DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

Mechanism Design with Public Goods: Committee Karate, Cooperative Games, and the Control of Social Decisions through Subcommittees

Charles R. Plott
California Institute of Technology

Brian Merlob
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1389

January 1, 2017

Mechanism Design with Public Goods: Committee Karate,
Cooperative Games, and the Control of Social Decisions through Subcommittees

Charles R. Plott and Brian Merlob

January 1, 2017

Abstract

Principles from social choice, voting theory and cooperative game theory are used to design an organization that influences a voting group to choose the alternative preferred by a designer. The designer can impose organization but cannot control options or dictate choice and has limited information about individual preferences. By choice of subcommittee memberships, alternatives available to subcommittees, chairpersons and voting rules, the designer can manipulate group decisions in favor of his preferences. Experiments demonstrate that the core of the resulting game theory model is an accurate predictor of the group choice. Participants think the mechanism is fair.

Mechanism Design with Public Goods: Committee Karate,
Cooperative Games and the Control of Social Decisions through Subcommittees

Charles R. Plott and Brian Merlob
California Institute of Technology

1. Introduction

This study extends the literature that explores how social and group organization can be used to influence group choice. The theme is captured by the word "influence" but the emphasis is on the word "how". Riker [1](1986) studied historical examples of political manipulation.[2] Gerrymandering, strategic voting, agenda and amendment control are recognized examples.[3] Riker coined the term "heresthetics" to describe what he viewed as art of manipulation that borders on science. However, "committee karate" or "organization karate" might be a more descriptive term to capture an underlying theoretical structure of political manipulation in which organization, as opposed to strategic voting behavior, has a fundamental role. Known properties of axiomatic social choice theory and developments in the effectiveness form of cooperative games (to be distinguished cooperative games in characteristic function form and non-cooperative games) are used as tools for organizational manipulations. Laboratory experimental methods are used in tests.

The problem is easy to state for situations that have similar properties. An organization consisting of a group of people (ten in our case) must choose one option from a set of options (fifteen in our case). Individual preferences differ, are well formed with certainty, and are not publically known. Individual preferences are partially known to an administrator, possibly as a result of interviews and private conversations, or other sources of data and observation available to administrators. Thus, preference revelation, the traditional challenge of design postulated in the academic literature, is not an issue. The administrator has personal preferences over the group choice and cannot alter the set of options. No private goods are available so the administrator cannot bribe or use personalized motivations to shape voting choices. Thus, traditional tools used for mechanism design do not exist. The practical objectives are to identify a

[1] William H. Riker, *The Art of Political Manipulation*. New Haven: Yale University Press, 1986.
[2] Riker emphasized the role of language in addition to institutions and the strategic use of procedures as important ingredients of the art of heresthetics. The emphasis of committee karate is on organizational structure. The influence of procedures on committee decisions is well established in the literature. See Plott and Levine (1977); Levine and Plott (1978).
[3] See Chambers and Miller (2013) for geometric illustrations of the consequences of gerrymandering and attempts to control it.

set of controls that could be made available to the administrator and to design an organization that can function under a wide range of parameters to consistently produce as an outcome the alternative the administrator wants. A process design that successfully manipulates should be perceived as being "fair" even though the goals implemented and the mechanism itself might be strongly rejected if the decision making body's informed opinion were to be consulted about design objectives and consequences.

The research purposes are to design and produce such a mechanism; characterize tools and principles that support the design, and to conduct experimental tests that demonstrate how and why the tools work. Many important issues are not addressed and other successful mechanisms might exist.[4] The mechanism studied here consists of two independent and separated decision making centers, viewed as subcommittees, each of which chooses from among a subset of the options. The two choices are elevated to a vote by the organization as a whole. Subcommittees follow a majority rule voting process with open proposals and voting following pre-defined procedures. The administrator influences the ultimate choice through the assignment of members to the centers/subcommittees, the chairpersons, and the options (the committee charges) from which the centers/committees must choose.

The paper consists of a body of theory and a set of experiments. The theory is taken from existing voting theory, cooperative game theory, and social choice theory. No new theorems are produced. At an abstract level of analysis the theory postulates individual behaviors, how they interact with the institutions to produce the social choices, and how they confer such power to an administrator. Experiments are used to place otherwise abstract concepts in a context of observable and measurable variables. The complexity of multiple, interacting variables and theories requires an experimental approach that differs from those traditionally constructed to

[4] Many key issues are not studied here including the possible uses of alternative principles like non-cooperative game models. Our focus is on a specific class of organizational controls and how they can be used on an unsuspecting group. The set of constraints on the administrator, the set of all controls available to the administrator or how the controls might be acquired or modified are important issues not addressed here. Indeed, the evolution of such constraints might be a product of multiple manipulations of different groups over time. Agents other than the administrator are not aware of the administrator's purpose, the implications of the organizational design the administrator imposes or the use to which information revealed to the administrator might be put. Indeed, keeping the public in the dark, exhibiting myopic behavior, might be a key to successful manipulation. Of course, possible defenses or protections from manipulation are interesting and relevant. Perhaps an understanding of how protections can be erected will follow from a better understanding of how manipulation through organization can be achieved, which is the focus here.

test a single theory or hypothesis. The experiments seek answers to two broad questions used to assess the success of complex institutional designs.[5] (1) A proof of principle: does the mechanism do what it is supposed to do? Basically, does the mechanism produce the group choice that it was designed to get? (2) Design consistency: does the mechanism operate according to the theoretical principles on which the design rests? The theoretical principles are important for reliability, scalability and the creation and study of organizations in addition to the one explored here. If the design does not work for the right reasons, if the success is accidental, then there is no reason to think that the mechanism can be scaled in terms of size, structure or environment.

Section 2. Overview

The theory builds from well-known axioms from social choice and game theory and will be recognized immediately by specialists. However, the modifications, interpretations, connections and integration may not be so well known. Section 3 outlines the basic concepts, elements and notations appropriate for the application of the theory. Section 4 is an example of the background theory and its application to voting.

Section 5 develops the institutional structure and behavioral theory. Two levels of theory are studied. One level applies to the behavior of centers (subcommittees) when operating in isolation and addresses how a subcommittee choice can be influenced. The tools are taken from theories of cooperative games, majority rule equilibrium and the core as formulated through cooperative game theory in effectiveness form.

The second level of theory applies to organizational structure, the relationship between different parts of an organization. Manipulation works through a type of "divide and conquer" strategy that works if the divided parts are kept divided. The administrator's control operates through multiple, separated centers. The theory addresses the properties of institutions needed to assure that subcommittee decisions are insulated from each other and thus avoid the emergence of coalitions, , communication, coordination, signaling and other variables that game theory demonstrates are important and serve to channel behaviors away from otherwise myopic behavior. The key property is an adaptation from classical, axiomatic social choice theory.

[5] The questions that guide the experimental procedures were posed by Plott (1994) as a methodology for assessing testbed experiments for newly designed mechanisms and decision processes.

The overall organizational design is simple. A social goal or "target" is identified and the organization is structured to produce the target as the group decision. The "organization chart" that explains the decision center organization is illustrated in Figure 1. The administrator creates two decision centers (subcommittees), assigns the membership, and defines the task of each. Each subcommittee will convene to choose one element from the set of options defined as the committee's task. The procedure used by the subcommittees is a strict majority rule according to a simplified version of Roberts Rules of Order with the default, presumably a status quo, designated by the administrator that will be the committee choice if the committee fails to choose any other alternative. The two options that emerge as the two subcommittee choices will then be voted on by the committee of the whole.

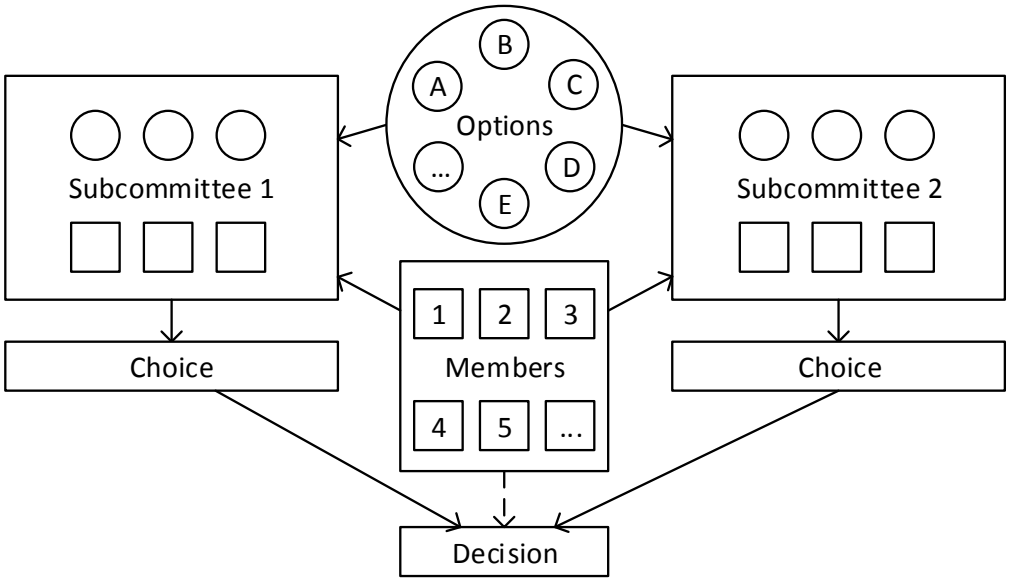


Figure 1. *Organization Chart* – Each subcommittee is assigned a subset of options and members. From among the options assigned, each subcommittee makes a choice. Then, the committee as a whole (the set of all members) makes a decision between the two choices.

Section 6 outlines the experimental testbed, the induced preferences, and the details of the institutions. The logic of the testbed is simply to create an institutional and preference environment and through a series of tests establish if the social choices respond as predicted.

The experiments consist of a fixed set of options and a fixed set of preference types, which remain unchanged throughout the experiments. The preference type assignments to individuals are systematically changed. Two different mechanisms/organizations are studied. One is based on a majority rule equilibration model and the other is based on a veto player game model. A set of target options is designated and tested in a series of trials. For each of the designated targets, the theory is applied to craft an organization and procedures to influence the group to choose the target option while leaving all individual committee members preferences unchanged.[6] The experiments demonstrate that the target option of a trial tends to be the group choice as predicted.

[6] The theories rest on well-established experimental results for committees operating under majority rule and elements of Roberts Rules of order. The predictive power of the majority rule equilibrium and Condorcet winner was experimentally demonstrated by Fiorina and Plott (1978) and replicated and explored further by Berl et.al (1976), McKelvey and Ordeshook (1984), and Herzberg and Wilson (1991). The impact of a single veto player and the core as a solution concept was established by Kormendi and Plott (1982) and extended by Isaac and Plott (1978). The analysis was extended to the case where all participants are veto players by Grether, Isaac and Plott (2001). Recent reviews of the power and versatility of the underlying theory and be found at Bottom, King, Handlin and Miller (2008) and at Wilson (2008a. 2008b.). Kagel, Song and Winter (2010) studied proposal and veto powers in the context of a three person bargaining model.

Section 7 contains the precise predictions of the model when applied to the testbed. As stated in the introduction, the questions are whether or not the mechanism does what it is supposed to do and whether or not it does it for the right reasons. Section 8 reports the details of the experimental procedures.

Section 9 is a statement of results. Overall, the process worked well to produce the target alternatives. The models tended to match the data. Section 10 is a summary of results and addresses several issues that are not addressed in the text. Actually, Section 10 could be read as an introduction. Appendices contain details of preference inducement and aspects of the technologies used in experiments.

Section 3. Notation, Institutions and Solution Concepts

A social choice function expressed as a function of the environment represents the theoretically predicted outcomes of a decision process. The environment consists of the feasible alternatives, individual preferences, decision making rules, and the status quo. The choice itself is a subset of the possible outcomes, interpreted as the solution of an underlying game as defined by the environment.

A. Notation

Y = a universal set of conceivable alternatives over which individuals have preferences.

X = the set of feasible alternatives available and from which a choice is made. The set of feasible alternatives is a subset of the set of conceivable alternatives, that is, $X \subseteq Y$.

N = the set of all individuals, with $i \in N$.

R_i = the preference relation of individual $i \in N$, with $x R_i y$ or as $x \succcurlyeq_i y$. R is the set of all total preorders over Y, $R_i \in R$ and $(R_1,\ldots,R_N) \in R^N$.

Let ***D*** be all asymmetric and irreflexive binary relations on Y. For a dominance relation $D \in \boldsymbol{D}$, xDy means x "dominates" y in a sense to be made more precise later.

A social dominance function, $D(R_1,\ldots,R_N) \in R^N$, is a mapping that assigns to each vector of preferences $(R_1,\ldots,R_N) \in R^N$ a dominance relation.

A (game form) social choice function, $C(X, x_0, D(R_1 ,..R_i ,..R_N)) \subseteq X$, is defined by the following properties: (i) a set of options X, (ii) a "status quo", x_0, which assures that the outcome will be one of the alternatives as opposed to the empty set, (iii) a vector of individual preferences

$(R_1,\ldots,R_N)$, and (iv) a dominance relation $D(R_1,\ldots,R_N)$ reflecting the underlying rules the group choice.

B. Institutions

Institutions reflect the rules of the game, the game form, and are represented by winning and blocking coalitions in a game in effectiveness form.[7] For the analysis developed here the rules, the winning and blocking coalitions, are fixed in the sense that they do not change as the preferences of a decision making group change.

We say that a coalition $c \subseteq N$ is a "winning" coalition if the rules grant the coalition the power to implement any element of the set of options, X, as the social choice. The family of all winning coalitions is W. The concept is defined relative to some fixed set of options X and its subsets. Thus, the winning coalitions for a given a set of options X need not be the same as the winning coalitions given a disjoint set of options X'. The games we will consider are "proper" games in the sense that if c is a winning coalition then its complement is not winning. A "blocking" coalition is a subset of every winning coalition. B is the notation for the family of blocking coalitions.

The families of coalition W and B have the following properties:

$W = c \subseteq N$ such that c is a winning coalition. So if $c \in W$, then, $c' = N \backslash c$ and $c' \notin W$.

$B = c \subseteq N$ such that if $b \in B$ then $b \subseteq c$ for all $c \in W$.[8] That is, a blocking coalition is a subset of all winning coalitions. It follows that if a coalition, c, is blocking for a pair (x,y) in X then it is blocking for all pairs in X.

---

[7] The basic connection with cooperative game theory in environments with public goods and no side payments evolves from Rosenthal (1972).

[8] The concept of blocking can be interpreted as flowing from group rights. While concepts of individual and group rights are not studied here, generalizations open the door for additional applications of committee karate.

Blocking coalitions can be defined relative to pairs of options. If blocking is defined relative to pairs, some of the restrictions of simple games are relaxed and the concept of blocking can become connected to a wider spectrum of institutions. Let $\triangle(x,y)$ be the set of coalitions for which x cannot dominate y unless the coalition prefers x to y. That is, if a coalition c is an element $\triangle(x,y)$ then the system cannot move from y to x (or "move to consider" x over y) unless all members of c prefer x to y. Notice that the family of coalitions in $\triangle(x,y)$ can change as the pair (x,y) changes and that the agents that have blocking power over (x,y) might have no blocking power or influence over the different pair (w,z). Such a representation is natural when the options are divided for choice among different subcommittees where the subcommittees are viewed as participating in different games.

The structure of $\triangle(x,y)$ can be used to define individual and group rights. For example, if one postulates that i is in all c in $\triangle(x,z)$ for all z, then i can prevent moves to x (the system cannot move to x unless i agrees) or in $\triangle(z,x)$ for all z, which means that the system cannot move away from x unless c agrees (if the system gets to x it cannot move away from x unless i agrees). The concept reflects a type of "private rights". One can think of x as having some dimension (such as a good or activity to be allocated or some type of externality) that affects i. The concept can be refined further to reflect sets of options $\cup(i)$ and $\cup(i)$ over

The concept of "dominance" is derived from the abstract concepts of "power" and "preference" of coalitions $c \subseteq N$. The concept of "power" reflects the nature of institutions or the "rules of the game" as defined by the concepts of winning and blocking coalitions. Alternative x "dominates" alternative y, written xDy, if there exists a winning coalition that unanimously prefers x to y.

More generally if $D(R_1 ,..R_i ,..R_N)$ is a dominance function as defined by decision making rules then xDy, if there exists a winning coalition that unanimously prefers x to y given the preferences $(R_1 ,..R_i ,..R_N)$.

C. Game Solutions as System or Organizational Choice[9]

The condition that the social choice be a subset of the feasible set of options, X, allows a connection to notions of equilibria and "game solutions". In this paper only the core will be used as a solution concept. While other solution concepts[10] exist, only the concept of the core of the game will be used.

The core is the set of x in X that are undominated: Core = $\{x \in X$: for all $y \in X, \neg yDx\}$.
So the social choice is the core if

$$C(X, x_0, D(R_1 ,..R_i ,..R_N) = \{x \in X: \text{for all } y \in X, \neg yDx\}.$$

Since N and X are finite, the core is nonempty if a blocking coalition exists. Of particular importance is a special institutional structure that Brown (1973) identifies as a collegium in which the blocking coalition becomes a winning coalition by being joined with selected, disjoint subsets of the individuals.

D. Design Goals and Objectives

---

which i has such control. That is, the relation xDy is blocked if $x \in \iota(i)$ and i does not prefer x to y and the relation yDx is blocked if $x \in \cup(i)$ and i does not prefer y to x. Hammond (1997) makes this distinction in terms of one way rights and two way rights. The concepts extend themselves to the rights of groups and in doing so can be used to formalize the concept of an "amendment control rule" found in Shepsle (1979) and in Shepsle and Weingast (1984). Additional, powerful rules that impact the structure of blocking powers derived from parliamentary procedures are found in Schwartz (2006, 2008).

Generalization can also be achieved through a relaxation of dominance to the case where blocking coalition is indifferent. For example, such a change of definition would be needed to allow exchanges between others, i.e. changes of the social state that do not change the component over which an individual has rights.

[9] The connection between the social preference as found in social choice theory and dominance as found in cooperative game theory is first introduced by Wilson (1972).

[10] An example is the Von Neumann-Morgenstern solution. It is a set, $VM \subset X$: $\{x,y \in VM \Longrightarrow \neg xDy$ and $\neg yDx\} \wedge \{ y \notin VM \Longrightarrow \exists x \in VM$ such that $xDy\}$ The VM solutions can be families of sets. Other solution concepts are outlined for committee process in Isaac and Plott (1978).

The concept of *design* reflects an effort to find underlying processes such that for every situation of feasible alternatives and individual preferences the resulting equilibria, the social choice, is what the designer wants the outcome to be. The goals of the design can be specified as a choice function representing what the social choice "should be" (or "should not be") according to the criteria of the designer. Where $X \subseteq Y$, the goals can be represented as

$$S(X, x_0, R_1,..R_i,..R_N) \subseteq X.$$

The design objective can be stated as S being equal to the social choice

$$S(X, x_0, R_1,..R_i,..R_N) = C(X, x_0, D(R_1,..R_i,..R_N)) \subseteq X.$$

The criteria of success could be stated as a subset, superset, or non-intersection as opposed to the equality sign, $S(\bullet) = C(\bullet)$. The subtle distinction plays a role in the concept of "error" when assessing the accuracy of the design. In the sections that follow, a goal of equality will be assumed so the nature of all model errors can be measured and examined.

Section 4. Example: Theory and Computation Applied to a Single Committee and Alternative Rules

The process to be studied consists of "centers" that operate separately and can be modeled as separate committees. This section analyzes one such decision center. The entire process to be tested is outlined in the next section, Section 5.

Consider a committee of three people, {1,2,3}, that must decide on one of three feasible alternatives, {x,y,z} and have preferences: {1: x≻y≻z; 2: x≻z≻y; 3: z≻y≻x}.The majority rule preference order is a transitive binary relation, x≻z≻y.

Two concepts are central to the analysis of the decision center. The first, a Condorcet winner, is an alternative that, when compared to any other alternative, is strictly preferred by a majority. It would beat all other options in a head to head majority vote. In the example the Condorcet winner is x.

The second concept is the core. Given the individual preferences in the example, the dominance relation is x D z, x D y and z D y. The core is x and it is also the Condorcet winner.

$$M = \begin{array}{c} \\ X \\ Y \\ Z \end{array} \begin{array}{c} \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} & 1 & 1 \\ -1 & & -1 \\ -1 & 1 & \end{bmatrix} \end{array} \quad C = \begin{array}{c} \\ X \\ Y \\ Z \end{array} \begin{array}{c} \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} & -1 & -1 \\ 1 & & -1 \\ 1 & 1 & \end{bmatrix} \end{array} \quad M + C = \begin{array}{c} \\ X \\ Y \\ Z \end{array} \begin{array}{c} \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} & 0 & 0 \\ 0 & & -2 \\ 0 & 2 & \end{bmatrix} \end{array}$$

Figure 2. *Dominance Relation Matrices* – Each cell (α, β) represents a dominance relation such that the cell is 1 if αDβ and – 1 if βDα. If we let the matrix M represent the dominance relation for simple majority rule and C represent the preference for the chairman, we derive the dominance relation $M + C$ for the game. An option is undominated if its associated column only has values less than or equal to 1; hence, in the example below, the core is the pair $\{X, Z\}$. For simple majority rule an option with a non-positive column is in the core, e.g. X in the matrix M.

A matrix representation of binary relations will be a useful tool for computing core. The computation can be difficult for large numbers of options and people. The dominance relation for simple majority rule is represented by the matrix M in Figure 2 in which the cell (α,β) is 1 if αDβ and it is -1 if βDα. If the cell (α,β) is blank or 0, then no dominance exists between the pair. In the matrix representation for simple majority rule, a column with only non-positive entries represents an undominated alternative. The column for x in the matrix M has no positive entries and thus x is in the core. The absence of positive entries means nothing dominates x.

Consider now a rule that gives some subset of agents blocking power such as a process that operates by majority rule but the rules give individual 3 blocking power, in the sense that the rules put individual 3 in every winning coalition. That is, no majority coalition can exercise its power unless individual 3 is in favor of it, but the preference of individual 3 must be backed by a majority in order to establish dominance. Dominance between a pair, for instance, x and y, requires a majority preference plus the preference of individual 3 so individual 3 alone does not constitute a winning coalition. In the matrix notation of Figure 2, the majority preference is represented by the matrix M. The preference of individual 3 is represented by C and the dominance relation is derived by adding M and C to get the matrix M+C. Recall that dominance requires unanimity of members of a winning coalition, which in this case can be stated as a majority plus individual 3. As before if the cell (α,β) is 1 if αDβ and it is -1 if βDα. An alternative is undominated if its associated column contains only numbers less than 2. In the matrix, M+C the columns associated with x and z have only numbers less than two and thus the core is the pair {x,z}. The dominance relation is only zDy because individual 3 and the majority prefer z to y but for all other pairs the majority is blocked because the preference of individual 3 is the opposite of the majority's. The alternative y is dominated by z but both x and z are

undominated. The computational method can be generalized to cases in which the blocking coalition contains more than one person.[11]

Note that if a blocking coalition is a single person, as will be the case in much of the theory applied in this paper, then the most preferred alternative of that person is always in the core. The column of the most preferred alternative of the blocking individual contains no positive elements in the matrix representation (the Z column for individual 3 in our example). The most preferred alternative of the blocking player cannot be dominated and is thus in the core, although the core can contain additional elements.

Section 5. An Organization Designed for Control: Structure and Theory

The manipulation works through the organization used by the decision making group. The focus is on how the organization can be constructed by the designer to implement the designer's preference. The principles that operate depend on institutional detail and how various parts of organization interact. In this section we introduce institutions (features of organization) that exist in various forms in naturally occurring organizations and might be used by the manipulator without arousing suspicion. The organization consists of (A) decision centers, (B) the powers, charges or tasks assigned to decision centers, (C) individuals assigned to decision centers, (D) procedures to be followed within decision centers and (E) the relationship between organizational choice and manipulation goal. In the absence of a language to define complex institutions, we follow the tradition of social choice theory and employ a system of axioms that characterize the variables that will and will not influence specific parts of the organization. Organization is defined in terms of function and behavior as opposed to institutional detail.

A. Decision Centers as Choice Functions.

A decision center is a group of agents that operate as a committee charged with the duty and power to make a choice from a well-defined set of options that we will call the center's "jurisdiction". The jurisdictions are fixed sets of options from which the center must choose one alternative. Jurisdictions do not overlap across centers.

[11] Suppose a set of agents block unless unanimous. Let the C entry be 1 if the multiple agent blocking coalition unanimously prefers α to β, -1 if the blocking coalition unanimously prefers β to α, and , 0 if the blocking coalition is not unanimous. The Pareto Optimal options for the blocking coalition will always be in the core but the core can contain additional elements.

Let N_k be the individuals assigned to the decision center k and let X_k be the jurisdiction of center k. Thus X_k is the set of alternatives assigned to the center and the task of the center is to produce a choice within its jurisdiction, the feasible set of alternatives assigned to it. The jurisdiction assigned to k is a subset of a larger set of options, X, available to the organization as a whole, $X_k \subseteq X \subseteq Y$ where Y is some universal set of alternatives.

Decision center choices reflect preferences of decision center members. Assume R_i is the preference relation for $i \in N_k$. The vector of preferences of the individuals of the center is $(R_1,..R_i,..R_{Nk}) = (R_i, i \in N_k)$. In addition, it has a default option $x_k \in X_k$, which is the choice of the committee should the deliberations of the committee lead to no decision. For purposes of design and modeling, we can represent the decision center k as a choice function that depends on jurisdiction, the default rule and the preferences of the center members.

$$C_k(X_k, x_k, (R_i, i \in N_k)) \subseteq X_k .$$

B. Decision Center Autonomy:

Predictability and control of center choices require that the decision centers operate with autonomy, independently from each other and from other parts of the organization.[12] The following axiom, Independence of Infeasible Alternatives (IFA), captures the idea that two centers act as independent games and is very familiar idea in social choice theory[13] It captures the property of a center as an independent system but it is also a necessary condition for "implementability" in the sense of social choice functions compatibility with the solution of a game.

Independence of Infeasible Alternatives (IFA):

If $R = (R_1,\ldots,R_N)$ are preferences over Y and if $X \subseteq Y$ and if $R' = R \backslash X$ then

[12] The rich and variable possibilities of coalition formation are reviewed by Ray and Vohra (2014). The structure of the theory predicts that coalition formation is a natural tendency that can occur spontaneously in groups. The role of the axiom is to capture the existence of institutions that prevent their formation. Similarly, the intent of the axiom is to capture the existence of institutions and information control that prevent strategic voting that is known to occur naturally (see Bonoit, 2006).

[13] The axiom is a modification of the Arrow axiom of independence of irrelevant alternatives with a status quo variable added and without the strong properties of rational social choice used by Arrow. Plott (1976) recognized that Independence of Infeasible Alternatives is a necessary condition for implementability of social choice functions as solutions to a game. To see the connection with implementability one need only notice that the concept of implementability, as found in the social choice literature, considers only preferences on the feasible set and defines the choice to be a (non-empty) subset of the feasible set. The property is also easy to see in the case of cooperative game theory since the dominance relation is not defined for infeasible alternatives. No coalition has the power to implement an alternative that is not feasible. Similarly, in the case of non-cooperative games no strategy can lead to an infeasible outcome so preferences for the infeasible cannot influence the outcome. Related discussions can be found at Koray and Yildiz (2013).

$C(X, x_0, R) = C(X, x_0, R')$. $R\backslash X$ indicates the restriction of R to the elements of X, i.e. [for all i and for all $x,y \in X$, $xR_i y \Leftrightarrow x R'_i y$].

A decision center is said to act with autonomy if its choice function satisfies IFA. If the choice function satisfies IFA, then the outcome of the choice is not influenced by individual preferences for items that are not feasible. The condition requires that the blocking powers of a decision center do not extend beyond the jurisdiction of the center and that for pairs of options in the jurisdiction only members of the center are members of winning coalitions. Autonomy requires that decisions made by members of a center do not anticipate or influence decisions in other centers. Subgroup meetings, or caucuses by subgroups of a given center or multiple centers, are not allowed. Presumably, their existence could cause the loss of control and a reduction in the success of the manipulation. The axiom rules out strategic behavior in decisions over subsets that will subsequently be part of another choice and rules out the influence of decisions made by other decision centers. It also rules out "strength of preference" variables other than those that might be reflected in a marginal rate of substitution.
The model is neutral about exactly what institutions guarantee the property. However, if the institutions do not guarantee choice behavior that satisfies IFA, then the model may not work.

C. Assignment of Individuals to Decision Centers
Membership in a decision center is assigned. Self-selection is not allowed, so that center membership is not endogenous. Basically, only the preferences of the group assigned to the center have standing as the individual preferences that determine the center's outcome.

Hence the organization of a decision center reflects two properties. First, only those assigned to the decision center contribute to the choice function. Second, due to IFA, the decision center's choice is insensitive to changes in the preferences of individuals outside the center. It is also insensitive to changes in members' preferences for options that are outside the center's jurisdiction.

D. Committee Procedures
Procedures are assigned to decision centers and cannot be changed by the center and do not respond to individual preferences. The procedures begin with a proposal stage where the status quo placed as the motion on the floor. The floor is open for amendments, which might be

recognized and seconded. An amendment that passes a strict majority becomes the motion on the floor. If a proposed amendment fails to get a strict majority the motion on the floor remains. The following rules are imposed and govern committee procedures.

D.1. The recognition process

Any member of the committee can seek recognition and if recognized, places an amendment. If multiple members seek recognition within the timeframe, recognition is exercised through an independent, equally likely random draw. Any motion can be proposed at any time. In particular, a motion that just failed can be proposed again.

D.2. Motion Seconds, the Role of Chairman and *open rules* vs. *closed rule* procedures.

All motions require a second. Two different rules are implemented for study and form the basis of the two different mechanisms. Under conditions of the *open rule* (a type of Roberts Rules), any member of the committee, other than the member that made the proposal, can second the motion. Once the motion is seconded, it goes directly to a vote by the committee. Under the *closed rule* or Committee Chairman sessions, only the chairman has the power to second. Any motion recognized is proposed, but only those motions that are seconded by the chairman can proceed to a vote by the committee. Motions not seconded fail and the floor is open for new motions. Since a Chairman can always "entertain" a proposal for someone else to propose, whether or not the chairman has the power to propose seems to be a minor issue.

D.3. Ending rules.

Any committee member recognized can propose to end the debate and vote on the motion on the floor as the final choice. The motion must receive a second in order to be voted on. The second can come from any member of the committee that is not the proposer.

D.4. Role of status quo.

Should the committee fail to make a decision, the status quo is chosen. The decision process starts at the status quo. If any motion passes, the decision process moves away from the status quo so the status quo will not be chosen unless it is proposed and returns as the motion on the floor.

If only two options exist, then a tie after repeated votes results in the choice of the status quo.

E. Organizational Choice and Manipulation

The discussion now turns to the relationship between center choices and the choice of the organization as a whole. Against the background of some universal set of alternatives Y, let the set or options available to the organization be $X \subseteq Y$, and partition the X into two sets, X_A and X_B. Choose two defaults, x_{0A} and x_{0B} plus x_0, which is the default for the committee of the whole. Let N be the members of the organization and partition them into two sets N_A and N_B. The organizational decision process consists of two separate subcommittees plus a committee of the whole. Subcommittees choose an alternative from their jurisdictions, respectively, and report the decisions to the committee of the whole. The committee of the whole then decides from the narrowed set of two options chosen by the subcommittees respectively. The organizational structure is formally represented as:

$$C(X, x_0, R_i\ i \in N) = C(C_A(X_A, x_{0A}, R_i\ i \in N_A) \cup C_B(X_B, x_{0B}, R_i\ i \in N_B), x_0, R_i\ i \in N)$$

The committee as a whole should be faced with two alternatives. The vote between the two alternatives is the final committee choice. The default is x_0 , the alternative that will be implemented if the voting fail to produce an outcome.

Successful manipulation follows if one of the two final options is the alternative preferred by the designer and the other is an alternative that will be defeated by a majority vote when placed against the alternative preferred by the designer. Thus the designer must know some options that the target option can beat in a majority contest. Given the two final options desired as the runoff, the focus of the design then folds back to the creation of subcommittee processes that will lead to their choice. To these ends, the designer chooses subcommittee tasks and subcommittee membership such that the core of each subcommittee is the alternative that the designer wants that subcommittee to choose. In essence, the target consists of three alternatives, one for each subcommittee and one of those two is the target for the committee of the whole.

Section 6. The Experimental Testbed

The testbed proceeds in a series of periods within a fixed environment and in each period applying the manipulation method outlined in the previous section to get a different outcome. In

each period, a target is designated and a group decision is made. From period to period, the set of alternatives and the set of preferences (the preference types) that exist in the overall group do not change. However, as the target changes, the committee assignments and committee jurisdictions change such that the underlying model predicts that the target will be the option chosen by the group. The two research questions (1) and (2) above, are answered by a study of the success rate with which the target alternative are chosen and by the capacity of the model to explain the behaviors exhibited by individuals and by groups in the voting.

The set of fifteen alternatives X = {A,B,…,O} remains fixed throughout the testbed, although as will be explained in the appendix, the labels are switched and rotated throughout the testbed to mask identical test situations from subjects. Preference types are induced. The set of preference types, T = {1,2,…,10}, in the environment are fixed throughout the experiment, although as explained in the appendix, the preference types are rotated among subjects to prevent obvious issues of long term strategies that could otherwise link tests. The preference types are contained in Table 1. Each type has a strict preference ordering over the fifteen alternatives as contained in the Table. The monetary incentives associated with the preference orderings are in Table 3 and will be explained in detail in Appendix 3. The preference types are illustrated in a two dimensional spatial configuration in Figure 3.

The fact that the options and preference types are fixed throughout the testbed means that the majority rule "preference" or "dominance relation" remains fixed throughout all parts of the testbed. The majority rule preference is contained in Table 2. Of significance is the fact that a Condorcet winner exists among all alternatives and preference types. Option A is preferred by a majority to any other option. Furthermore, the strict majority rule preference is acyclic in the sense that it contains no preference cycles. However, the majority rule preference order is not strictly transitive since it can be that x is preferred to y and y to z but x ties with z under majority rule given the even number of people.

Member Type	Ordinal Rank of Options														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	L	G	D	B	J	I	A	E	F	C	N	H	M	K	O
2	L	I	D	G	N	A	B	F	C	J	E	K	H	M	O
3	I	N	D	F	L	A	G	C	B	K	E	H	J	M	O
4	N	F	I	K	C	A	D	H	B	L	G	E	M	J	O
5	K	F	N	C	H	A	I	D	E	B	M	G	L	J	O
6	K	H	C	F	M	A	E	B	N	D	I	J	G	L	O
7	M	H	E	C	K	A	J	B	F	D	G	I	N	L	O
8	M	E	J	H	B	A	C	G	D	K	F	L	I	N	O
9	J	E	M	B	G	A	H	C	D	L	F	I	K	N	O
10	J	G	B	L	E	D	A	M	C	I	H	F	K	N	O

Table 1. *Experimental Preferences* – Throughout the experiment, the underlying preferences of members over the fifteen options remains the same. The option O is the least preferred option for all members and serves as the initial status quo for all periods. The majority rule order is A≽B≽C≽D≽E≽F≽G≽H≽I≽J≽K≽L≽M≽N≽O where ≽ indicates a majority preference or a tie.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A		6	6	6	6	6	6	6	6	6	6	6	6	6
B	4		5	5	5	6	6	5	6	6	6	7	6	6
C	4	5		6	5	5	5	6	6	6	7	6	6	6
D	4	5	4		5	6	6	5	6	6	6	7	6	6
E	4	5	5	5		5	5	6	5	6	6	5	7	6
F	4	4	5	4	5		5	5	6	5	6	6	6	7
G	4	4	5	4	5	5		5	5	5	6	6	5	6
H	4	5	4	5	4	5	5		5	5	5	6	6	5
I	4	4	4	4	5	4	5	5		5	6	5	5	7
J	4	4	4	4	4	5	5	5	5		5	5	5	5
K	4	4	3	4	4	4	4	5	4	5		5	5	6
L	4	3	4	3	5	4	4	4	5	5	5		5	5
M	4	4	4	4	3	4	5	4	5	5	5	5		5
N	4	4	4	4	4	3	4	5	3	5	4	5	5	

Table 2. *Majority Rule Dominance Relations* – The option A is the Condorcet winner of the committee as a whole because it is preferred by a majority to all other options. For the remaining options, there is a weak Condorcet ordering in that options are dominated only by a subset of options (represented via letters) that precede it in the alphabet.

Ordinal rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
$	2.50	2.33	2.17	2.00	1.83	1.67	1.50	1.33	1.17	1.00	0.83	0.67	0.50	0.33	0.17

Table 3. *Subject Value Schedule* – This table conveys each period's monetary payoffs (in $) for all participants. Payoffs were generated by equation $Payoff = 2.76 - 0.16 \times (Ordinal\ Rank)$.

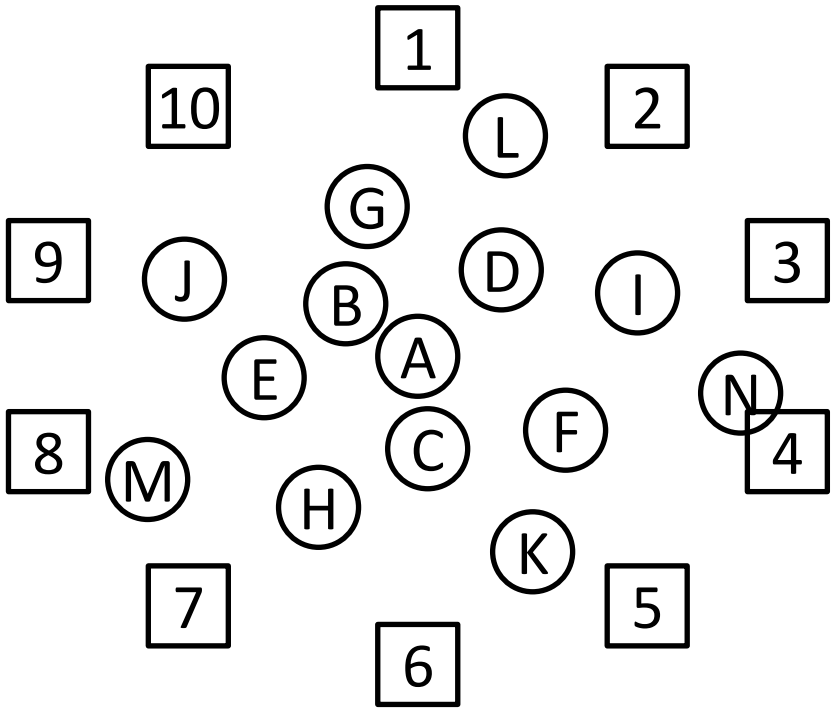


Figure 3. *Spatial Representation of Preference Profile* - In the two dimensional representation above, members (□) prefer options (○) that are spatially closer to options that are further away. The option A is Condorcet winner because a majority (6 out of 10) of members prefers it to any other option. The option O (not pictured) is everyone's least preferred option and the initial status quo.

The existence of a Condorcet winner among all alternatives plays two background roles as part of the testbed design. First, it is an attractive theory of behavior and in ordinary majority rule settings the Condorcet winner is known to serve well as a theory of equilibrium. It is a natural outcome of the committee process. Thus, the ability to design an organization that consistently and predictably chooses some alternative other than the Condorcet winner is a measure of organizational control and the power of the underlying influence model. Secondly, the condorcet winner when supplemented with majority rule preference acyclicity provides an intuitive measure of "distance" of an outcome from the Condorcet winner. This distance is a measure of organizational manipulation and control since intuition suggests that the process would naturally equilibrate "up" the majority rule order toward the Condorcet winner, which is alternative A. With the intuitive theory of dynamics and social preference in mind, one would expect the voting outcome would gravitate toward the Condorcet winner. The power of the "karate" theory to avoid the intuitively appealing outcome can be appreciated. In the course of the testbed, various

alternatives other than the overall Condorcet winner, alternative A, will be designated as a target and when so designated will be the alternative chosen by the relevant decision center/ subcommittee.

A. Testbed Structure

The testbed design is outlined in Figure 4. Two separate, identical experiments were conducted. Each experiment consisted of ten subjects, ten preference types, and fifteen alternatives. The pattern of preferences for all fifteen alternatives was identical across both experiments. One alternative was designated as the status quo, and at the same time the least preferred alternative by all preference types. Each experiment consisted of two separate sessions, specified by the underlying models, assignments, and procedures used to accomplish control: a Condorcet Mechanism and a Chairman Mechanism. Each of the two sessions focused on the ability of the mechanism to influence the organization to choose the specified target, which the organization was designed to choose. As will be discussed below, from the point of view of design the implementation of the Condorcet Mechanism required more information about individual preferences than does the Chairman Mechanism.

		Session 1: Condorcet Mechanism			**Session 2: Chairman Mechanism**		
Experiment 1	Organization configuration	Subcom. 1	Subcom. 2	Committee of the Whole	Subcom. 1	Subcom. 2	Committee of the Whole
	decisions	15 votes	15 votes	15 votes	15 votes	15 votes	15 votes
		45 total decisions			45 total decisions		
Experiment 2	Organization configuration	Subcom. 1	Subcom. 2	Committee of the Whole	Subcom. 1	Subcom. 2	Committee of the Whole
	decisions	15 votes	15 votes	15 votes	15 votes	15 votes	15 votes
		45 total decisions			45 total decisions		

Figure 4. *Testbed Design* – The testbed consists of two experiments, each with two sessions. Within each session, both subcommittees and then the committee of the whole vote 15 times. Those 15 votes are composed of 3 repetitions of 5 configurations (described in Tables 5 and 6). Subjects had preferences induced by monetary incentives. Voting followed the organization and rules imposed for the design. The tests are designed to reveal the influence of the organization and the accuracy of the models that predict the influence. In all, 20 subjects made 180 decisions.

The broad organizational framework was the same for all experiments and all sessions. Two subcommittees (Subcommittee 1 and Subcommittee 2) were created consisting of five members each. Committee jurisdictions consisted of seven alternatives each and the subcommittees were charged with choosing one option from the jurisdiction assigned to the subcommittee. Subcommittee choices were transmitted to a committee of the group as a whole, which made the final decision from the two options presented to it. This final decision served as the basis for payment to all subjects.

Five different group targets were chosen in each of the two sessions (the Condorcet mechanism and the Chairman mechanism) for each of the two experiments. The core was the behavioral model used to configure subcommittee organizations that would result in the group target as the final choice of the committee as a whole. Since the final decision depended on the choices made by subcommittees, the selection of a group target implied a selection of a target for each of the two subcommittees in addition to the target for the group when acting as a whole. The organization was configured such that the subgroup target was an element of the core and each subgroup choice contributed to the overall test of the underlying behavioral model.

Thus, within a session as defined by the mechanism (Condorcet or Chairman), each of the five group targets were actually decomposed into three (sub) targets. Tests were replicated three times for each of the five group targets producing 15 decisions for each of the three choosing groups within a session. That is, a session consisted of three voting groups (two subcommittees and a committee of the whole) so each session produced 45 committee decisions (fifteen by each of the two subcommittees and fifteen by the committee as a whole). Given two sessions per experiment and two experiments, the total is 180 decisions.

The two sessions are defined in terms of the mechanisms that were applied to control the group choices (Condorcet Sessions and Chairman Sessions). Choice manipulation using the equilibrium or Condorcet winner design required information about all preferences. In the Condorcet sessions, the assignment of types and the alternatives allocated as the jurisdiction were chosen such that the sub-target for each subcommittee existed as a Condorcet winner. Each group target was accompanied by changes in committee preference profiles and alternatives such that a Condorcet winner existed and was predicted as the choice. Of course, the Condorcet

winner is the core if it exists, but it is not always possible, or might be difficult to assign preference types and alternatives such that the sub-goal would be a Condorcet winner. Thus, the Condorcet sessions are a solid but limited test from the point of view of a testbed of the manipulation power of the mechanism.
From an institutional design and "committee karate" perspective, the Chairman mechanism has advantages over the Condorcet mechanism because under the Chairman mechanism, the core always exists, and finding an element of the core requires only the information about the most preferred option of the chairman as opposed to the whole preference ordering. However, it might not be possible to configure the institutions such that the core is exactly the target alternative.

In the Chairman mechanism sessions, the organizational configuration and the use of the closed rule procedures in particular, gives the Chairman special powers and thus can be interpreted as blocking powers in the model. That is, the chairman can be modeled as veto player, or, "blocking coalition" who becomes decisive only if joined by a majority. Possible design limitations are created by two features. First, the chairman's optimum alternative is always in the core but it could be the case that no possible committee member has the target alternative as an optimum. Second, other alternatives can also be in the core. If a Condorcet winner exists, it is also in the core as can be alternatives that are "between" the chairman's optimum and the Condorcet winner. As a result, in the Chairman sessions, the core always existed but sometimes contained multiple alternatives that include the chairman's optimum, a Condorcet winner, and possibly additional alternatives.

The testbed asks whether the chosen elements will be in the core and if so, which elements. The Chairman sessions focus directly on the empirical issues by choosing targets such that the target is always an element of the core. On some occasions, the chairman's most preferred was the unique alternative of the core and on others was part of a multi-element core. Defining the target to be the most preferred alternative of some member of the subcommittee and changing the identity of the chairman to align the target with the design created a test of the ability of the organization configuration to control the alternative chosen by the subcommittee.

For the Chairman sessions, the target alternative was changed across sessions while the committee jurisdictions and preference types remained unchanged in the two subcommittees. The configuration for Subcommittee 1 had no Condorcet winner and the core was sometimes a

unique alternative, the chairman's optimal. By contrast, alternative A, the overall Condorcet winner, was always available for Subcommittee 2 and was always in the core for that subcommittee. This pattern of configurations provides a study of conditions when the Condorcet winner does not exist and the core is unique, when the Condorcet winner does not exist and the core is not unique and when the Condorcet winner does exist and the core contains the Condorcet winner together with other alternatives. Thus, the power of the Chairman mechanism and the core is tested under a variety of circumstances and the possibility that the core favors some elements over others, e.g. chairman's most preferred versus Condorcet winner, can be studied.

The Condorcet mechanism is structured such that the majority rule relation in each committee contains no majority rule cycles. The weak majority rule relation is transitive but due to ties the strict majority rule relation is not. Classical voting theory with the open rule suggests that in each committee an application of a simple version of Roberts Rules will lead to a choice of the Condorcet winner. Each subcommittee will choose the option the designer wants and the runoff vote between subcommittee choices will result in the desired outcome.

B. Committee Procedures

Initially, the default option, option O is designated as the motion on the floor. It is the least preferred option for all members. The motion on the floor is then amended through an amendment process that occurs in four stages.

Stage 1: Motions - During this 10 second stage, members may do one of the following: nothing; propose an amendment to the motion on the floor; or propose that the subcommittee recommends the current motion on the floor. Once the timer expires, a proposal will be chosen at random for further consideration as an alternative to the motion on the floor.

Stage 2: Seconds to Motions- During this 5 second stage the floor is open for a second to the proposed option.

> Under the Condorcet mechanism, any member, except the proposer, may second the recognized proposal and bring it to a vote. If no one seconds the proposal, the process returns to the previous stage.
>
> Under the Chairman mechanism, only the chairman of the committee has the power to second a motion. Thus, the chairman has blocking power in the sense that the Chairman

is in all winning coalitions. However, the chairman alone is not a winning coalition. It is important to notice that this single rule is the only difference between the Condorcet mechanism and the Chairman mehanism.

Stage 3: Majority Rule- Any motion that passes by a majority of the committee members' votes, will pass. Voting is required.

Stage 4: Motions to end- Any member can propose a motion to end.

C. Preference Inducement

Traditional methods of using money to induce and control preferences were used. The control for long term strategies and interdependence of strategies among centers required rotations of the preference types and permutation of the alternative names/labels. These are reviewed in appendix 3.

Section 7. Model Predictions

As outlined in Section 5, the testbed procedures reflect the underlying normative goals and theory of the organizational decisions. Tables 4, 5, and 6 contain the targets and behavioral predictions.[14] Results are outlined in Section 8.. The testbed environment has four features that challenge the design and individually or collectively could be reasons why the model might fail to work. They also suggest the existence of interdependencies that could mask a clear view of the reasons for any observed model failures. Thus, the analysis and model accuracy will be assessed primarily from the point of view of the final organizational decision and to the extent that the final decisions match the model predictions the challenges are overcome.

(1) There is a large number of feasible options (15), so an ability to predict a specific option must overcome any inherent randomness that might be reasonably expected to produce any outcome with some probability;

(2) The experiments are conducted in an environment in which the preferences of the entire collective have a Condorcet winner, which serves as a natural standard against which the success of achieving other targets can be compared. That is, if the agents were to engage in a majority rule process following Roberts Rules with open proposals and

[14] Some alternatives were not chosen as potential targets due to the lack partitions of agents and alternatives that would produce the target as an alternative given the induced preferences. Experimental resources also played a role in limiting the number of experiments. Options EJM were not chosen as targets in the Condorcet organization and options BHN were not chosen as targets in the Chairman sessions.

voting, the group deliberations would converge to an equilibrium - the Condorcet winner, which is alternative A. The Condorcet winner from among all alternatives is a natural equilibrium and thus it could be difficult to have a design that reaches any other outcome. Thus the Condorcet winner is a possible outcome against which successful influence can be compared;

(3) The design depends on the reliability of subcommittee processes. If one of the subcommittees deviates, then the whole system can fail to hit the target. The design can employ a degree of robustness in the sense that the target will be chosen even if a subcommittee fails to choose as predicted.

(4) The core often has multiple elements but the target is always a single alternative, so the theory itself predicts that the target option will not be chosen with certainty.

Table 5 contains the predictions of committee decisions when operating under the Condorcet mechanism for both Experiment 1 and Experiment 2. Each experiment lasted for fifteen periods (five configurations repeated three times) and the sessions that existed in a given period were the same in both experiments. Each period consisted of three committee decisions: Subcommittee 1, Subcommittee 2, and the Overall Committee consisting of the members of both subcommittees. The core and the target alternative were the same for both experiments for a given committee and a given period as shown in Table 5. Under the Condorcet mechanism, the core and the target are always the same alternative. This is because in the Condorcet mechanism, the core is a single element and the flexibility of the design allowed a configuration of the parameters such that the core and the target coincided. Thus, the normative target was always the predicted outcome for the Condorcet sessions.

Table 6 contains the predictions of committee decisions and the core of the underlying social choice model when operating under Chairman mechanism for both experiment 1 and experiment 2. Similar to the Condorcet sessions, each of the Chairman sessions lasted for fifteen periods (five configurations repeated three times) and the sessions that existed in a given period were the same in both experiments. Again, each period consisted of three committee decisions: Subcommittee 1, Subcommittee 2, and the Committee as a whole consisting of the members of both subcommittees. The core and the target alternative were the same for both experiments for a given committee and a given period as shown in the table.

#	Status Quo	Motions					Motion Recognized	Voting				
		1	3	5	7	9		1	3	5	7	9
1	O	L	I	K	M	J	K	K	K	K	K	K
2	K	L	I	K	M	J	L	L	L	K	K	L
3	L	L	I	N	M	J	J	L	L	L	J	J
4	L	L	I	N	M	J	M	L	L	M	M	M
5	M	J	I	K	M	J	K	M	K	K	M	M
6	M	J	I	N	M	J	M	x	x	x	M	x
7	M	J	I	N	M	J	J	J	J	M	M	J
8	J	L	I	K	H	J	I	J	I	I	J	J
9	J	L	N	K	H	J	H	J	J	H	H	J
10	J	L	N	K	K	J	K	J	J	K	K	J
11	J	L	I	N	J	J	J	x	x	x	J	J
12	J	L	I	N	J	J	L	L	L	J	J	J
13	J	J	I	N	J	J	J	J	x	x	J	J

Table 4. *Dynamics of Committee Process* – This table accompanies Figure 5. These dynamics follow from two assumptions: (i) members follow the myopic strategy of motioning for their most preferred option that has not yet been compared to the current status quo, and (ii) members vote for their more preferred option in any binary vote. Members only vote to end if all more preferred options have already been compared to the current status quo.

		Condorcet Session							
		Subcommittee 1				Subcommittee 2			
	#	Target	Core	Members	Options	Target	Core	Members	Options
Configuration	1	B	B	3, 4, 7, 8, 9	BDFIKNL	G	G	1, 2, 5, 6, 10	GACJEHM
	2	D	D	1, 2, 3, 5, 9	DGAEKHM	C	C	4, 6, 7, 8, 10	CJBFINL
	3	F	F	4, 5, 6, 9, 10	FBNDJGL	I	I	1, 2, 3, 7, 8	IAECHMK
	4	L	L	1, 2, 4, 5, 10	LDCJEKM	H	H	3, 6, 7, 8, 9	HABFGIN
	5	K	K	1, 5, 6, 7, 10	KABDGIL	N	N	2, 3, 4, 8, 9	NFCEJHM

Table 5. *Experimental Configurations for Condorcet Session*– For each configuration, a target option is selected and an assignment of members and options to subcommittees is determined such that the core of each subcommittee contains the target. In the Condorcet session, the core is unique, so the target and core are equivalent. Within each subcommittee, there is a strict Condorcet ordering among the options (there are no cycles or indifference among the options). The options are presented in the sequence of the strict Condorcet ordering

		Chairman Session							
		Subcommittee 1 (Odd Members)				Subcommittee 2 (Even Members)			
	#	Target	Core	Chairman	Options	Target	Core	Chairman	Options
Configuration	**1**	L	L	1	LJINHMK	D	D, A	2	DGABFCE
	2	I	I	3	INLKHJM	F	F, C, A	4	FCADBGE
	3	K	K	5	KNHIMLJ	C	C, A	6	CFAEBDG
	4	M	M, H	7	MHKJINL	E	E, B, A	8	EBACGDF
	5	J	J, H	9	JMHLIKN	G	G, B, A	10	GBEDACF

Table 6. *Experimental Configurations for Chairman Session* – For each configuration, a target option is selected and an assignment of members and options to subcommittees is determined such that the core of each subcommittee contains the target. In the Chairman session, the assignment of members and options to subcommittees remains the same throughout all configurations: (i) all odd-numbered members (1, 3, 5, 7, 9) and options H – N are assigned to Subcommittee 1, and all even-numbered members (2, 4, 6, 8, 10) and options A – G are assigned to Subcommittee 2. The only change between configurations is the identity of the chairman. Note that the core is not always unique and that within Subcommittee 2, the option A is preferred by a majority to all other options (but is never the chairman's favorite). The target is always the most preferred alternative of the chairman for that subcommittee. The options are presented in the sequence of the chairman's preferences.

For the Chairman sessions, the alternative most preferred by the chairman is always an element of the core but a Condorcet winner is also an element of the core. Depending on the majority rule relation and the preference of the chairman, other alternatives can also be elements of the core. Thus, the core need not be unique as is illustrated in Table 6. Shown there for all periods and all committees are all elements of the core with the most preferred of the chairman shown at the left hand side of the list and the Condorcet winner listed on the right hand side. The core can be a single alternative as it is in periods 1, 2, and 3 of Subcommittee 1 but in all other cases the core contains more than one element.

The underlying conditions of the testbed when combined with the tools available for influencing the decisions create limitations on the design. In particular, in some occasions, without substantial reconstruction of committee assignments and jurisdictions, it is difficult to construct a configuration of the parameters such that the core is the single alternative designated as the target. However, the target is always an element of the core. Table 6 lists both the target for each period and the core.

Given institutional constraints, it may not be possible to design a process that always results in a choice of the target alternative. Nonetheless, the resulting design inaccuracy can be readily anticipated and measured. In particular, as the core becomes larger, it becomes less likely that the target will be selected.

Specifically, consider the case of two subcommittees and let C_i, be the core of subcommittee i and let x be the target alternative, which is one of the alternatives considered by Subcommittee 1. The measure assumes that the core occurs with certainty. Let Z be the alternatives considered by Subcommittee 2 that are in C_2 and also that x dominates in a majority rule sense and let $\Sigma_{z\in Z}$ Pr(z| C_2) be the sum over the relevant events. Under those assumptions the theoretical accuracy is measured by

$$\text{design accuracy} = [\Pr(x|C_1)\Pr(C_1)][\ \Sigma_{z\in Z}\ \Pr(z|\ C_2)\Pr(C_2)].$$

Given that the models predicting the core outcomes occur without error, the probabilities $\Pr(C_1)$ and $\Pr(C_2)$ are both 1 for purposes of this measurement. For the uniform distribution case where ties are ignored, the formula simplifies to design accuracy = (1/n)(k/m) where: n is the number of elements in the core of Subcommittee 1 (which has our preferred option x); m is the number of elements in the core of Subcommittee 2; k is the number of elements in the core of Subcommittee 2 that are dominated by x.

Note that if one of the subcommittees fails to produce the target alternative, then the overall target may be missed. Given our setup, even if Subcommittee 1 fails, as long as Subcommittee 2 succeeds, then the choice of the overall committee will be on target. This is because every option considered by Subcommittee 2 is preferred by a majority to every option considered by Subcommittee 1. Hence, the rate at which the overall committee is predicted to reach the target alternative is 33% in the Chairman sessions with three elements in the core and 50% when there are two elements in the core. There are two instances of two element cores and three of three element cores, so the hit rate is predicted to be 40% or a miss rate of 60%.

The results of the testbed and the outcomes of all experiments will be discussed in Section 8. The discussion includes design success and the accuracy of the underlying behavioral model, together with insights about the dynamics that became understood only after studying the committee behaviors.

Section 8. Experimental Procedures

Subjects were students at the California Institute of Technology and were recruited for 2-hour sessions through laboratory subject databases and dormitory announcements. All subjects were inexperienced (participated only once). As a group, they had no knowledge of the complexities of social choice (the possibility of cycles, the existence of a core, etc.).

Both experiments were conducted via computer terminals in the Caltech Laboratory for Experimental Economics and Political Science. Discussion was not permitted during the experiment and partitions existed that prevented clear views of other subjects. Instructions (Appendix 1) were printed, distributed, and read to all subjects. All values were stated in experimental currency units. Individual subject exchange rates differed across subjects but remained the same for a given subject throughout the experiment and were private information. Average earnings were $51 for the session ($1.70 average per decision).

Experiments were programmed and conducted with the software z-Tree (Fischbacher 2007). Screenshots of the program are shown in Appendix 2. As stated in the instructions, options appear on the subject's screen from most preferred on the left, to the least preferred on the right. Subjects were also able to see a history of all past recognized motions and votes.

Section 9. Results

It is useful to recall that the only difference between the Condorcet process and the Chairperson process is that the latter only the chairperson can second a proposal. Two classes of results are outlined. Section A addresses the reliability of the organization to manipulate the outcome as intended. Section B is focused on the dynamic processes at work. Three facts create a challenge for performance. First, the group had an overall Condorcet winner. It is the core given the majority rule dominance relation and is a natural equilibrium should the entire group be governed by majority rule without the subcommittees and the subcommittee procedures. Any successful design needs to overcome that tendency. Secondly, success of the design is defined by a single, desired outcome. Performance is measured as either 0 or 1. Furthermore, there are many alternatives in the sense that if the outcomes are random from among the alternatives then the probability that any one alternative would result is small, 1/15. So, any underlying randomness works against the design. Third, the process design consists of several separate

processes and if one of those processes fails to function the performance will not be according to design.

The outcomes of experiments are reported in Tables 7 and 8. Before reviewing the outcomes, an example of the dynamics of a single committee process might be useful. The entire process for one period is contained in Figure 5. The options are displayed in a two dimensional representation in a manner that maintains the consistency of a quadratic loss function in the sense that if the point of maximum is known for an individual then the preference between options is captured by the distance from the individual's location. The grey lines represent the dominance relation and the black, dashed lines represent motions that passed (an arrow) and failed (an x). Table 4 contains the details of the proposals and votes of the decision represented in Figure 5. The first proposal to be considered is for option K, which receives a majority to become the new motion on the floor. Option L is proposed and passes. A motion to move to J fails and then alternative M passes. A motion to change the motion on the floor from M to K fails but a motion to move to J passes. With J as the motion on the floor several motions for alternatives I, H, K fail, as does a motion to end debate and choose J. A motion to move to L fails and a motion to end and accept J wins. The final committee choice is J.

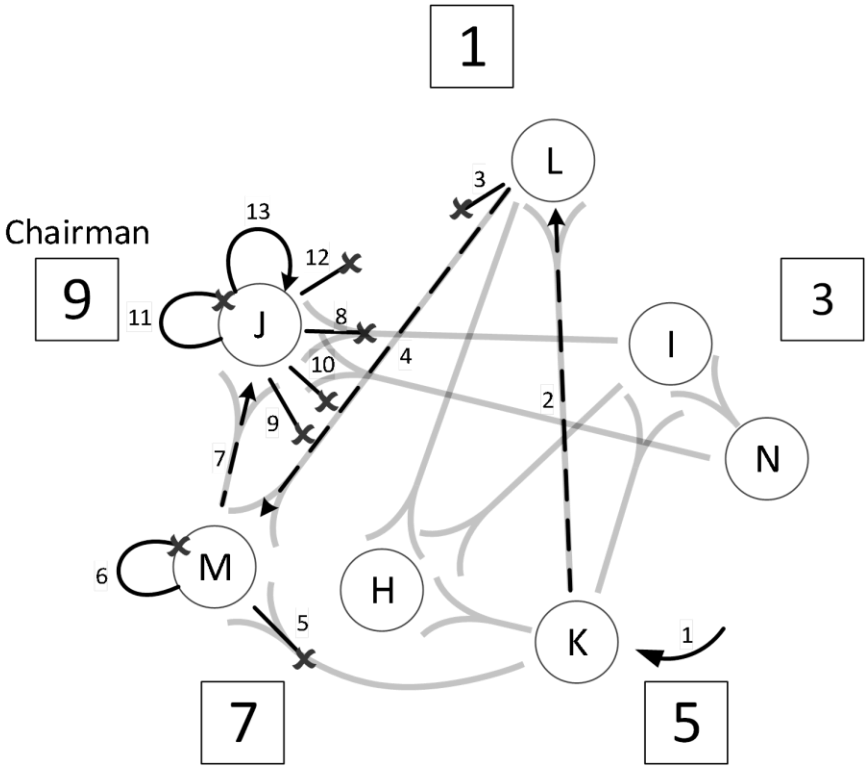


Figure 5. *Dynamics of Committee Process* – In the two dimensional representation below, members (□) prefer options (○) that are spatially closer to options that are further away. The dominance relations (≺) show which options are preferred by *both* the chairman (here, member 9) and the majority. Note that there are two elements in the core (H and J are both undominated), so the dynamics of the committee process influence the resulting decision. Note that the gray lines represent dominance relationships and black lines represent motions. The motions and votes from the figure are reported in Table 4. Initially, the status quo is O. The first movement (by unanimous vote) is to K [(1)], which is recognized randomly from among the 5 motions (one from each member). From K, it moves to L [(2)], and then fails to move to J [(3)] because the majority prefers L to J. It proceeds to move to M [(4)], fails to move to K [(5)], fails to end at M [(6)], moves to J [(7)], fails to move to I [(8)], H [(9)], and K [(10)], fails to end at J [(11)], fails to move to L [(12)] and finally succeeds to end at J [(13)].

		Condorcet Session														
		Subcommittee 1					Subcommittee 2					Overall Committee				
	#	Target	Core	Outcomes			Target	Core	Outcomes			Target	Core	Outcomes		
Experiment 1	1	B	B	B	B	B	G	G	G	G	G	B	B	B	B	B
	2	D	D	A	D	D	C	C	C	C	C	C	C	A	C	C
	3	F	F	F	F	N	I	I	I	I	I	F	F	F	F	I
	4	L	L	L	L	D	H	H	H	H	H	H	H	H	H	D
	5	K	K	K	K	K	N	N	N	N	N	K	K	K	K	K
Experiment 2	1	B	B	B	B	B	G	G	G	G	G	B	B	B	B	B
	2	D	D	D	D	D	C	C	C	C	C	C	C	C	C	C
	3	F	F	F	F	F	I	I	M	I	I	F	F	F	F	F
	4	L	L	L	L	L	H	H	H	H	H	H	H	H	H	H
	5	K	K	K	K	K	N	N	N	N	N	K	K	K	K	K
		Expected		(30/30)			Expected		(30/30)			Expected		(30/30)		
		Experimental		(27/30)			Experimental		(29/30)			Experimental		(27/30)		

Table 7. *Results for Condorcet Session* – The results are listed in the "Outcomes" columns. The core was unique in every configuration, so the expected accuracy is 100% (30/30). The results closely match this prediction, with an accuracy of 90% for Subcommittee 1, 97% for Subcommittee 2, and 90% for the overall committee.

		Chairman Session								
		Subcommittee 1			Subcommittee 2			Overall Committee		
	#	Target	Core	Outcomes	Target	Core	Outcomes	Target	Core	Outcomes
Experiment 1	1	L	L	L J L	D	D, A	A D A	D	D, A	A D A
	2	I	I	I I I	F	F, C, A	C A C	F	F, C, A	C A C
	3	K	K	K H K	C	C, A	C A C	C	C, A	C A C
	4	M	M, H	K H M	E	E, B, A	A B A	E	E, B, A	A B A
	5	J	J, H	J J J	G	G, B, A	B B G	G	G, B, A	B B G
Experiment 2	1	L	L	L I L	D	D, A	B A D	D	D, A	B A D
	2	I	I	I I N	F	F, C, A	C C F	F	F, C, A	C C F
	3	K	K	K H K	C	C, A	C C C	C	C, A	C C C
	4	M	M, H	H M H	E	E, B, A	A E A	E	E, B, A	A E A
	5	J	J, H	J J M	G	G, B, A	G G B	G	G, B, A	J J B
		Expected		(24/30)	Expected		(12/30)	Expected		(12/30)
		Experimental		(20/30)	Experimental		(12/30)	Experimental		(10/30)

Table 8. *Results for Chairman Session* - The results are listed in the "Outcomes" columns. The core was not unique in every configuration, so the expected accuracy is not necessarily 100%. Instead, if we assume that each element of the core is equally likely to be chosen, then when there are two elements, the expected accuracy is 50% and when there are three elements the expected accuracy is 33%. Across configurations, we get an average expected accuracy of 80% (24/30) for Subcommittee 1 and 40% (12/30) for Subcommittee 2. The target for the overall committee is always the target of Subcommittee 2, so we also have a 40% (12/30) average expected accuracy for the overall committee.

A. The Mechanism is Successful

As shown in Tables 7 and 8, the decision process performed substantially as it was designed to perform. Result 1 states that the performance of the system conformed to the design, but was imperfect. Result 2 explains that those imperfections were not due to a lack of reliability in the underlying theory, but instead a consequence of having multiple elements in the core.

RESULT 1: The organization systematically influenced the group choices to choose the target option. The Condorcet sessions were the most successful, followed by the Chairman sessions.

Support. The target alternative was chosen substantially more frequently than can be explained as random, which would have the target being chosen on the order of 6% of the trials. For subcommittees in the Condorcet sessions, 56 of the 60 trials (93%) resulted in a choice of the target. As mentioned before, the target success of the committee as a whole is sensitive to the

target success of the separate committee decisions that constitute the choice of the committee as a whole. For the committee as a whole, with subcommittees operating in Condorcet sessions, 27 of the 30 trials (90%) resulted in a choice of the target. For subcommittees operating in the Chairman sessions, 32 of 60 trials (53%) resulted in a choice of the target. For the committee as a whole, with subcommittees operating in Chairman sessions, 10 of the 30 trials (33%) resulted in a choice of the target, which is on the order of the 40% hit rate predicted.

Result 1 says that the overall group choice was influenced by the organization as predicted. Result 2 leads to a better understanding of the reliability of the tools. The question posed is focused on accuracy of the core. Given the alternatives available to a committee how well are the committee choices predicted by the core (as opposed to the target). The subcommittees have all of the assigned alternatives available but the committee as a whole has only the alternatives available that happened to filter through the subcommittees. In addition, the core has multiple alternatives. The result says that given the condition for the application of the theory, it does very well as a predictive tool.

RESULT 2: The system outcomes tended to be in the core of the dominance relation among the options available to the appropriate committee. When the core contained multiple elements, the winning alternative was not biased toward some particular alternative in the core such as the Chairman's optimum or the Condorcet winner. The tendency of the choice to include any element of the core, as opposed to just the target, resulted in proportionate degradation of design accuracy.

Support. The choices tended to be the alternatives in the core. In the Condorcet sessions, the core (the Condorcet winner) was a single element and in those experiments, 56 of 60 subcommittee trials (93%) resulted with the only alternative in the core. Thus, under the Condorcet mechanism, the alternative targeted as the subcommittee choice was the choice. In the chairman sessions, 52 of 60 trials resulted with an alternative in the core. Of those 52 trials, 32 were the chairman's optimum (i.e. the target), 9 were the Condorcet winner, and the remaining 11 were additional elements in a multi-element core. These results are reported in Table 7 and Table 8 respectively. Across both sessions, all binary decisions made by the committee of the whole were the core (the majority preferred option) of the two alternatives.

While the chosen alternatives were in the core the design accuracy was less than perfect because the core often contained elements in addition to the target alternative. Of the eighteen cases for which the core had one alternative (the target), 13 of the eighteen committee decisions were in the core (.72). When the core had two alternatives, there were 24 committee decisions of which 21 (.875) were in the core and 14 out of 24 were the target (..67). When the core had three alternatives, there were 18 committee decisions of which 18 (1.0) were in the core and 5 (.277) were the target. The core was an accurate predictor of the outcome and among the elements of the core the target was chosen with about the same probability as were other elements of the core. Of course, whether this feature of proportionality is a general property or not requires additional theory and experiments.

The model accuracy of the committee when voting as a whole reflects the fact that subcommittees must choose the target alternative in order for the committee as a whole to choose the target. In the Condorcet sessions, the target alternative was the choice for 27 out of the 30 committee of the whole choices (.90). In the chairman sessions, the committee as a whole chose the target alternative 10 out of the 30 decisions (.333), which is in line with the special case accuracy prediction of .333.

Section B. Properties of Dynamics

The success of the core as a behavioral model leads to questions about the micro principles of decision and the properties of the dynamics. What types of voting and sequences of motions lead to the core? The results below suggest that non-strategic (myopic) models of behavior dictate the dynamics. In particular, Result 3 shows that the next step in a dynamic path of motions is dictated by the dominance relation. Result 4 shows that the myopic behavior stems from both the decisions to propose motions and from the voting behavior. If strategic voting was predominant, one might have deviations from dominance.

RESULT 3: The dynamic movement of the motion on the floor follows the path of the dominance relation. That is, a motion to move the alternative on the floor to a different alternative succeeds if the proposed change is to a dominating alternative and the motion fails if the destination is not a dominating alternative.

Support. Across all experiments, a total of 301 motions were made to change the motion on the floor to some other alternative. Of the 301 motions, 85 motions proposed movements to an

alternative that dominated by the motion on the floor. All 85 failed, resulting in only 216 movements of the motion on the floor. Of the 216 movements, 211 (98%) followed the dominance relation. The 5 movements that did not follow the dominance relation are discussed as errors in Result 5.

RESULT 4: Individual proposals and individual voting tend to be sincere (preference revealing).

Support. The sincere nature of proposals is supported by the fact that nearly every motion proposed is sincere in the sense that the proposed alternative is preferred to the motion on the floor by the proposer: 98.5% of motions (3928 of 3988) are for an option preferred by the proposer or to end deliberations. Moreover, proposed motions follow a predictable pattern: 62% directly follow the strategy of proposing their most preferred option not yet considered against the current motion on the floor. An additional 13% propose their most preferred option despite previous consideration. Another 11% propose their most preferred not yet rejected (a strategy that is rational in an environment without cycles). Thus, over 86% of motions are part of a myopic "hill climbing" ("sincere") strategy. Of the 60 insincere motions, 52 were made after a failed attempt to end with the individual's preferred option. This instance of insincerity can be interpreted as strategy to compromise on a slightly less preferred option rather than endure protracted debate that might end with an even less preferred option.

The voting on motions proposed also follows a pattern of sincere, preference revealing similar to proposals. Nearly all votes are sincere: 94.5% of votes (1422 of 1505) are cast for the preferred option and only 83 were not.

Observation 1: (evidence of dynamic strategies). Instances of insincere voting appear only under special circumstances. Of the 83 insincere votes, 60 are cast for O – the initial motion on the floor, which is everyone's least preferred option. The observation is that these votes for O may reflect a dynamic strategy to retain a "bad" option that has no chance of being chosen, while attempting to influence a motion in the preferred direction. Regardless, these individual votes were never effective in blocking an amendment.

Given the possibility of strategic behavior, it is useful to study the detail of instances in which the core model was inaccurate. The next result is that these exceptions to model predictions do

not contain evidence that suggests they resulted from strategies more sophisticated than sincere voting.

RESULT 5. The core failed to contain the group choice in 11 of 180 committee decisions. These inaccuracies of the model reveal no systemic departure from the basic principles of sincere voting behavior that support the use of the core as a behavioral model.

Support. (i) In 8 of the 11 model errors, the error reflects a premature termination in the sense that an agent voted to end without proposing a preferred alternative that had not been defeated by the motion on the floor. (ii) In the remaining 3 errors, 1 was due to insincere voting and 2 were due to the chairman seconding motions that were for a less preferred option.

The patterns of subcommittee (centers) behavior suggest that overall design created behavior consistent with the Independence of Infeasible Alternatives axiom. The setting of subcommittee decisions within the larger organization had no effect on subcommittee behavior. That is, information and incentives remained local.

Observation 2. Decision Centers acted with autonomy.

Support. The design produced no direct test of the axiom since preferences and committee jurisdictions systematically changed throughout the testbed. A direct test would require the jurisdiction and preferences in one center to remain constant while changed in the other center. However, Results 4 and Result 5 amount to restatements of the axiom at the individual level. Individual decisions tended to be sincere. The decisions reflected only the immediate, local environment. The exceptions to sincere voting were few and exhibited no relationship with the preferences or options under consideration by the members of other centers.

The combined results indicate that the design did what it was supposed to do and did it according to the principles that lead to the design in the first place. Both Result 1 and Result 2 report outcomes that were substantially predicted by the model. Indeed, the accuracy is accurately predicted by the model. The success of the design cannot be attributed to accident or randomness. The reliability of the basic principles are supported by a study of the dynamics (Result 3) and individual behavior (Result 4). Behavior inconsistent with the model can be traced

to specific errors of the model (Result 5) as opposed to some broad failure of the principles. Observation 2 connects the behavior with the most fundamental of the organizational properties that enable the success of the design.

The final observation addresses participant perceptions of bias. Is the organization perceived as biased in the sense that participants realize the purpose implicit in the organizational design and reject the organization as a result? A questionnaire distributed after the experiment asked the questions directly. It appears that the participants did not perceive that the group as a whole was manipulated by the organizational design and did not observe the overall process as being unfair.

Observation 3. Participants did not regard the process as "unfair". Participants did not perceive the overall process as biased and to the extent that bias was perceived, it was confined to the detailed operations of a committee.

Support. Manipulation occurred equally in Condorcet and Chairman sessions, yet perceptions of fairness were narrowly focused on whether or not one person was perceived as having an advantage. Among the subjects 80% thought that the results in Condorcet sessions accurately reflected preferences and 55% thought that the Condorcet mechanism was fair. By contrast, 25 % of the subjects thought that the decisions in the Chairman experiments "accurately reflected preferences" and 15% thought that the decisions in the Chairman experiments were "fair".

From the perspective of implementation, it appears as though the groups were unaware of the manipulation that took place. Thus, the operation of Committee Karate can proceed through the implementation of a process that appears fair but is not.

Section 10. Summary of Conclusions

This section addresses issues and questions that are not fully explored in the text and can best be addressed in the light of results. Prominent features of the experimental environment were typical of naturally occurring environments where we suspect manipulation is possible but private goods and associated potential for tailored incentives are not possible. Indeed, the environment has features of examples studied in the literature such as the flying club (Levine and Plott, 1977), the airport slot committees (Grether, Isaac and Plott, 2001), as well as familiar parts of everyday life such as clubs, home owners associations, businesses or even universities

where deliberative bodies are free to develop their own procedures. We studied groups of ten people, with fixed, well-formed and conflicting preferences over fifteen alternatives. The environment contained neither money nor any other form of private goods that could be used as side payments (transfers). The groups used well defined, organizational procedures for conducting meetings and decision.

Our interest is in underlying principles. However, as orientation we outline a simple rule of thumb algorithm that a committee karate practitioner might try.

(1) Determine a target alternative, the alternative the practitioner wants chosen.

(2) Agenda theory suggests that the manipulation power is enhanced by a type of "divide and conquer" strategy. Partition options and choose an option from each of the two sets such that one of the two chosen options is the target alternative and the second of the two options is one that the target option will beat in a majority vote of the committee as a whole (or by the committee that will be assigned to make a final resolution of choices).

(3) Key steps are the allocation of people to subcommittees, the appointment of a committee chairman and the designation of rules for the subcommittee decision process. The objective of the steps is that the two alternatives be the respective committee choices. Use available information about individual preferences to appoint committee members such that for each group the desired alternative is the core of the respective voting group. If the Condorcet winner (majority rule core) does not exist, then appoint chairpersons such that the alternative the designer wants to be the winner of subcommittee voting is the alternative most preferred by the chairperson. Give the chairperson blocking power, such as the unique power of a second or a power that prevents votes on certain proposals (such as recognition or germaneness). This arrangement assures that the target alternative will be in the core of the implied game representation of the subcommittee decision process.

(4) Make sure that the subcommittees are autonomous and that there is no coordinated voting across subcommittees and no side payment within a subcommittee. If the independence of infeasible alternations is violated by the emergence of informal organization or coordinated strategies across the subcommittees then manipulative control could be compromised.

(5) Pass the subcommittee choices back to the group as a whole or designate a special committee to make the final, runoff decision. Since the two options emerging from subcommittee votes

were strategically chosen such that the non-target alternative would lose in a runoff, the choice of the group as a whole will be the target alternative.

A. Framework and Principles

While manipulation through organizational properties is mechanism design, the purpose and principles employed diverge from the traditional approach to mechanism design. (i) The task is to design a system, a "mechanism", to manipulate the group to choose an alternative we want as opposed to an alternative that would be efficient and that would be the outcome if actions revealed individual preferences. Thus, design purpose is not closely tied to the behavior of individuals as is the case if preference revelation is important (see Section 3-C). (ii) The behavioral principles are tied to collective decisions, coalitions or macro-type, solution concepts typical of cooperative game theory (see Section 3-A). Specifically, the predicted system behavior is tied to the core of an appropriately structured cooperative game as opposed to, say, the Nash equilibrium of a non-cooperative game. While the core can be supported as a Nash equilibrium in some games, that relationship is not a fundamental feature of the principles used here. In part, this departure from tradition avoids limitations of non-cooperative games when applied to voting models[15] . For example, the Condorcet equilibrium is the core of an underlying cooperative game and if preferences are common knowledge it is also a subgame perfect Nash equilibrium but in the absence of common knowledge the subgame perfect Nash equilibrium need not exist. Thus, while the Condorcet equilibrium is used in the design, it is not a consequence of an assumption about Nash equilibria and it is not a consequence an assumption about myopic behavior. (iii) The cooperative game model is not based on the classical characteristic function form of cooperative games. Instead, the dominance relation is heavily dependent on the effectivenss form in which dominance is derived from an abstract concept of "power" and "preference" (see Section 3-B) that can systematically restrict coalition formation to subcommittee members. That feature of the model allows the separation of committee decisions as required by autonomy motivated concepts and the decentralized nature of the overall organization, central issues when considering different organizational designs.

[15] Under majority rule almost all outcomes can be supported as a Nash equilibrium and thus that equilibrium concept is consistent with all observations.

Of course, it might be possible to use the classical characteristic function form to do the work but the effectiveness form of cooperative games is flexible, easily applied and produces a reliable model as is demonstrated by the experimental results. The effectiveness form lends itself to integrate concepts from different literatures (axiomatic social choice theory, cooperative game theory, non-cooperative game theory, voting theory, experimental methods and mechanism design). It also places few restrictions on the organizational forms explored. For example, the model could be applied to more than three committees, creating either flatter or vertical organization.

B. Model Assessment Methods

The fact that the model is about systems behavior requires a departure from typical experimental tests and measurements. The mechanism is a system with many interacting parts and numerous variables. The system is interdependent in the sense that outcome of the system cannot be correct unless it is correct at each subsystem. An error of one variable can cause errors in many others, if not all other variables. The facts of organizational behavior with interacting individual behaviors must be acknowledged in the testing.

Experimental methodology for testing mechanism designs can be posed as two key questions concerning the connection between observed performance and the principles used in the design. (1) Did the system do what it was supposed to do – proof of principle? (2) Did it do what it did for understandable (theoretical) reasons – design consistency? The design reflected a "stress test" in which a reasonable theory of adjustment would work against the purposes of the design[16]. Against that background the results demonstrate overwhelming support for the design success. The target alternative typically emerged thus establishing proof of principle (Result 1). Furthermore, the success of the design was consistent with the principles used in the design (design consistency). The outcomes were as predicted in the core (Result 2) using the dominance relation as a path (Result 3). The cases where the target did not emerge had multiple equilibria and the departures from target were predicted by the model. While the model itself was not constructed from a theory of individual choice a deeper examination of such tools is

[16] The group as a whole always had a majority preferred alternative. The research task was to use principles of group decision found in broad models to develop procedures that would cause the group to choose various alternatives (the target alternatives) that were ranked low in a (weak) majority rule order. Thus, the target alternatives were alternatives that one would expect would not be chosen by the group if left to its own.

invited by the fact that the behaviors can be understood in terms of a best response model from non-cooperative game theory (Result 4 and Result 5). Consistent with the goals of successful design post experiment questionnaires suggested that the committee members regarded the process as fair and did not perceive to have been manipulated. In every respect, the committee karate seems to have worked.

C. Limitations

With proof of principle and design consistency established for a simple case, natural questions exist about robustness of mechanism performance as complexity, participant experience and scale expand and as variations of the institutions are employed. The axiomatic theoretical structure suggests additional rules and procedures that can shape the dominance relation to give different veto powers to different groups and unrelated to their size of majoritarian status, see footnotes 8, 9 and 10. The axiomatic development alone (Section 5) suggests no limitations imposed by scale or size. Theory suggests organizational structure can accommodate more subcommittees. However, the theory does suggest limitations imposed by the environment. For example, if the status quo is unanimously preferred to all other options manipulation might not be possible at all while a unanimously disliked default can play a powerful role. Conflict plays a role and in its presence, theory suggests additional procedures, such as agendas, can be used as tools for manipulation, especially when majority rule cycles exist.[17]

Practical considerations are important. For actual mechanism creation the axioms must be operational in terms of observables and the actions used in models. For example, the veto power that shapes the abstract concept of a dominance relation in the Chairman experiments is made operational by the requirement that all proposals be seconded by the chairman. However, the chairman might not understand how to use the power or might find that its exercise creates complex reactions among others. Another example is the key axiomatic feature, Independence of Infeasible Alternatives. The axiom reflects institutions that function to keep the population in the dark with restricted communication between different parts of the organization so "informal organizations", coalitions, or strategic voting across different parts of the organization cannot

[17] McKelvey (1976), Schofield (1978) along with a well-crafted sequence of options implemented and fixed for consideration and voting can be used to move a group form any alternative to any other alternative.

emerge as vehicles to coordinate voting away from the manipulator's purpose. Operational implementation of the axiom could range from enforced rules to physical separation and (electronic) communication controls but success is not guaranteed. "Institutional failure" from a manipulation point of view is a possibility.

Manipulation requires supporting institutions. In some cases organization might be imposed by an administrative authority but other channels exist. For example, rules are often delegated to a small group, such as an agenda and rules committee, that is given substantial powers to implement procedures. Or, rules and organization can evolve over time and change in response to problems with conferred powers of manipulation being an unintended consequence. Neither manipulation nor its source need be obvious. Interestingly, the subjects in the experiment did not notice the manipulation that took place and only expressed a sense of unfairness due to obvious asymmetries of powers in the Chairman experiments (Observation 3). Rules that facilitate manipulation might be "sold" to an unsuspecting group as simplifications or rules to reduce decision costs of otherwise complex or challenging group decision problems.

Manipulation appears in many forms. Some are obvious (ballot stuffing, fraud, physical threats) while others are subtle and can take different, possibly hidden organizational forms. Here, manipulation was used as a formal framework for exploring how organizations work. Thus, while the idea of committee karate or manipulation carries with it a tone of Machiavellianism or antisocial theory, the opposite can be the case. On such matters, the science is morally neutral. The theories can be used for manipulation as well as for tools for protection against manipulation, hidden agendas, and poorly designed organizations or decision processes. This paper simply outlines how cooperative game theory, axiomatic social choice theory, voting theory and experimental methods can be added to the design toolbox.

January 1, 2017

REFERENCES

Aleman, Eduardo and Thomas Schwartz, (2006) "Presidential Vetoes in Latin American Constitutions" *Journal of Theoretical Politics,* 18(1):98-120.

Aleskerov, Fuad and Andrey Subochev (2013), "Modeling optimal social choice: matrix-vector representation of various solution concepts based on majority rule", *Journal of Global Optimization* 56(2):737-756.

Austen-Smith, David and Jeffrey S. Banks (1999), *Positive Political Theory I: Collective Preference,* Ann Arbor, University of Michigan Press.

Austen-Smith, David and Jeffrey S. Banks (2004) *Positive Political Theory II: Strategy and Structure,* Ann Arbor, University of Michigan Press.

Bergin, James and John Duggan (1999) "An Implementation – Theoretic Approach to Non-cooperative Foundations" *Journal of Economic Theory* 86(1):50-76.

Berl, Janet, Richard D. McKelvey, Peter C. Ordeshook, and Mark Winer (1976), "An Experimental Test of the Core in a Simple. N-person Cooperative, Non-sidepayment Game", *Journal of Conflict Resolution*, 20,453-79.

Benoit, Kenneth (2006). "Duverger's Law and the Study of Electoral Systems." *French Politics* 4(1):69-83.

Bottom, William P., Ronald A. King, Larry Handlin, and Gary J. Miller (2008). "Institutional Modifications of Majority Rule" *Handbook of Experimental Economics Results*, pp.857-879.

Brown, Donald J. (1973). "Acyclic Choice," Cowles Foundation discussion paper.

Chambers, Christopher P. and Alan D. Miller, (2013), "Measuring Legislative Boundaries". *Journal of Mathematical Social Sciences*, 66, 268-275.

Fiorina, Morris and Charles R. Plott, (1978). "Committee Decisions under Majority Rule: An Experimental Study", *American Political Science Review*, 72,575-598.

Fishburn, Peter C. (1973), *The Theory of Social Choice*. Princeton, N. J.: Princeton University Press.

Fischbacher, Urs (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics 10(2), 171-178.

Gomes, Armando and Philippe Jehiel (2005) "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies" *Journal of Political Economy* 113(3):626-667.

Grether, David, Marc Isaac and Charles R. Plott (2001), "The Allocation of Landing Rights by Unanimity among Competitors", *American Economic Review*, 71 (2), May 1981, pp.166-71. Reprinted in *Public Economics, Political Processes and Policy Applications*, Vol. 1, Charles R. Plott, ed. Edward Elgar, 2001.

Hammond, Peter J. (1997) "Game Forms versus Social Choice Rules as Models of Rights" *Social Choice Re-examined, Vol. II (IEA Conference Volume No. 117*) (London: Macmillan) 2(11):82–95.

Isaac, Mark, R. and Charles R. Plott (1978) "Cooperative Game Models of the Influence of the Closed Rule in Three Person, Majority Rule Committees: Theory and Experiment," *Game Theory and Political Science*, edited by P. C. Ordeshook. New York University Press.

Kagel, John H., Hankyoung Sung, and Eyal Winter (2010) "Veto Power in Committees: An Experimental Study", *Experimental Economics* 13: 167-188.

Konishi, Hideo and Debraj Ray (2003) "Coalition formation as a dynamic process" *Journal of Economic Theory* 110(1):1-41.

Koray, Semeh and Kemal Yildiz (2013) "Implementation via Code of Rights" NYU Working Paper.

Kormendi, Roger C. and Charles R. Plott (1982) "Committee Decisions under Alternative Procedural Rules: An Experimental Study Applying a New Nonmonetary Method of Preference 3:175-195.

Levine, Michael E and Charles R. Plott (1977) "Agenda Influence and Its Implications," *Virginia Law Review* 63(4):561-604.

McKelvey, Richard and Peter Ordeshook, (1984) "An Experimental Study of the Procedural Rules on Committee Behavior", *Journal of Politics*, 46, 1,182-205.

McKelvey, Richard D. (1976). "Intransitivities in multidimensional voting models and some implications for agenda control". *Journal of Economic Theory*. 12 (3): 472–482.

Moulin Herve (1998); *Axioms of Cooperative Decision Making*, Cambridge University Press.

Plott, Charles R. and Michael Levine (1978) "A Model of Agenda Influence on Committee Decisions," American *Economic Review* 68:146-160.

Plott, Charles R. (1994) "Market Architectures, Institutional Landscapes and Testbed Experiments." *Economic Theory* 4(1):3-10.

Plott, Charles R. (1976) "Axiomatic Social Choice Theory: An Overview and Interpretation." *American Journal of Political Science*, XX, 3 (August):511-596.

Ray, Debraj and Rajiv Vohra (2014) “Coalition Formation”, *Handbook of Game Theory*, Volume 4, ed. Peyton Young and Shmuel Zamir, North-Holland.

Riker, William H. (1986), *The Art of Political Manipulation*. New Haven: Yale University Press.

Rosenthal, Robert W. (1972) “Cooperative Games in Effectiveness Form”, *Journal of Economic “Theory*, 5:88-101.

Schwartz, Thomas (2008) “Parliamentary Procedure: Principal Forms and Political Effects”, *Public Choice*, 136:353-377

Schofield, N. (1978). "Instability of Simple Dynamic Games". *The Review of Economic Studies*. 45 (3): 575–594.

Shepsle Kenneth A., (1979) “Institutional Arrangements and equilibrium in Multidimensional Voting Models, *American Journal of Political Science*, 23:27-59.

Shepsle, Kenneth A. and B.R. Weingast, (1984) “Uncovered Sets and Sophisticated Voting Outcomes with Implications for Agenda Institutions” *American Journal of Political Science*, 28:49-7.

Wilson, Rick, (2008a) “Endogenous Properties of equilibrium and Disequilibrium in Spatial Competition Games”, *Handbook of Experimental Economics Results*, pp.872-879.

Wilson, Rick, (2008b) “Structure Induced Equilibrium in Spatial Committee Games” *Handbook of Experimental Economics Results*, pp.880-897.

Wilson, Robert B., (1972) “The Game Theoretic Structure of Arrow’s General Possibility Theorem”, *Journal of Economic Theory*, 5:14-20.

Appendix 1: Software, Procedures and Instructions

A. An outline of committee procedures

A history of past proposals and their results is maintained for all committee members to review as deliberations take place.

Motion on the floor: The process starts with the default option as the motion on the floor.

Recognition: When stage opens participants have 10 seconds to choose an option and seek recognition. A random selection is made from those who seek recognition.

Proposed Amendments: The participants recognized submit their selected alternative as an amendment to the motion on the floor. If the proposal is seconded it is presented for a vote. The proposed amendment is highlighted for all to see and understand that it is proposed.

Seconds to proposed amendments: Under the Condorcet mechanism, seconding of the motion can be done by any committee member other than the member that made the motion. Under the Chairman mechanism, only the chairman can second motions. The floor remains open for 5 seconds or until seconded. If there is no second, the proposed motion fails. The previous motion on the floor remains as the motion on the floor. If the proposed amendment is not seconded, the system returns to the Recognition Stage.

Voting: If the amendment is seconded, the screen changes color (orange) to indicate that a vote is to take place between the proposed amendment and the motion on the floor. Voting is open until all have voted. Do you want x to become the new motion on the floor and replace y?

If the amendment fails, the original motion on the floor remains. If the amendment passes it becomes the new motion on the floor. The system returns to the Recognition Stage.

Ending debate: During the recognition stage, a motion to end debate can be offered. If recognized, it must also receive a second, which can be done by anyone other than the person making the motion. If the motion to end debate is seconded, then the screen asks "We now vote on the motion on the floor. Would you like to end the amendment process and accept the motion on the floor as the committee decision?" Subjects must choose yes or no. Pass is determined by majority. If no, the system returns to the Recognition Stage.

B. Instructions

Purpose and Payoffs

You will participate in an experiment on group decision making. You will be a member of a committee that must choose one letter from a set of letters. Only one of the letters will be chosen and the payment you receive for participation depends *entirely* upon which letter it is. People's preferences for the letters may differ, so the letters you prefer may not be preferred by others.

Preferences

On your screen, the letters are ordered from your most preferred to your least preferred.

Here, L is your most preferred letter, I is your second most preferred letter, and so on. Below each letter is your payoff in experimental currency should that letter be chosen as the committee's decision. Thus, if L is chosen as the committee's decision, you will get 400 in experimental currency; if K is chosen, you will get 150. Your exchange rate from experimental currency to dollars is located in the top left. This member's exchange rate is 100, so a payoff of 400 would amount to $4.

Other members may differ in their orderings, payoffs, and exchange rates. For example:

Her exchange rate is 25 and her most preferred alternative K has a payoff of 75, which is worth $3.

Procedure

In order for the committee to choose a letter, it must follow some rules of order. Initially,

- The committee is split into subcommittees
- Each subcommittee is assigned a subset of the letters from which they must choose a recommendation. All recommendations go to a final committee that chooses among the recommendations as the committee's decision via majority rule.
- It is this final decision on which your payoff depends, not the subcommittee's decision.
- One letter O is designated as the initial "motion on the floor" (which can be considered the subcommittee's tentative decision).

The subset of letters from which your subcommittee can choose is depicted via outline. Specifically, the letters your subcommittee can choose between in the example above are L, I, B, E, A, F, K and O. The motion on the floor is O and the subcommittee will have an opportunity to change the motion on the floor through an amendment process.

Amendment Process

Initially, the floor is open for proposals and all members may propose an amendment to the motion on the floor by selecting an alternative letter. This proposal stage will occur multiple times. After the timer expires, only one proposed amendment will be chosen for further consideration. So if multiple people submit a proposal that lowers the chance that your proposal is considered by the subcommittee at this stage. But suppose that you are the only person submitting a proposal – then your proposal will definitely be considered. During the seconding stage, the proposed letter will be bolded and all other letters deemphasized. If you were the proposer, the box is **blue**. Otherwise, it will simply be **bolded**. Anyone except the proposer may

second the proposal. If not seconded within the time limit, the process returns to the recognition stage. If seconded, all members must choose whether or not the proposed amendment should become the motion on the floor. If not, the existing motion on the floor remains.

History

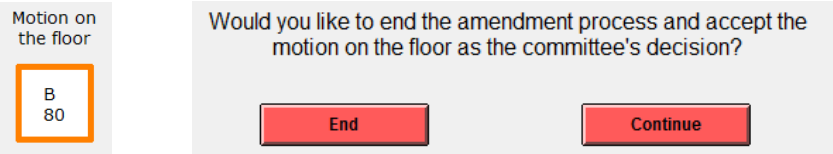
Throughout the period, a history of all proposals made and their votes is displayed in the bottom left.

Motion on the Floor	Proposed Amendment	Proposer	Second		Member 1	Member 2	Member 3	Member 4	Member 5
G B	B A	5 2	2 4		G B	B A	G B	B A	B B

As other members vote, their votes are displayed in real time. After each vote on a proposed alternative, the process returns to the proposal stage. In the example above, the majority of members voted for the proposed amendment B over the initial motion on the floor G, so the motion on the floor has changed to B. If the timer expires and no member has proposed an amendment, the next proposal will be immediately considered.

Ending a Period

During the proposal stage, members may also propose to end the amendment process by selecting the motion on the floor for consideration as the subcommittee's recommendation. If the proposal to end is seconded, all members must choose to either accept the motion on the floor or continue the amendment process.



However, remember that while your subcommittee was choosing between a subset of letters, the other subcommittee was choosing between the other subset of letters. After both committees have recommended a letter, the committee as a whole will vote between the two recommendations. Again, your payoff depends *entirely* upon the decision by the committee as a whole.

Ties in the Committee as a Whole

In the event of a tie, the voting process will repeat itself, but a timer will begin to count down during which members may revote. If the timer expires and the committee still has not reached a decision, O will be the committee's decision. If a decision is reached, everyone is paid accordingly.

Multiple Periods

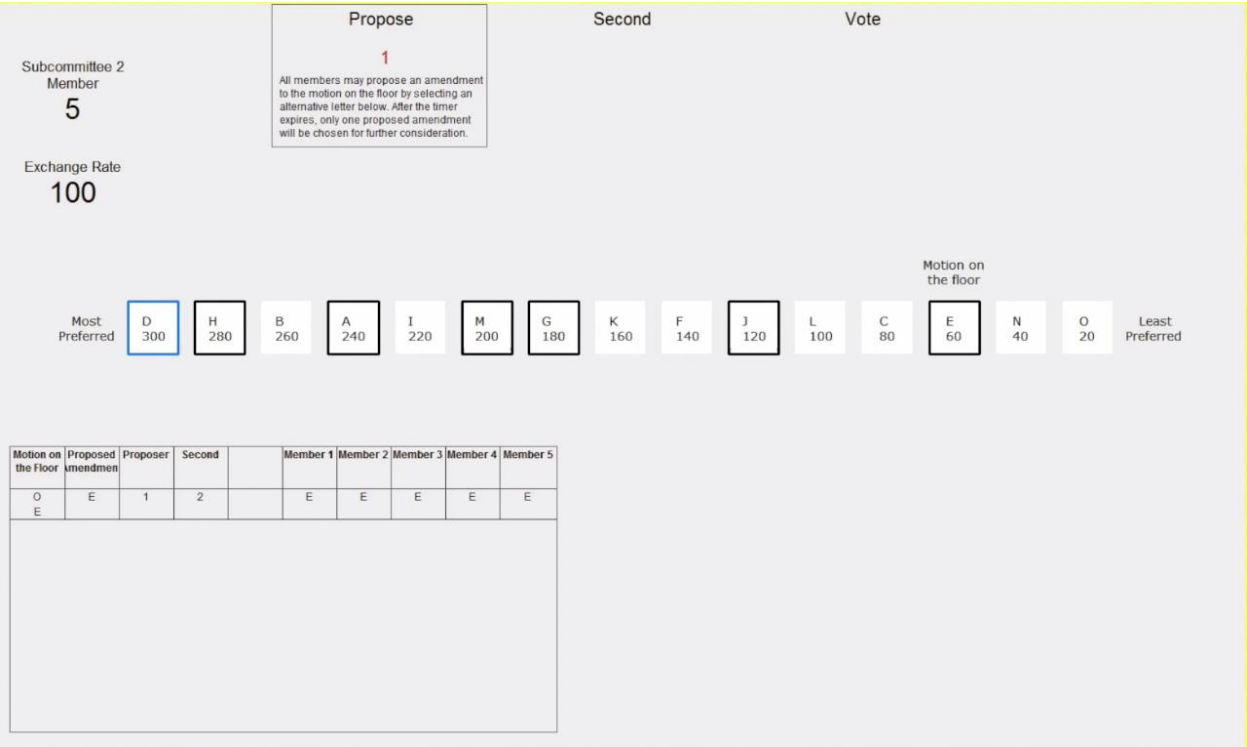
The entire process will be repeated over multiple periods. In each period, you will have different preferences, but you will retain the same member number and exchange rate. You will be paid the sum of the profits you make in each period.

Phase 2 (Phase 1 was the previous phase): Introduction of a Convener (Committee Chairman)

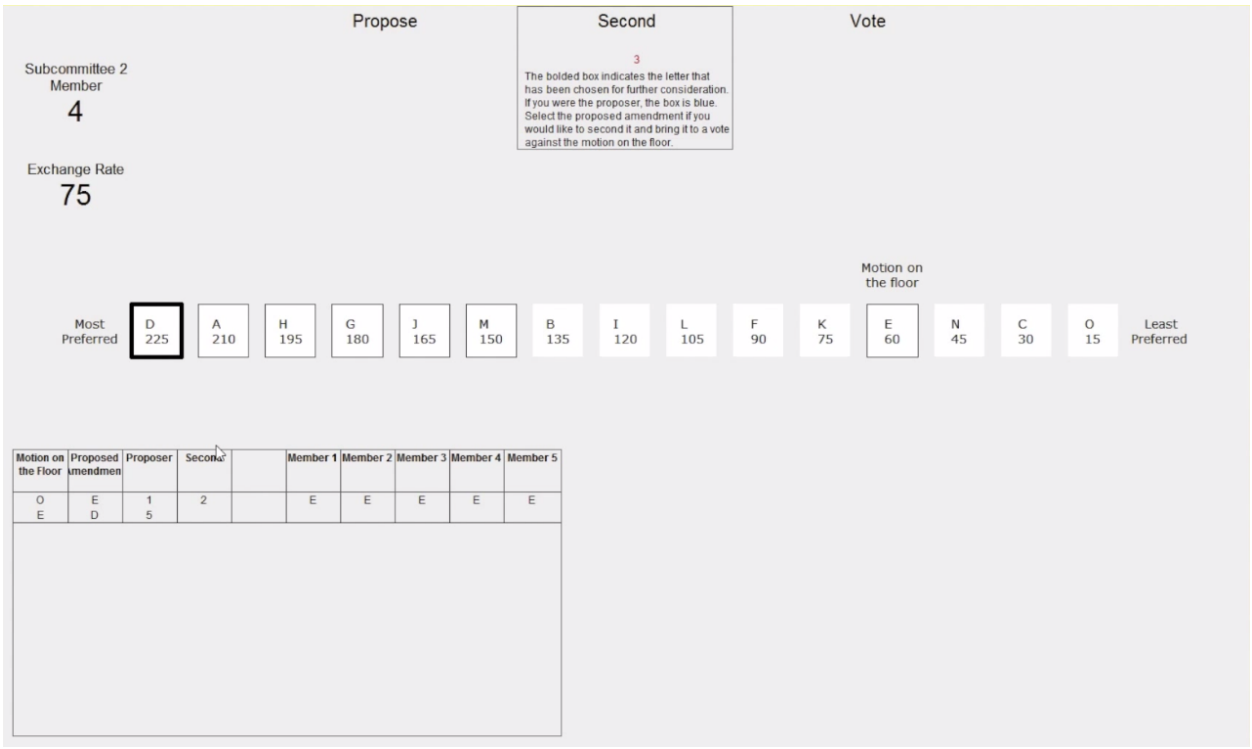
Similar to the previous phase, any member can propose an amendment. However, one member will be designated as a convener, who has the sole power to second a proposed amendment to the motion on the floor (except their own, which must be seconded by someone else). That is, you will only vote on proposed amendments seconded by the convener. Proposals to end and accept the motion on the floor differ in that any member may second a proposal to end (except the proposer).

In the top left of the screen, you will be notified whether or not you are the convener.

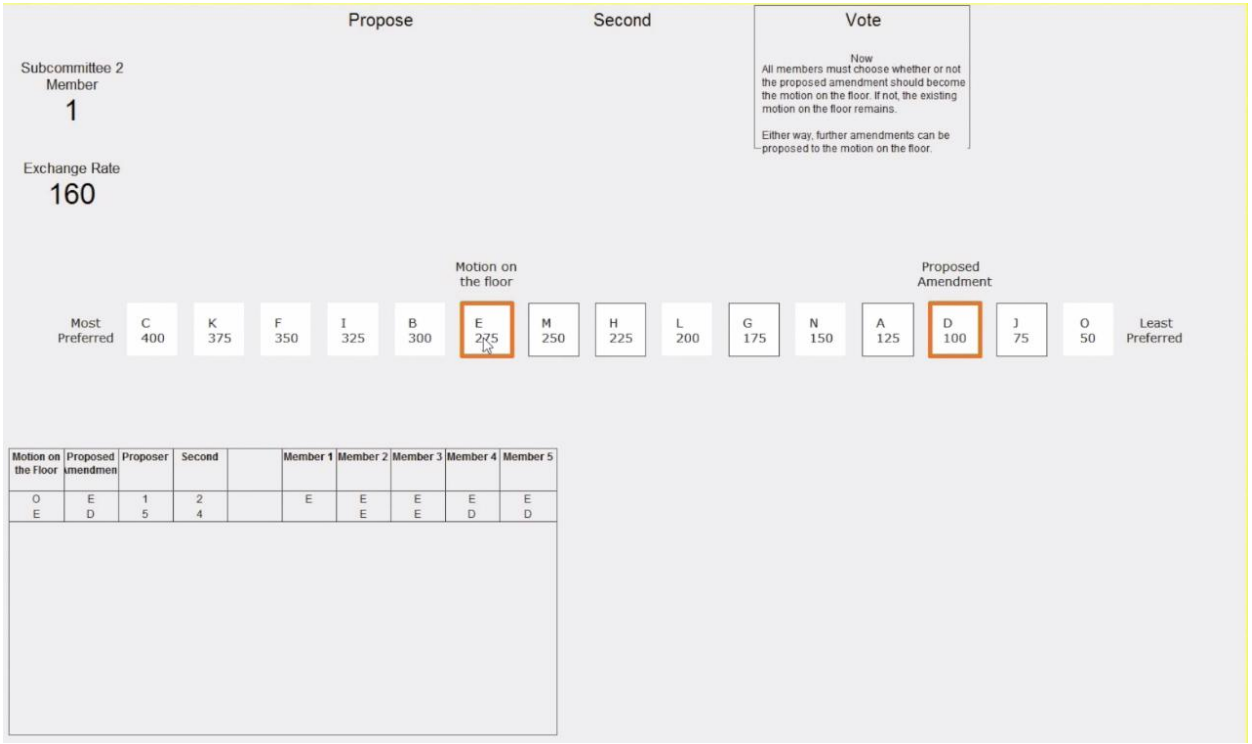
Appendix 2: Screenshots



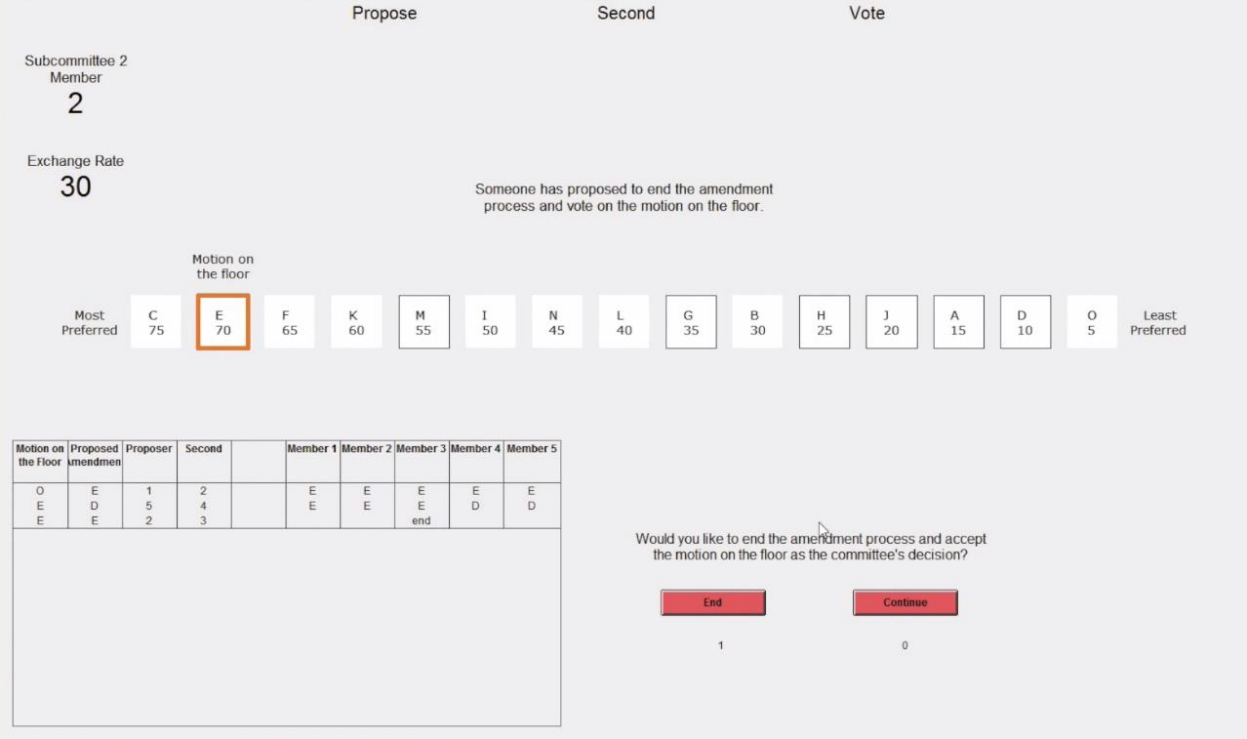
Propose – Subjects may choose (with their mouse) to seek recognition by selecting an option from among the subset of feasible options (as shown in a thick outline). A timer counts down until it either reaches zero or everyone has sought recognition. From those seeking recognition, the computer chooses one at random, and that subject's selection becomes the proposal. If no one seeks recognition within that timeframe, the program waits and accepts the first selection as the proposal.



Second – The proposed motion is bolded (and the other options deemphasized). If a subject that was not the initial proposer chooses to second the option, it proceeds to the next stage of voting. Otherwise, when the timer reaches zero, it returns to the previous *Propose* stage.



Vote – The two options that are being voted on are highlighted in orange. All subjects must choose one of the two options to vote for before the program proceeds. The winning option becomes the motion on the floor. The program then proceeds back to the *Propose* stage.



Motion to End Debate – If the current motion on the floor is proposed, then there is a motion to end debate. All subjects must vote to either end discussion and choose the motion on the floor as the subcommittee's choice or continue discussion (hence returning to the *Propose* stage).

Appendix 3: Preference Inducement

Preference Types and Preference Inducement

A preference type is an ordering of the 15 alternatives. There are ten different types, contained in Table 1, each of which is assigned to a subject and induced as a preference in a given period. Figure 3 is a two dimensional spatial configuration of the ten types, with the numbered boxes as the location of the maximum for each type.

For example, the preference ordering of the alternatives for Type 1 is in the first row of Table 1 and has alternative L as the most preferred, G as the second most preferred, D as the third, etc. with the least preferred alternative for Type 1 being the alternative O. Type 1 is also illustrated in Figure 3, where the box containing the number 1 is the spatial location of the optimum for Type 1 and the further the letter is from the optimum the lower is the level of preference. As shown, alternative L is the closest to the Type 1 optimum and the location of alternatives G and D are further, indicating that they are less preferred. All individuals have alternative O ranked lowest and thus alternative O is not shown in Figure 3.

Table 3, the Subject Value Schedule, contains for each subject the monetary value of the fifteen alternatives given as a function of the rank of preference. For example, if the committee's decision is a subject's most preferred alternative, that subject receives a payoff of $2.50. If the decision is the subject's second most preferred, the payoff is $2.33 and so on. Each step has a difference of 16.6 cents. For example, if the committee's decision was L, the subject with Type 1 would receive $2.50 while Type 3 would receive $1.83 and Type 6 would receive $0.33.

The Assignment of Types to Subjects for Experimental Design and Testbeds

The experimental design, outlined in Section 6, calls for two subcommittees. Each of the two subcommittees is assigned a subset of the fifteen alternatives. The status quo, alternative O, is available to both subcommittees as the default of no decision. In the experiments studied here, each subcommittee considered seven alternatives plus alternative O. Each subcommittee has five members and no one serves on both subcommittees. Thus, the ten subjects are partitioned into two subcommittees of five members each.

Each of the ten types contained in Table 1 is assigned to one of the two subcommittees and will become the preference ordering of a committee member. The committee's mechanism – simple Roberts Rules or Chairman – together with the five preference types and seven alternatives, are sufficient to determine the dominance relation discussed in the theory of Section 6.

Rotations and Permutations

Two procedures were employed to avoid the possibility that the assignment of types to subjects carried information about the environment and the possible preferences of other subjects. First, the letter assignments were permuted each period. That is, the letter assigned to the original alternative A became some other letter, such as F. Second, whenever the same environment was repeated, the types were rotated among the same subjects. For example, subjects 1,2,3,4,5 would be assigned types 3,4,6,9,10 in one repetition and 4,6,9,10,3 in another.