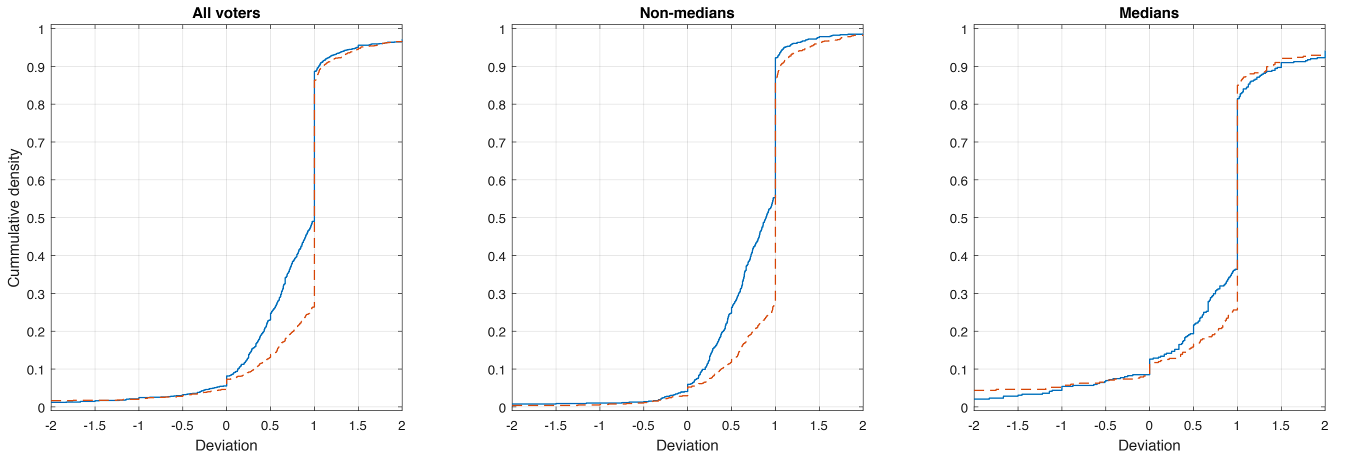


Figure 8: Subjects' part B proposals. The graphs represent proposals as deviations from the corresponding part A outcome. These are normalized to 0 when the proposal coincides with the part A outcome and 1 when the proposal coincides with the subject's corresponding starting point. The left panel shows the empirical cumulative density of the deviations for each treatment. The middle and right panel break down the data by type of voter: non-medians and medians. Solid blue lines correspond to the *MA* treatment and dashed red lines to the *SM* treatment.

inspection of these graphs indicates that a treatment effect is present for both types of voters, although it does appear to be stronger for the non-medians.

From the above findings it appears that subjects' preferences move closer to the Part A outcome in the *MA* mechanism. We investigate this further by fitting linear regressions that explain proposals.

In the first column of Table 4, we report results from a linear regression that explains subjects' proposals as a convex combination of their starting points and part A results, and the interaction of these variables with the treatment dummy (which takes value 0 in *MA* and value 1 in *SM*). The estimated values verify what we see in the left panel of Figure 8, and we find all coefficients to be highly significant. Subjects put a substantial weight on the result of part A in choosing their proposal in treatment *MA*. In treatment *SM* this effect is reduced by about two-thirds. In the second and third columns we run the same regressions separately for median and non-median voters. This reveals that while the attractor effect of the Part A result exists for both types, the treatment effect (a stronger attractor effect in the *MA* treatment) is only present for non-median voters.

Subsequently, we introduce two new explanatory variables. The first is the *efficiency maximizer*, which corresponds to the point that, if chosen, maximizes the group's sum of payoffs. This coincides with the median starting point. If a subject cares about efficiency, she would be expected to put some weight on this point in her proposal. The second variable is the *inequality minimizer*, i.e. the mid-point between the two more extreme starting points in each group. Overall, a positive weight on its coefficient should capture subjects' concerns for inequality. Given that we find some differences in part A outcomes with respect to inequality, we want to examine whether they can explain the treatment effects we find in part B.

In the fourth and fifth columns, we report the regression results separately for each voter type.[19] The coefficient for the *efficiency maximizer* is essentially zero, leading us to conclude that subjects are not concerned about efficiency. The coefficient of the *inequality minimizer* is positive but only becomes slightly significant for median voters. Nevertheless, all other coefficients remain highly significant and at essentially the same magnitudes as in the previous two columns. We conclude that there is some degree of inequality aversion among the subjects, but this can in no way explain the large treatment effect that we find.

Next, we instrument the outcome of part A by the means of the Nash equilibrium prediction. Formally, since the part A result appears in the basic specification of the first column both alone and in an interaction with the treatment we need to utilize two instruments: the Nash equilibrium prediction and its interaction with the treatment dummy. We present the second stage of the 2SLS estimation process in the second to last column which gives us largely the same results as our benchmark specification.[20] The additional insight provided by this exercise is that the large differences in the attraction of the group outcome across treatments cannot be attributed to any outcome-related differences. Indeed, this regression shows that even if one focuses only on the part of the group-decision of part A that is explained by the Nash equilibrium prediction, one finds similar results to our benchmark specification.

Finally, in the last column, we present the reduced form of this two-stage approach: We report results from a linear regression that explains subjects' proposals as a convex combination of their starting points and the Nash equilibrium prediction, and the interaction of these variables with the treatment dummy. These results are particularly important, given that the part A outcome similarity between the *MA* and *SM* treatments is established in a stochastic manner. That is, despite the demonstrated affinity of outcome distributions –even when one controls for the exact preference profile– this coincidence is not deterministic: Two groups with identical preferences hardly ever arrive at exactly the same outcome,

[19] We do not include the *efficiency maximizer* for median voters as it too often coincides with their starting point.

[20] As is standard in the literature (see Angrist and Pischke, 2008) when we have more than one endogenous regressors –and, hence, we utilize two (or more) exogenous instruments– the relevant statistic for the first stage regression becomes the Cragg-Donald Minimum Eigenvalue Vector statistic. In our case this statistic takes a value of about 700, which is well above the critical threshold of 7.

both across– and within–treatments. These reduced form results establish that the predicted outcome –the Nash equilibrium one– is a stronger predictor of the subjects' proposals in *MA* than in *SM*, hence reassuring us that the smaller weight that subjects assign to their starting point in *MA* compared to *SM* is not driven by outcome-related differences. It should be noted that these last findings carry an independent interpretation and broader implications as far as implementation of welfare optima is concerned. When a mechanism designer wants to implement a certain welfare optimum –in our case, this is the median of the set that contains the subjects' starting points plus points 33 and 67– and expects that after the voting procedure individuals will try to revise the outcome to their liking, then the mechanism designer should opt for a congruent mechanism. Such a mechanism will enhance the probability that the post-revisions outcome will be as close as possible to the desired policy alternative.

The findings from the regressions in Table 4 together with what is observed in Figure 8 support the following result:

Result 3: *We find strong support for Hypothesis 3. The attractor effect of the Part A outcome differs significantly across treatments, being larger in the* MA *treatment. The treatment effect appears in the preferences submitted by non-median voters. For median voters we do not observe a differential attractor effect across treatments. Overall, the treatment effects cannot be explained by any differences in the outcomes across treatments.*

TABLE 4: Regression results

Dependent variable: *proposals*

	OLS					2SLS 2nd stage	Reduced form
	All	Non-median	Median	Non-median	Median	All	All
Constant	0.857 *[-1.231, 2.932]*	0.105 *[-2.515, 2.758]*	2.048 *[-0.858, 4.946]*	-0.423 *[-3.126, 2.315]*	-0.194 *[-2.930, 2.556]*	0.051 *[-2.383, 2.485]*	2.672^{*} *[0.347, 5.018]*
Starting point	0.730^{**} *[0.652, 0.809]*	0.720^{**} *[0.637, 0.804]*	0.827^{**} *[0.713, 0.931]*	0.717^{**} *[0.635, 0.801]*	0.810^{**} *[0.700, 0.913]*	0.721^{**} *[0.636, 0.808]*	0.727^{**} *[0.643, 0.811]*
Part A result	0.261^{**} *[0.165, 0.357]*	0.287^{**} *[0.176, 0.397]*	0.135^{*} *[0.027, 0.250]*	0.262^{**} *[0.136, 0.388]*	0.112^{*} *[0.001, 0.230]*	0.289^{*} *[0.197, 0.383]*	
Nash							0.228^{**} *[0.129, 0.327]*
Efficiency maximizer				0.011 *[-0.044, 0.067]*			
Inequality minimizer				0.026 *[-0.033, 0.085]*	0.081^{*} *[0.010, 0.151]*		
Starting point $\times$ Treatment	0.179^{**} *[0.083, 0.275]*	0.201^{**} *[0.105, 0.297]*	-0.029 *[-0.206, 0.151]*	0.200^{**} *[0.103, 0.297]*	-0.035 *[-0.211, 0.144]*	0.194^{**} *[0.091, 0.290]*	0.188^{**} *[0.091, 0.284]*
Part A result $\times$ Treatment	-0.181^{**} *[-0.278, -0.084]*	-0.209^{**} *[-0.314, -0.104]*	0.044 *[-0.137, 0.220]*	-0.207^{**} *[-0.312, -0.102]*	0.050 *[-0.131, 0.227]*	-0.200^{**} *[-.300, -0.102]*	
Nash $\times$ Treatment							-0.188^{**} *[-0.290, -0.087]*
Observations	2400	1584	784	1584	784	2400	2400

Notes: Confidence intervals (95%) and p-values are obtained using wild cluster bootstrap, with robust errors clustered at the Subject level (see Cameron, Gelbach, and Miller, 2008, MacKinnon and Webb, 2018 and Roodman, MacKinnon, Nielsen, and Webb, 2018). *Efficiency maximizer* is the point that maximizes the sum of payoffs for the group. *Inequality minimizer* is the point that minimizes inequality as measured by the Gini coefficient or the maxmin criterion. Treatment is a dummy variable that takes value 1 for treatment *SM*. In two instances in each session two voters had the same starting point. These observations are excluded from the regressions run by type.

**: $p - val < .01$

$* : p - val < .05$

5.2.3 The role of apparent consensus with the outcome

According to our theoretical predictions, the increase in the level of actual agreement is achieved in the *MA* due to the increased levels of apparent consensus with the outcomes this mechanism achieves. The attractor effect of the Part A outcome established in Results 2 and 3 is consistent with this prediction, but this does not necessarily mean that it actually holds. As stated in Hypothesis 4, we need to test whether the attractor effect depends on the apparent level of consensus with an outcome. We investigate this here.

In the first column of Table 5, we report results from a regression that explains the deviation of subjects' proposals from the part A outcomes as a function of the distance between the part A results and their starting points, and the interaction of this variable with the treatment dummy.[21] If subjects were behaving in a simple payoff-maximizing manner, there would be no treatment effect (in other words, the coefficient of the interaction would be insignificant, and the dependent variable would be fully explained by our first independent variable). In line with the results presented in the previous section, we do find a strong treatment effect: The coefficient of the interaction is positive and significant, indicating that proposals are farther away from the part A result and closer to the subjects' starting points in the *SM* treatment compared to the more moderate changes in the outcome observed in the *MA* treatment.

Next, we introduce our measure of individual *apparent consensus* (see section 5.1.3) in to the regression (column 2). Recall that this measures a subject's apparent consensus with the outcome, as expressed through her vote. As we observe, the treatment effect vanishes. Indeed, *apparent consensus* seems to fully pick up the differences observed across the two treatments. As seen in Figure 6, *apparent consensus* differs significantly across treatments. One worry here might be that this variable is correlated with some unobserved factor that differs across treatments and the effect captured in this regression is unrelated to *apparent consensus*. To address this, we run regressions restricting the sample to each treatment separately (columns 3 and 4) and obtain very similar results.[22]

If we use *group apparent consensus* (column 5) we observe that this measure cannot explain the degree of deviations of the proposals from the original group's choice. The direct treatment effects remain significant.[23] In fact, one can break down *group consensus* into two components: individual

[21]Due to the nature of our dependent variable, we use the negative binomial estimator.

[22]Results are robust to constructing the measures of consensus using *average* instead of *minimum* distance of vote to outcome (see section 6). The only difference is that, while with minimum distance the treatment effects become unambiguously insignificant (see column 2), using average distance the interaction ceases to be significant at our chosen 1% level, but remains significant at the 5% level.

[23]We also run within-treatment regressions with this variable and find the same to be true. These are not presented

apparent consensus and the agreement of other group members with the outcome. When these variables are introduced separately into the regression, both the treatment effect and *other's apparent consensus* are not significant, while *individual apparent consensus* remains highly significant.

The following result summarizes our findings:

Result 4: *We find qualified support for Hypothesis 4. High levels of individual apparent consensus with an outcome reinforce the outcome's attractor effect. This is true both across and within treatments and explains away the observed treatment effect. On the other hand, the apparent consensus of other group members with the outcome does not influence the strength of the outcome's attractor effect for a specific voter.*

6 Concluding remarks

Our experimental approach tests whether procedures have an effect on participants' preferences. After the first part of our experiment –in which a collective choice is reached through two mechanisms– the second part of the experiment elicits individuals' preferences after a collective decision is taken in the first part of the protocol. The behavior we observe in the second part of our experiment is at odds with self-regarding payoff maximization. As expected, participants' preferences move closer together and away from the payoff-maximizing optima. What we did not anticipate is that this behavior does not appear to be driven by *other-regarding* factors. Frey et al. (2004), for instance, propose that procedural utility is derived from the allocative and redistributive properties of a mechanism, or from how one is treated in interaction with others. In the neutral context used in the lab, subjects remained *self-regarding*, even if not rational: their preferences seem to change in response to their own behavior during the collective choice process, but are not affected by the behavior of others. This is more in line with some preference for consistency, or aversion to cognitive dissonance (we refer the reader to Kamenica, 2012 for a discussion of these concepts). Of course, we do not preclude a role for other-regarding factors in different settings, especially outside of the lab environment. But our findings suggest that procedures can have an effect on preferences for reasons unrelated to procedural justice. Further research in this direction is definitely warranted.

here to save space and are available by the authors upon request.

TABLE 5: Negative binomial regression results

Dependent variable: *proposal deviation from part A result*

	Full Sample (1)	(2)	Treatment MA (3)	Treatment SM (4)	Full Sample (5)	(6)
Constant	1.682^{**} (.000)	2.035^{**} (0.000)	2.409^{**} (0.000)	2.010^{**} (0.000)	1.843^{**} (0.000)	1.954^{**} (0.000)
Deviation from starting point	0.043^{**} (.000)	0.044^{**} (.000)	0.044^{**} (.000)	0.045^{**} (.000)	0.044^{**} (.000)	0.044^{**} (.000)
Deviation from starting point $\times$ Treatment	$.0048^{**}$ (.003)	.0004 (.839)			$.003^{*}$ (.197)	.001 (.570)
Individual consensus		$-.004^{**}$ (.003)	$-.008^{**}$ (.008)	$-.004^{*}$ (.011)		$-.004^{**}$ (.001)
Group consensus					$-.002$ (.309)	
Other's consensus						.002 (.342)
Observations	2400	2400	1200	1200	2400	2400

Notes: The variable *Other's consensus* is calculated as *Group consensus*, but ignoring one's own apparent consensus. Treatment is a dummy variable that takes value 1 for treatment *SM*. The numbers in parenthesis are the p-values that are obtained using wild cluster bootstrap, with robust errors clustered at the subject level. The calculation of confidence intervals is computationally infeasible in the case of the negative binomial regression.
**: $p - val < .01$
*: $p - val < .05$

References

Aaron, Henry J (1994). "Distinguished lecture on economics in government: Public policy, values, and consciousness". In: *Journal of Economic Perspectives* 8.2, pp. 3–21.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Becker, Gary S (1993). "Nobel lecture: The economic way of looking at behavior". In: *Journal of political economy* 101.3, pp. 385–409.

Bolton, Gary E and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition". In: *American economic review* 90.1, pp. 166–193.

Bowles, Samuel (1998). "Endogenous preferences: The cultural consequences of markets and other economic institutions". In: *Journal of economic literature* 36.1, pp. 75–111.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008). "Bootstrap-based improvements for inference with clustered errors". In: *The Review of Economics and Statistics* 90.3, pp. 414–427.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests". In: *The Quarterly Journal of Economics* 117.3, pp. 817–869.

Corazzini, Luca, Kube, Sebastian, Maréchal, Michel André, and Antonio Nicolo (2014). "Elections and deceptions: an experimental study on the behavioral effects of democracy". In: *American Journal of Political Science* 58.3, pp. 579–592.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and behavior: Experimental evidence on the effects of democracy". In: *American Economic Review* 100.5, pp. 2205–29.

De Dreu, Carsten KW and Laurie R Weingart (2003). "Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis." In: *Journal of applied Psychology* 88.4, p. 741.

De Wit, Frank RC, Lindred L Greer, and Karen A Jehn (2012). "The paradox of intragroup conflict: a meta-analysis." In: *Journal of Applied Psychology* 97.2, p. 360.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *The quarterly journal of economics* 114.3, pp. 817–868.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental economics* 10.2, pp. 171–178.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013). "Lies in disguise?an experimental study on cheating". In: *Journal of the European Economic Association* 11.3, pp. 525–547.

Frey, Bruno S and Alois Stutzer (2005). "Beyond outcomes: measuring procedural utility". In: *Oxford Economic Papers* 57.1, pp. 90–111.

Frey, Bruno S, Matthias Benz, and Alois Stutzer (2004). "Introducing procedural utility: Not only what, but also how matters". In: *Journal of Institutional and Theoretical Economics JITE* 160.3, pp. 377–401.

Gerardi, Dino and Leeat Yariv (2007). "Deliberative voting". In: *Journal of Economic Theory* 134.1, pp. 317–338.

Gershkov, Alex, Benny Moldovanu, and Xianwen Shi (2017). "Optimal voting rules". In: *The Review of Economic Studies* 84.2, pp. 688–717.

Gneezy, Uri (2005). "Deception: The role of consequences". In: *American Economic Review* 95.1, pp. 384–394.

Goeree, Jacob K and Leeat Yariv (2011). "An experimental study of collective deliberation". In: *Econometrica* 79.3, pp. 893–921.

Greiner, Ben (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". In: *Journal of the Economic Science Association* 1.1, pp. 114–125.

Grossman, Guy and Delia Baldassarri (2012). "The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda". In: *American Journal of Political Science* 56.4, pp. 964–985. ISSN: 1540-5907.

Hartnett, Tim (2011). *Consensus-Oriented Decision-Making: the CODM model for facilitating groups to widespread agreement*. new society publishers.

Jehn, Karen A (1994). "Enhancing effectiveness: An investigation of advantages and disadvantages of value-based intragroup conflict". In: *International journal of conflict management* 5.3, pp. 223–238.

Jehn, Karen A and Elizabeth A Mannix (2001). "The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance". In: *Academy of management journal* 44.2, pp. 238–251.

Jehn, Karen A, Clint Chadwick, and Sherry MB Thatcher (1997). "To agree or not to agree: The effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes". In: *International journal of conflict management* 8.4, pp. 287–305.

Kamei, Kenju (2016). "Democracy and resilient pro-social behavioral change: an experimental study". In: *Social Choice and Welfare* 47.2, pp. 359–378.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran (2015). "State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods". In: *Experimental Economics* 18.1, pp. 38–65.

Kamenica, Emir (2012). "Behavioral Economics and Psychology of Incentives". In: *Annual Review of Economics* 4.1, pp. 427–452.

Kellermanns, Franz W, Jorge Walter, Christoph Lechner, and Steven W Floyd (2005). "The lack of consensus about strategic consensus: Advancing theory and research". In: *Journal of Management* 31.5, pp. 719–737.

Kelman, Herbert C (1958). "Compliance, identification, and internalization three processes of attitude change". In: *Journal of conflict resolution* 2.1, pp. 51–60.

Lind, E Allan and Tom R Tyler (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

MacKinnon, James G and Matthew D Webb (2018). "The wild bootstrap for few (treated) clusters". In: *The Econometrics Journal* 21.2, pp. 114–135.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2013). "Self-organization for collective action: An experimental study of voting on sanction regimes". In: *Review of Economic Studies* 81.1, pp. 301–324.

Markussen, Thomas, Ernesto Reuben, and Jean-Robert Tyran (2014). "Competition, Cooperation and Collective Choice". In: *The Economic Journal* 124.574, F163–F195. ISSN: 1468-0297.

Mason, Richard O. and Ian I. Mitroff (1981). *Challenging strategic planning assumptions: Theory, cases, and techniques*. John Wiley & Sons Inc.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The dishonesty of honest people: A theory of self-concept maintenance". In: *Journal of marketing research* 45.6, pp. 633–644.

Mellizo, Philip, Jeffrey Carpenter, and Peter Hans Matthews (2014). "Workplace democracy in the lab". In: *Industrial Relations Journal* 45.4, pp. 313–328. ISSN: 1468-2338.

Moulin, Herve (1980). "On Strategy-proofness and Single Peakedness". In: *Public Choice* 35, pp. 437–455.

Núñez, Matias and Dimitrios Xefteris (2017). "Implementation via approval mechanisms". In: *Journal of Economic Theory* 170, pp. 169–181.

Ockenfels, Axel and Alvin E Roth (2006). "Late and multiple bidding in second price Internet auctions: Theory and evidence concerning different rules for ending an auction". In: *Games and Economic behavior* 55.2, pp. 297–320.

Priem, Richard L., David A. Harrison, and Nan Kanoff Muir (1995). "Structured conflict and consensus outcomes in group decision making". In: *Journal of management* 21.4, pp. 691–710.

Renault, Régis and Alain Trannoy (2005). "Protecting minorities through the average voting rule". In: *Journal of Public Economic Theory* 7.2, pp. 169–199.

Roodman, David Malin, James G MacKinnon, Morten Ørregaard Nielsen, and Matthew Webb (2018). *Fast and wild: Bootstrap inference in Stata using boottest*. Tech. rep. Queen's Economics Department Working Paper.

Schweiger, David M., William R. Sandberg, and James W. Ragan (1986). "Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus". In: *Academy of management Journal* 29.1, pp. 51–71.

Schweiger, David M, William R Sandberg, and Paula L Rechner (1989). "Experiential effects of dialectical inquiry, devil's advocacy and consensus approaches to strategic decision making". In: *Academy of Management Journal* 32.4, pp. 745–772.

Sen, Amartya (1997). "Maximization and the Act of Choice". In: *Econometrica: Journal of the Econometric Society*, pp. 745–779.

Sutter, Matthias, Stefan Haigner, and Martin G Kocher (2010). "Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations". In: *The Review of Economic Studies* 77.4, pp. 1540–1566.

Thibaut, John W and Laurens Walker (1975). *Procedural justice: A psychological analysis*. L. Erlbaum Associates.

Tyran, Jean-Robert and Lars P Feld (2006). "Achieving compliance when legal sanctions are non-deterrent". In: *The Scandinavian Journal of Economics* 108.1, pp. 135–156.

Walker, James M, Roy Gardner, Andrew Herr, and Elinor Ostrom (2000). "Collective choice in the commons: Experimental results on proposed allocation rules and votes". In: *The Economic Journal* 110.460, pp. 212–234.

Yamamura, H. and R. Kawasaki (2013). "Generalized Average Rules as stable Nash mechanisms to implement generalized median rules". In: *Social Choice and Welfare* 40, pp. 815–832.

A Instructions

The experiment was run in Greek. We present here a translation of the instructions done by the authors. Original instructions in Greek are available upon request.

A.1 Treatment *SM*

Thank you for participating in this session. Please remain quiet. The experimental session will be run using a computer and all answers will be given through it. Please do not talk to each other and keep quiet during the session. Please note that the use of mobile phones and other electronic devices is not permitted. Please read the instructions carefully, and if you have any questions, raise your hand. The answer that will be given will be announced to everyone.

General Instructions

During the experiment, you can win points. The points will be converted into euros. **1 euro = 15 points.** Each participant will receive a payment. The exact amount you will receive depends on the decisions you will make during the experiment, the decisions of other participants and also on luck. In addition, you will receive the amount of **€3 as a show-up fee**. Following the completion of the experimental session, a fee will be paid privately in cash to each one of you. The experiment consists of two parts. The instructions below are for part A of the experiment. Following the completion of part A, the instructions of part B will be given. Your final earnings will be:
€3 show-up fee + earnings in part A + earnings in part B

Part A

Aim

Part A of the experiment consists of 20 periods. In each period, you will be in a three-member group with two other participants. The aim of the group is to choose a common **destination** from 100 consecutive locations (that is, an integer from 1 to 100) which will be the final decision of the group in the end of each period. The composition of the groups will change in every single period, and you will not be able to know the identity of the members of the group. The way the destination is chosen from the group will be explained below. First, we will explain the way in which the payoffs of each player are determined.

Starting points and payoffs

In each period, a specific destination (that is, an integer from 1 to 100) will be chosen as an **individual starting point**. The payoffs in each period depend on the distance between the final destination that will be chosen and your individual starting point: **The farther the destination is from the starting point of each player, the smaller his/her payoffs will be**. Specifically, each player's payoff (in points) will be calculated as follows:

$$Profits = 100 - |destination - starting\ point|$$

Example:

The group chose the location 49 as a destination.

Player 1's starting point is 20. Player 1's payoff is **71 points**. (The distance between the starting point of player 1 and the final (common) destination is 49 - 20 = 29. So, the payoff is 100 - 29 = 71).

Player 2's starting point is 50. Player 2's payoff is **99 points**. (The distance between the starting point of player 2 and the final (common) destination is 50 - 49 = 1. So, the payoff is 100- 1 = 99)

Player 3's starting point is 95. Player 3's payoff is **54 points**. (The distance between the starting point of player 1 and the final (common) destination is 95 - 49 = 46. So, the payoff is 100 - 46 = 54)

The calculations above will be conducted automatically from the computer, and on the screen you will see your starting point, your group members' starting points, and their payoffs, depending on the chosen destination.

Attention!

- The starting point of each player will be different (unique).
- In each period, the starting points will change.
- All players' starting points will be shown in the screen with arrows.
- Each player's payoff will be indicated by a bar. The greater the payoff, the taller the bar.

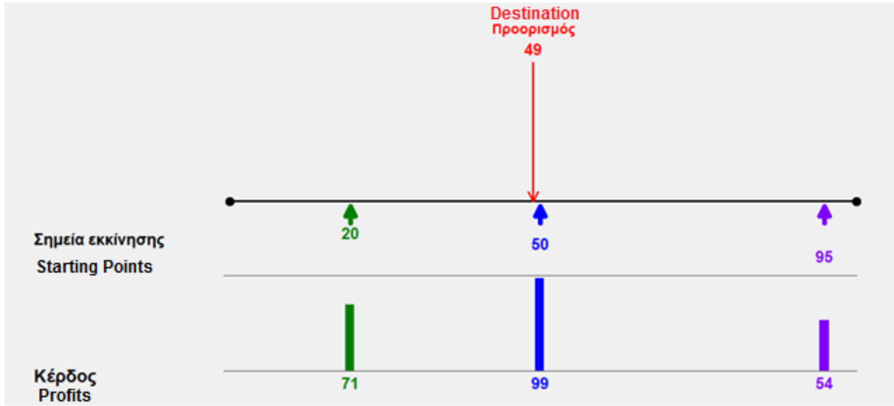


Figure 9

Selection

The selection of the destination will be done as follows: Each group member can vote for exactly one location. The final destination will be the average of all locations voted for by the group members. (In case of a non-integer average, the final destination will be calculated by rounding to the nearest integer).

Example 1:

Player 1 votes 5.

Player 2 votes 80.

Player 3 votes 95.

The final destination will be the location 60 because the average of the chosen locations is: $\frac{5+80+95}{3} = 60$

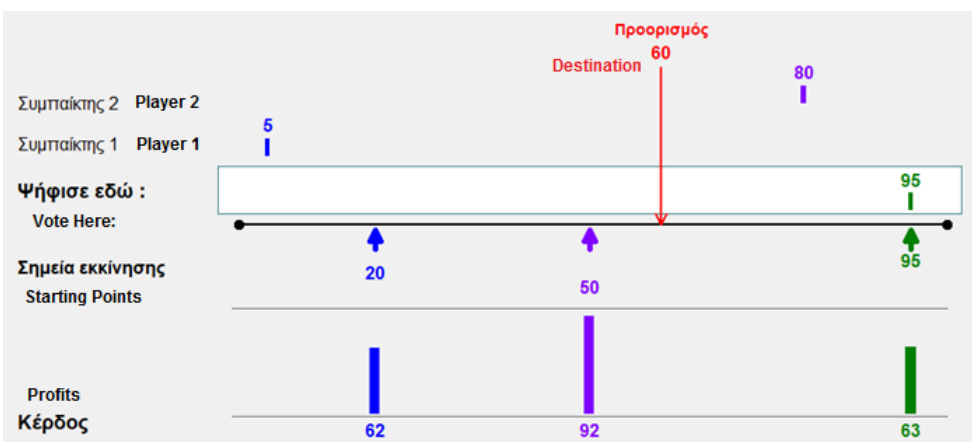


Figure 10

Example 2:

Player 1 votes 30.

Player 2 votes 80.

Player 3 votes 95.

The final destination will be the location 68 because the average of the chosen locations is: $\dfrac{30+80+95}{3} \approx 68,33$

Figure 11

Voting Procedure

Every single period, the voting procedure will last 60+x seconds, where x is a random number from 1 to 10. In other words, following the completion of the voting, the procedure will stop randomly in one of the next 10 seconds.

You can specify the location you are voting by clicking on the white frame you will see on your screen.

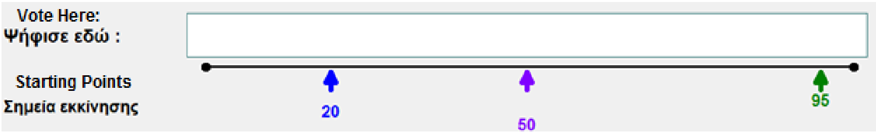


Figure 12

At the same time, you will see what other group members are voting for and how the common destination is shaped. You can change your vote as many times as you want until the voting procedure is over.

The destination of each period **will be determined after the completion of the voting procedure**. Hence, make sure you have made your choice before the end of the 60-second period. In the first two periods, the duration will be 90+x seconds, so as to allow you plenty of time to get used to the procedure. The remaining time will be shown at the bottom of your screen.

At the end of the experiment, **one period from part A will be selected randomly and your payment will be based on your earnings in this period**. Hence, we encourage you to pay attention to all your decisions in all periods, since each of them can determine your final payment.

Before we start, there will be two trial periods to make sure everything is understood. These two trial periods cannot be chosen, and your decisions in those periods will not affect your payment.

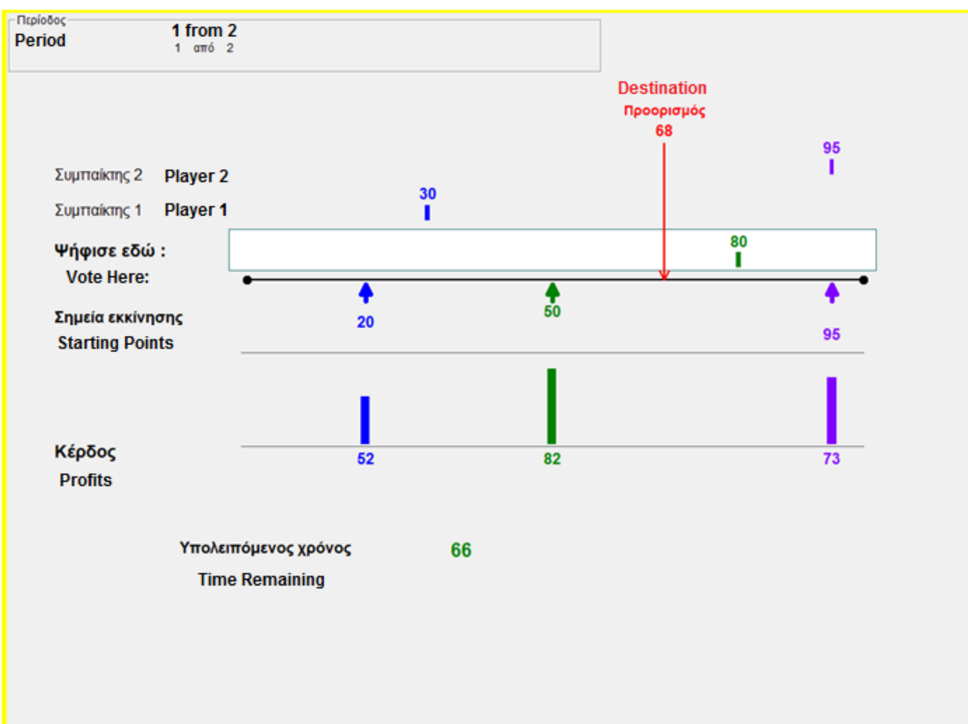


Figure 13

Part B

This part again consists of 20 periods.

In each period you will be placed in the same group you were in the corresponding period of part A, with the same individual starting points.

The aim of the group is to make a collective decision regarding your common destination.

At the top of your screen, you will see a figure in which the starting points, the votes of your teammates and the final chosen destination of the corresponding period of part A will be displayed.

This time the final destination will be determined according to a new procedure. Click in the white frame at the middle of the screen to propose a destination. Your proposal will appear as well as each group member's payoff if your proposal is selected. **You can change your proposal as many times as you like until you press the red button 'Submit'**. When you press the button, your submission will be confirmed, and you will move to the next period. In this part, the proposals of your teammates will be unknown to you until the end of the experiment.

One of the three proposals made by the members of each group will be chosen randomly and become the new common destination for this period, according to which group members' payoffs will be determined.

The profits will be calculated in the same way as in part A by taking into consideration the new destination.

At the end of the experiment, one period from part B will be selected randomly and your payment will be based on your earnings in this period. Hence, we encourage you to pay attention to all your decisions in all periods, since each of them can determine your final payment.

Following the completion of the experiment, you will see on your screen the chosen periods for the

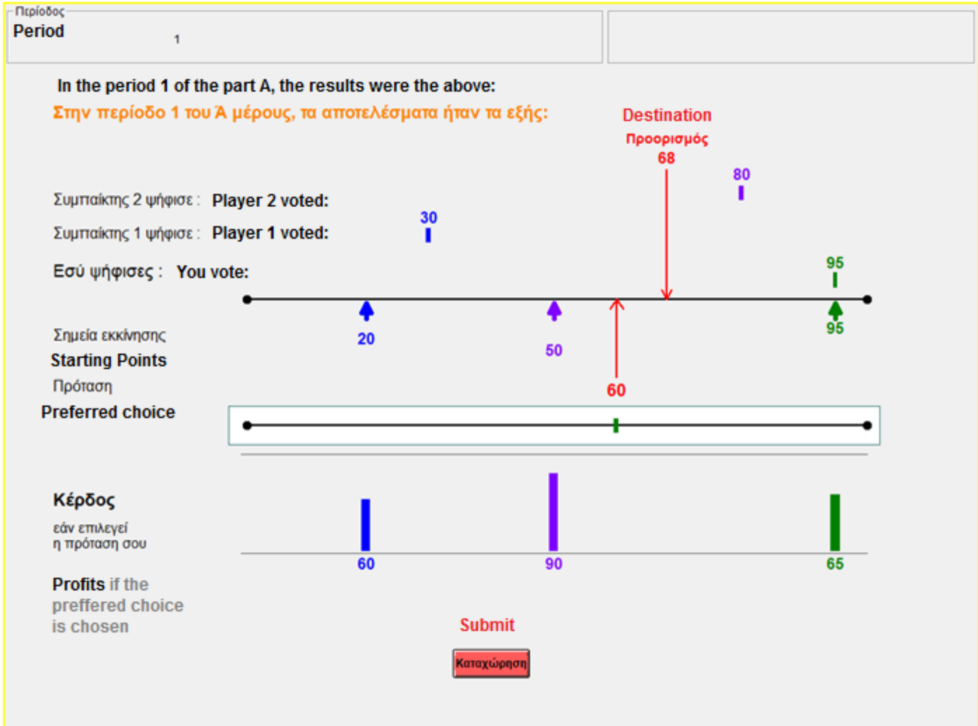


Figure 14

calculation of your profit, as well as your final profit.

B Treatment *MA*

Thank you for participating in this session. Please remain quiet. The experimental session will be run using a computer, and all answers will be given through it. Please do not talk to each other and keep quiet during the session. Please note that the use of mobile phones and other electronic devices is not permitted. Please read the instructions carefully and if you have any questions, raise your hand. The answer that will be given will be announced to everyone.

General Instructions

During the experiment, you can win points. The points will be converted into euros. **1 euro = 15 points.** Each participant will receive a payment. The exact amount you will receive depends on the decisions you will make during the experiment, the decisions of other participants, and also on luck. In addition, you will receive the amount of **€3 as a show-up fee**. Following the completion of the experimental session, a fee will be paid privately in cash to each one of you. The experiment consists of two parts. The instructions below are for part A of the experiment. Following the completion of part A, the instructions of part B will be given. Your final earnings will be:

€3 show-up fee + earnings in part A + earnings in part B

Part A

Aim

Part A of the experiment consists of 20 periods. In each period, you will be in a three-member group with two other participants. The aim of the group is to choose a common **destination** from 100 consecutive locations (that is, an integer from 1 to 100) which will be the final decision of the group in the end of each period. The composition of the groups will change in every single period, and you will not be able to know the identity of the members of the group. The way the destination is chosen from the group will be explained below. First, we will explain the way in which the payoffs of each player are determined.

Starting points and payoffs

In each period, a specific destination (that is, an integer from 1 to 100) will be chosen as an **individual starting point**. The payoffs in each period depend on the distance between the final destination that will be chosen and your individual starting point: **The farther the destination is from the starting point of each player, the smaller his/her payoffs will be**. Specifically, each player's payoff (in points) will be calculated as follows:

$$Profits = 100 - |destination - starting\ point|$$

Example:

The group chose the location 49 as a destination.

Player 1's starting point is 20. Player 1's payoff is **71 points**. (The distance between the starting point of player 1 and the final (common) destination is 49 - 20 = 29. So, the payoff is 100 - 29 = 71).

Player 2's starting point is 50. Player 2's payoff is **99 points**. (The distance between the starting point of player 2 and the final (common) destination is 50 - 49 = 1. So, the payoff is 100- 1 = 99)

Player 3's starting point is 95. Player 3's payoff is **54 points**. (The distance between the starting point of player 1 and the final (common) destination is 95 - 49 = 46. So, the payoff is 100 - 46 = 54)

The calculations above will be conducted automatically from the computer, and on the screen you will see your starting point, your group members' starting points, and their payoffs, depending on the chosen destination.

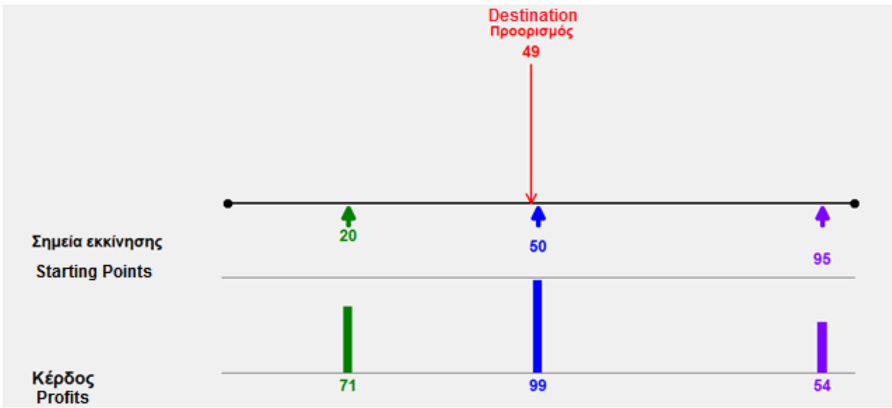


Figure 15

Attention!

- The starting point of each player will be different (unique).
- In each period, the starting points will change.
- All players' starting points will be shown in the screen with arrows.
- Each player's payoff will be indicated by a bar. The greater the payoff, the taller the bar.

Selection

The selection of the destination will be done as follows: Each group member can vote up to 100 locations. These locations should be consecutive (e.g. someone can vote from 23 to 56). The destination will be the largest median of the distribution of the votes. That is, destination X will be selected if at least half of the votes have been given to locations to the left of X + 1 and at least half of the votes have been given to locations to the right of X-1. If there are more than two locations with this feature, the largest one will be selected.

Example 1:

Player 1 votes from 1 to 20 (20 votes).

Player 2 votes from 71 to 75 (5 votes).

Player 3 votes from 91 to 100 (10 votes).

The sum of the votes is 35.

Every single location from 1 to 20, from 71 to 75, and from 91 to 100 has been voted for once. The rest of the locations have not been voted for.

The destination will be location 18 (because the votes that have been given to locations to the left of 18 + 1 are 18 $> 35/2 = 17.5$, and the votes that have been given to locations to the right of 18-1 are 18 $> 35/2 = 17.5$).

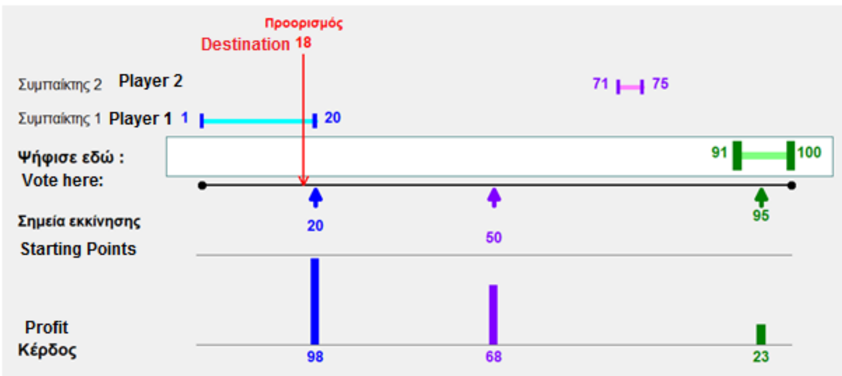


Figure 16

Example 2:

Player 1 votes from 1 to 20 (20 votes).

Player 2 votes from 71 to 75 (5 votes).

Player 3 votes from 71 to 100 (30 votes).

The sum of the votes is 55.

Every single location from 71 to 75 has been voted for twice (by player 2 and player 3).

Every single location from 1 to 20 and from 76 to 100 has been voted for once.

The rest of the locations have not been voted for.

The destination will be location 74 (because the votes that have been given to locations to the left of 74 + 1 are 28 > 55/2 = 27.5, and the votes that have been given to locations to the right of 74-1 are 28 > 55/2 = 27.5).

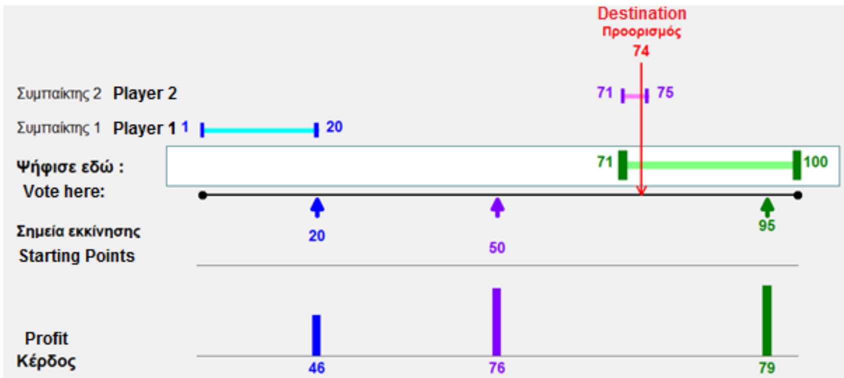


Figure 17

Voting Procedure

Every single period, the voting procedure will last 60+x seconds, where x is a random number from 1 to 10. In other words, following the completion of the voting, the procedure will stop randomly in one of the next 10 seconds.

You can specify the locations you are voting for by clicking and by moving the green bars that you will see on your screen.
At the same time, you will see what your teammates are voting for and how the common destination is shaped. You can change your vote as many times as you want until the voting procedure is over.

The destination of each period **will be determined after the completion of the voting procedure**. Hence, make sure you have made your choice before the end of the 60-second period. In the first two periods, the duration will be 90+x seconds, so as to allow you plenty of time to get used to the procedure. The remaining time will be shown at the bottom of your screen.

At the end of the experiment, **one period from part A will be selected randomly, and your payment will be based on your earnings in this period**. Hence, we encourage you to pay attention to all your decisions in all periods, since each of them can determine your final payment.

Before we start, there will be two trial periods to make sure everything is understood. These two trial periods cannot be chosen, and your decisions in those periods will not affect your payment.

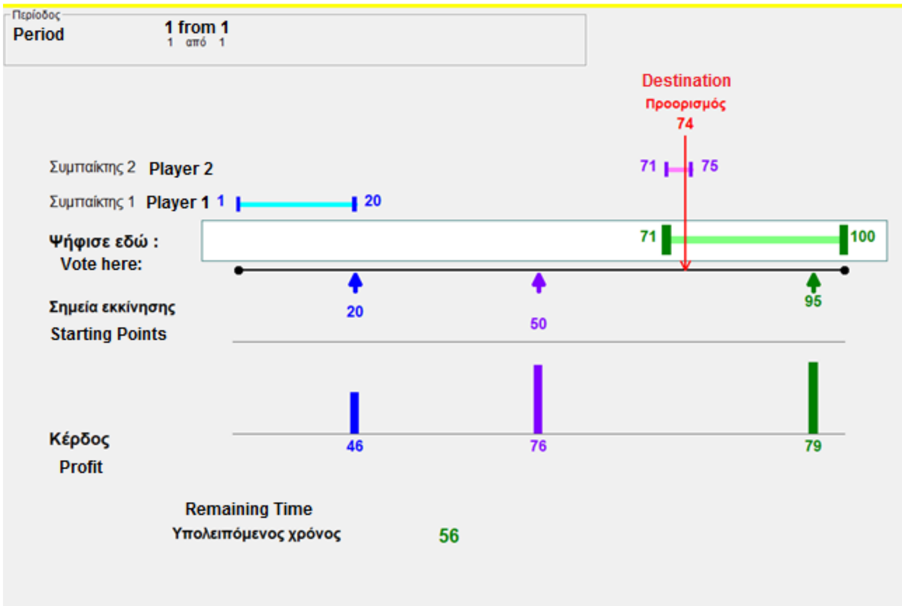


Figure 18

Part B

This part, again, consists of 20 periods.

In each period you will be placed in the same group you were in the corresponding period of part A, with the same individual starting points.

The aim of the group is to make a collective decision regarding your common destination.

At the top of your screen, you will see a figure in which the starting points, the votes of your teammates, and the final chosen destination of the corresponding period of part A will be displayed.

This time the final destination will be determined according to a new procedure. Click on the white frame at the middle of the screen to propose a destination. Your proposal will appear, as well as each group member's payoff if your proposal is selected. **You can change your proposal as many times as you like until you press the red button 'Submit'**. When you press the button, your submission will be confirmed and you will move to the next period. In this part, the proposals of your teammates will be unknown to you until the end of the experiment.

One of the three proposals made by the members of each group will be chosen randomly and become the new common destination for this period, according to which group members' payoffs will be determined.

The profits will be calculated in the same way as in part A by taking into consideration the new destination.

At the end of the experiment, one period from part B will be selected randomly, and your payment will be based on your earnings in this period. Hence, we encourage you to pay attention to all your decisions in all periods, since each of them can determine your final payment.

Following the completion of the experiment, you will see on your screen the chosen periods for the calculation of your profit, as well as your final profit.

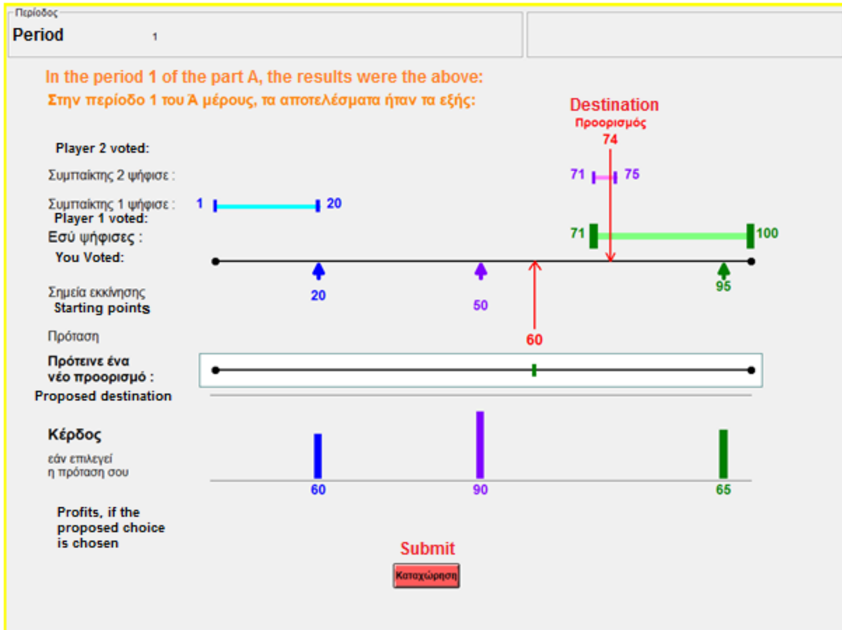


Figure 19