Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Mathematical Economics xxx (2006) xxx–xxx
JOURNAL OF Mathematical ECONOMICS
www.elsevier.com/locate/jmateco

From decision problems to dethroned dictators☆

Jason Kronewetter, Donald G. Saari*

Department of Mathematics, Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA 92697, United States

Received 17 January 2006; received in revised form 09 June 2006; accepted 10 June 2006

Dedicated to Roko Aliprantis on the occasion of his 60th Birthday.

Abstract

Economic models as well as aggregation and decision problems with "holes" in the domain can be difficult to analyze because, unexpectedly, they are related to Arrow's Impossibility Theorem: embedded within the model may be "topological dictators." But, just as it is possible to remove the negative impact of Arrow's dictator by recognizing that the problem is caused by not using crucial, available information (about voter preferences), the obstacles confronting these economic decision problems can be removed by identifying what kind of available information is not being used.

Keywords: Social choice; Economic decision problems; Dictators; Averaging

1. Introduction

Difficulties can occur when decision problems are encumbered with obstacles. Based on various criteria, for instance, we may wish to locate a plant in a square-shaped region. Complexities arise if the region's interior has "holes" where the plant cannot be located; e.g., a lake may be in the middle and private land might occupy another spot. A simpler example has a group planning to picnic on the beach of a circular lake: how should they pick the spot? While these problems appear to be simple, complications can occur because, unexpectedly, the holes subtly connect the decision problem with Arrow's (1962) seminal impossibility result. Namely, even seemingly simple constraints can force admissible decision rules to involve "topological dictators"—a term

☆ The research of D. Saari was supported by an NSF Grant DMI-0233798.

* Corresponding author.
E-mail address: dsaari@uci.edu (D.G. Saari).

0304-4068/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.jmateco.2006.06.004

that we describe technically and intuitively. More generally, as decision rules are aggregation rules, we must anticipate that similar negative consequences also plague general aggregation and statistical methods as well as social science models that involve aggregation: they do.

Rather than establishing still another "dictator" assertion, our intent is to offer positive contributions; e.g., our main goals include describing, in an assessable manner, why these negative results occur, why this serious situation need not be as negative as the terms suggest and are typically interpreted, and, of particular importance, how to sidestep these difficulties. While the implications of our results cross a variety of decision and social science issues, our arguments derive from social choice. In terms of choice theory, our contributions include the following:

1. Chichilnisky (1982) introduced the "topological dictator" in her seminal paper describing a topological extension of Arrow's result. (Baigent (2006) reviews this area.) Her main thrust is to establish the impossibility of finding a societal von Neumann–Morgenstein utility function. But as we show, a much richer, wider assortment of issues are also stymied by "topological dictators" and other restrictions. Beyond group decisions these topics involve multicriteria decision problems and even models with multivariable mappings. To include a wide variety of extensions, we relax the traditional assumptions made in choice theory.
2. Confusion accompanies the interpretation of a "topological dictator;" e.g., comments at conferences and in the literature suggest that this term requires one person to dictate, or play a dominant role in selecting the societal outcome. Extending the arguments of (Saari, 1993, 1997) we show that this need not be the case. After explaining why the term is seriously misleading, we replace it with a more meaningful choice that captures the initial intent.
3. Our main goals are to develop intuition (with more assessable proofs), to explain why these problems arise, and to indicate how to eliminate them. To do so, rather than presenting our more general results, we emphasize the simpler two agent setting. Our explanation about the source of "topological dictators" mimics an explanation for the Arrovian dictator (Saari, 2001). The idea is that problems must be expected whenever the admissible decision rules are forced to ignore crucial information: information that, by explicit assumption, is intended to be used. This explanation also identifies how positive conclusions can be obtained; just modify the conditions so that they allow the intended information to be used. We develop one approach, but there are many others.

2. Obstacles in finding solutions

In an insightful way, Chichilnisky converted certain problems from economics into choice problems of selecting points on an n-dimensional sphere. Think of this in terms of gradients. At a particular point, the level set of a utility function can be locally described by its gradient where only the direction, not the length, is of interest. Selecting a gradient direction, then, can be equated with selecting a point on the sphere (denoted by S^n) given by $x_1^2 + \cdots + x_{n+1}^2 = 1$. As far as we know, Chichilnisky emphasized only spheres where her primary interest continued to be motivated by this class of economic interpretations and extensions. But, as we show, there is a very rich class of other economic, decision, and social science problems that suffer similar difficulties.

To develop intuition about what kinds of problems suffer these restrictions, notice that all of the following examples have "holes" in connected regions. To characterize them, think of the region as an elastic material that can shrink to a point. Before the shrinking process, fill all holes with a solid, inflexible material, say iron. If, without tearing, the flexible material can slip around the solid material to shrink into a point, the region is *contractible*; if it cannot, it is not contractible. All of the following examples involve regions that are *not contractible*.

Locating a plant: The "holes" created by the lake and private property as described above create obstacles that prohibit the flexible material from sliding around to contract the square region into a point.

Beach party: Suppose a group plans to have a party on the shore of a lake. While the shore is not a perfect circle, it can be continuously pushed and pulled into a circle denoted by S^1. But a circle, say an elastic ring on a finger, cannot be shrunk to a point without tearing, so S^1 is not contractible.

To enhance the beach story, suppose the party is at a resort with activities 24 h a day. This means that the choice of where to have the party depends on when the party is to occur; e.g., some people may wish to gather at 2 p.m. where they can sunbathe, others may wish to meet at 7 p.m. for dinner, while still others may wish to convene at 1 a.m. to gamble or dance. Thus, each potential beach location is accompanied by a circle of possible times (described on a 24 h basis). Moving the "time circle" through all locations on the circular beach traces out a "torus" – denoted by $T^2 = S^1 \times S^1$ – for the choice domain; it resembles the surface of a donut. The "hole" to be filled with iron is the actual donut. Clearly, the skin of the donut cannot be shrunk to a point without tearing.

To add further complications, suppose that along with the chosen spot and time, we must select an appropriate orientation of a table, which defines a direction given by a point on a circle. If the orientation can be treated as being the same after rotating the table 180°, then opposite points on a circle are identified; this defines what is called a *real projective 1-space* $\mathcal{R}P^1$. The complete description leads to the domain $\mathcal{V} = T^2 \times \mathcal{R}P^1$.

Satellite location: Envision several nations negotiating over the location of a satellite that will remain in a fixed position over the Earth. The location, then, is a point on a sphere S^2: the surface of a ball. The "hole" is the actual ball, so, clearly, S^2 is not contractible. To add complications, suppose that the functional advantage of the satellite depends on its position *and* its orientation. (With a communication satellite, for instance, different orientations may provide varying advantages to different countries.) As the orientation also is determined by a point on the sphere, the domain becomes the non-contractible space $\mathcal{V} = S^2 \times S^2$. If only the orientation of the satellite, not the direction of a particular point, is of interest, then points on opposite sides of the sphere are identified: the orientation is given by the real projective sphere $\mathcal{R}P^2$. This space is described by non-contractible $\mathcal{V} = S^2 \times \mathcal{R}P^2$.

2.1. General representation

Each example defines a region where $\mathcal{V}$ represents the designation of this choice domain. As each of the n agents has a particular favored point, the decision problem is to use these inputs to determine a group outcome. In mathematical terms, the decision rule is a mapping F

$$F : \mathcal{V} \times \cdots \times \mathcal{V} \to \mathcal{V}. \tag{1}$$

If F models a standard multicriteria problem, or even a multivariable mapping, then replace the agents with criteria or variables. Each criterion – maybe taxes, availability of a labor supply, etc. – identifies an ideal location. The choice rule determines the optimal location by aggregating the choices based on the various criteria. The first step is to impose natural conditions on F. To allow our results to hold for even larger classes of problems, in Section 6 we indicate how to relax the assumptions.

2.1.1. Assumptions about F

The following assumptions involve the mapping F.

Continuity: Let F be *continuous*. After all, if F is not continuous, then an infinitesimal change in the inputs – determined by voter preferences or criteria – could cause an undesired jump in the outcome.

Unanimity: A natural assumption (that can be generalized) is that F respects *unanimity*: when there is total agreement, this is the decision. Namely, *for any* $\mathbf{p} \in \mathcal{V}$,

$$F(\mathbf{p}, \ldots, \mathbf{p}) = \mathbf{p}. \tag{2}$$

The above two assumptions do not suffice for our purposes as they allow functions with undesirable outcomes. With beach party example, for instance, if angle θ_j describes the jth person's location choice, then, while $F(\theta_1, \theta_2) = 6\theta_1 - 5\theta_2$ satisfies the continuity and unanimity condition, it generates unreasonable outcomes. To explain, suppose that the second person wants the party held directly to the east, $\theta_2 = 0$, while the first wants it slightly more to the north, $\theta_1 = \pi/6$ or 30°; with this F the outcome is directly to the west, $F(\pi/6, 0) = \pi$. The next condition avoids such pathology.

Pareto condition: On a circle, a line segment, or circles connected with line segments, and with two agents (variables, etc.), when there is a shortest distance between two points, the *Pareto condition* requires the outcome to be in this region. With the beach party example, the Pareto condition requires the outcome to be in the $[0, \pi/6]$ region.

Our results extend to several other topics because, with the possible exception of the unanimity condition, the above framework already applies to a variety of aggregation settings ranging from statistics to allocation problems. To be even more inclusive, we describe in Section 6 how the unanimity[1] and Pareto conditions can be significantly relaxed.

2.1.2. The domain

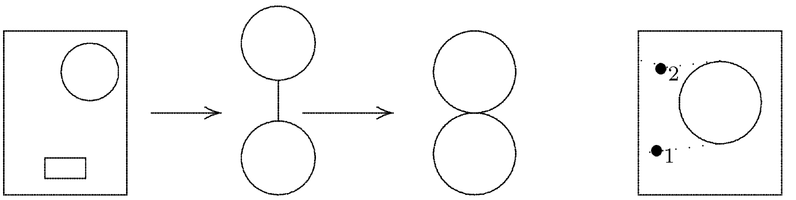
Important for our arguments is the ability to contract the domain $\mathcal{V}$ into objects that are easier to analyze.

Contractible to a point: To determine whether $\mathcal{V}$ is contractible to a point, the intuitive approach treats all holes as being filled with a solid material, and then determines whether it is possible, without tearing, to continuously shrink the region to a point $x_0 \in \mathcal{V}$. In mathematical terms, *region $\mathcal{V}$ is contractible to subregion $\mathcal{V}_1 \subset \mathcal{V}$* if there is a continuous function:

$$\begin{aligned} &G(\mathbf{x}, t) : \mathcal{V} \times [0, 1] \to \mathcal{V}, \quad \text{where } G(\mathbf{x}, 1) = \mathbf{x} \text{ and} \\ &\text{the image of } G(-, 0) = \mathcal{V}_1, \quad \text{where } G(\mathbf{x}, 0) = \mathbf{x} \text{ for all } \mathbf{x} \in \mathcal{V}_1. \end{aligned} \tag{3}$$

So, region $\mathcal{V}$ *is contractible to point* $\mathbf{x}_0$ when $\mathcal{V}_1 = \mathbf{x}_0$. In this definition, treat the t values as describing the different stages of "shrinkage:" at level $t = 1$ there is no shrinkage as the map is the identity map, but at $t = 0$ the whole region collapses to $\mathcal{V}_1$, which may be a point $\mathbf{x}_0$. A square, for instance, is contractible to any point in its interior; e.g., for the square $[-1, 1] \times [-1, 1]$, $G(\mathbf{x}, t) = t\,\mathbf{x}$ is the identity map when $t = 1$, and it is the mapping with a single image point, the origin $\mathbf{0}$, when $t = 0$. While a square is contractible, a square with an interior point removed is *not* contractible because the region cannot slide around the hole without tearing: this rupture

[1] We just need $F(\mathbf{p}, \ldots, \mathbf{p}) = g(\mathbf{p})$ where $g : \mathcal{V} \to \mathcal{V}$ is a continuous one-to-one and onto function, e.g., $g(\mathbf{p}) = -\mathbf{p}$ or $\mathbf{p} + (\pi/3)$. Even the diagonal can be replaced with other curves, e.g., $F(\mathbf{p}, -\mathbf{p}) = g(\mathbf{p})$.



(a) (b)

Fig. 1. Reduction. (a) Homotopic equivalents and (b) generalized Pareto.

corresponds to a discontinuity for $G(\mathbf{x}, t)$. For all of the problems described above, the domains are *not contractible to a point*.

Fig. 1a, for instance, shows how the problem of locating a plant on a rectangular region with a lake and private property is contractible not to a point, but to two circles connected by a cord, and then to two circles touching at a point, which is called the "wedge product of the circles." As shown later, these kinds of regions restrict the choice of decision rules.

Generalized Pareto condition: If $\mathcal{V}$ is contractible to a line segment, circle, or circles connected by line segments, as in Fig. 1a, the *generalized Pareto condition* for F is where *the contracted image of* $\mathbf{p}_1, \mathbf{p}_2$, *and* $F(\mathbf{p}_1, \mathbf{p}_2)$ *satisfy the Pareto condition.* Namely, if $G(\mathbf{x}, t)$ is the contraction mapping of $\mathcal{V}$ to the reduced $\mathcal{V}_1$, then $G(F(\mathbf{p}_1, \mathbf{p}_2), 0)$ must satisfy the Pareto condition with respect to $G(\mathbf{p}_1, 0)$ and $G(\mathbf{p}_2, 0)$.

This generalized Pareto condition admits a surprisingly wide class of functions. If, for instance, $\mathcal{V}$ is contractible to a point, then any smooth F satisfies the generalized Pareto condition. The flexibility is a direct consequence of the many ways a region can be contracted onto another; each way defines a different class of Pareto conditions. So, if a function F fails to satisfy the Pareto condition with one choice of a contraction, it might with another. To illustrate, let the labeled bullets in Fig. 1b represent the two $\mathbf{p}_j$'s. One way to contract this rectangular region onto the circle is along lines toward the center of the circle. Another way, as indicated in the figure, pulls the upper and lower bullets, respectively, along the dotted lines near the top and bottom of the circle. If the $F(\mathbf{p}_1, \mathbf{p}_2)$ image is anywhere within the region defined by the dotted line, then, at least for these points, F satisfies the generalized Pareto condition.

2.2. *An explanation*

It is worth explaining the importance of contracting $\mathcal{V}$ to a subregion. Remember, we want to establish that holes in the domain $\mathcal{V}$ restrict the choices of admissible Eq. (1) mappings. A merit of using contractions is that a single theorem establishes our conclusions for a wide variety of settings. If, for example, a result holds for a domain $\mathcal{V}$ that can be contracted to a circle, then the same conclusion applies to an annulus, a rectangular region with a hole in the interior, a three dimensional ring, and so forth.

The second purpose is to simplify the analysis by replacing $\mathcal{V}$ with a simpler geometric object. After all, if $F : \mathcal{V} \times \cdots \times \mathcal{V} \to \mathcal{V}$ has desired properties, and if G contracts $\mathcal{V}$ into a mathematically simpler $\mathcal{V}_1$, then, by continuity and for each t, the mapping

$$F^*(\mathbf{p}_1, \ldots, \mathbf{p}_n, t) = G[F(G(\mathbf{p}_1, t), \ldots, G(\mathbf{p}_n, t)), t] \tag{4}$$

(where domain and range points are being contracted) also enjoys these properties. Thus, to prove that F suffers difficulties, we only need to show that when $t = 0$, which corresponds to a F^* mapping from the simpler $\mathcal{V}_1 \times \cdots \times \mathcal{V}_1$ to $\mathcal{V}_1$, has problems. Thus the geometric complexities of the original domain can be avoided by analyzing what occurs with much simpler geometric structures.

2.3. Homotopic

The concept of a choice function being *homotopic to a dictator* is closely related to the notion of a region being contractible to a circle or a point. Indeed, the F^* mapping in Eq. (4) is a homotopy.

To motivate the definition of a "homotopy," consider the modeling approach used to analyze a complicated model. A first cut at a problem is to use a simpler model obtained by ignoring certain variables. Doing so carries a tacit sense of a continuum of models through which we can continuously move from the original setting to the simplified one. The "continuous transition" is important because a discontinuity represents a disruption in the relationship between the original and simplified models. In particular, a discontinuity may indicate that a particular variable cannot be ignored as it is essential. In mathematical terms, this approach is captured by "homotopy."

More precisely, choice rules F_0, F_1 are homotopic if there exists a continuous[2]

$$\begin{aligned} &F(\mathbf{p}_1, \ldots, \mathbf{p}_n, t) : \mathcal{V} \times \cdots \times \mathcal{V} \times [0, 1] \to \mathcal{V}, \quad \text{where} \\ &F(\mathbf{p}_1, \ldots, \mathbf{p}_n, 0) = F_0(\mathbf{p}_1, \ldots, \mathbf{p}_n), \quad \text{while} \\ &F(\mathbf{p}_1, \ldots, \mathbf{p}_n, 1) = F_1(\mathbf{p}_1, \ldots, \mathbf{p}_n). \end{aligned} \tag{5}$$

It is convenient to view $F(\mathbf{p}_1, \ldots, \mathbf{p}_n, t)$ as describing the tth rule in a continuum of choices. Namely, similar to the "shrinkage" interpretation offered for Eq. (3), informally treat the t values in $F(\mathbf{p}_1, \ldots, \mathbf{p}_n, t)$ as describing changes in the level of empowerment of certain agents (or variables) in the choice rule during the transition between the rules F_1 and F_0. The F^* of Eq. (4), then, is a homotopy relating the original mapping F to a mapping effectively defined on the domain $\mathcal{V}_1$.

As another example, when locating a plant on the line interval [0, 1], let x_j be the choice of the jth agent and suppose that agents 1 and 2 are more knowledgeable, respectively, about conditions near the 0 and the 1 endpoint. Such a choice rule might be the weighted average

$$F_1(x_1, x_2, \ldots, x_n) = \frac{(2 - x_1)x_1 + (1 + x_2)x_2 + 0.1(x_3 + \cdots + x_n)}{3 + x_2 - x_1 + 0.1(n - 2)}. \tag{6}$$

A first, crude approximation, is the simplified model $F_0 = ((x_1 + x_2)/2)$, which selects the average choice of the two dominant agents. The above comments suggest that to partially justify the F_0 approximation, we must show that F_0 is homotopic to F_1. One choice to do so is

$$F(x_1, \ldots, x_n, t) = \frac{(1 - t)(x_1 + x_2) + t[(2 - x_1)x_1 + (1 + x_2)x_2 + 0.1(x_3 + \cdots + x_n)]}{t(3 + x_2 - x_1 + 0.1(n - 2)) + 2(1 - t)};$$

this homotopy describes a continuous decrease in the influence of the values of $x_3, \ldots, x_n$.

Dictators and topological dictators: A *dictator* is a choice rule where the outcome always agrees with a specified agent's (the dictator) preferences. Namely, *decision function F designates*

[2] While the definition does not require it, one might require each rule in the transition to satisfy certain conditions. For instance, one could require all rules for each t value to satisfy unanimity, etc.

agent j as the dictator if for all $\mathbf{p}_i \in \mathcal{V},\ i = 1, \dots, n,$

$$F(\mathbf{p}_1, \dots, \mathbf{p}_j, \dots, \mathbf{p}_n) = \mathbf{p}_j. \tag{7}$$

If F is a multicriteria problem where criteria replace agents, then criterion j is the "dictator" if the outcome always coincides with the choice identified by this criterion. In a multivariable problem, variable x_j is the "dictator" if $F(x_1, \dots, x_j, \dots, x_n) \equiv x_j$. More generally, when generalizing the unanimity assumption from x to $g(x)$, much of what we say holds by replacing the "dictator" to where F is a function of a single variable; e.g., $F(x_1, \dots, x_j, \dots, x_n) \equiv g(x_j)$.

Function F is *homotopic to a dictator* if F can be continuously deformed into a dictatorial decision rule. To illustrate with the beach party, suppose the first person selects the location for the picnic, while a second person makes modifications—perhaps moving the choice slightly to the left or right to avoid sitting directly in the sun. By continuously diminishing to zero the second person's right to modify the final position, we arrive at a dictatorial setting. Namely, if F_1 is homotopic to a dictator, then it is possible to continuously deform F_1 into a dictatorial one by continuously diminishing to zero the power and influence of all other agents. In a topological sense, then, F_1 can be approximated by the specified dictator's choice.

2.4. Domain structure affecting admissible decision functions

As asserted next, should the domain have holes, then expect the admissible rules to be homotopic to a dictator. Beyond showing that "holes" in domains restrict the selection of a decision and allocation rules, we use Theorem 1 to explain (Section 6) why these restrictions occur and how to weaken the basic assumptions. While our actual result is more general (involving more variables, more complicated domains, etc.), the simpler Theorem 1 suffices for our principal objective, which is to explain that there are difficulties and how to circumvent them. Also, Theorem 1 has a simpler, transparent proof using geometry, rather than topological arguments that may not appeal to the intuition of readers.

Theorem 1. *Let* $F(\mathbf{p}_1, \mathbf{p}_2)$ *be a continuous function satisfying unanimity*, Eq. (1), *and the generalized Pareto condition. Suppose* $\mathcal{V}$ *is homotopic to a connected, one-dimensional chain of k circles where the jth circle meets the* $(j-1)$*th and the* $(j+1)$*th each in one point*, $1 < j < k$, *while the first circle meets only the second circle at a point, and the kth circle meets only the* $(k-1)$*th circle. Function F is homotopic to a specific dictator.*

As stated, our more general results will appear elsewhere. Chichilnisky (1982) has a similar conclusion when $\mathcal{V}$ is S^n. While useful, her specialized setting of spheres fails to address the problems that motivate this paper (e.g., the location of a plant). A pragmatic advance of Theorem 1, then, is to address a wider range of practical decision and social science problems. More important than extending this theorem is to understand what causes the restriction on the choice rules (functions, etc.), and to use this information to remove these obstacles. In this spirit, our proof (Section 6) uses the standard concept of level sets of a mapping to identify why these problems occur.

3. Dethroning "topological dictators"

Asserting that a rule is homotopic to a dictator carries the distinct image of a specified agent who either is a dictator, or plays a dominant role in making the decision. However, one of us (Saari, 1993, 1997) introduced the beach party example as a way to show how easy it is to create rules that are homotopic to a dictator where, in fact, other agents play significant and even dominant roles in determining the decision over large regions of the domain.

An illustrating example is what we call the "upset child" rule. Suppose a well-behaved child and his mother plan to picnic on the beach. Their choices are, respectively, θ_1, θ_2 where the mother's choice dominates except in extreme settings. Namely, for a small $\eta > 0$ value, let $F(\theta_1, \theta_2) = \theta_2$ for $\theta_2 - (\pi - \eta) \leq \theta_1 \leq \theta_2 + (\pi - \eta)$; i.e., unless the child's preferred outcome is almost directly opposite his mother's, the mother's choice dictates. As an example, if the child's wishes are within 179° of θ_1, then the mother's choice dictates; i.e., her choice holds except possibly for 2° arc of the circle. But as any parent appreciates, extreme differences in opinions can create an upset child and an accommodating parent. So near this extreme, for $\theta_2 - \pi \leq \theta_1 \leq \theta_2 - (\pi - \eta)$, let the decision rule linearly change from θ_2 at $\theta_1 = \theta_2 - (\pi - \eta)$ to θ_1 at $\theta_1 = \theta_2 - \pi$ ranging through the values on the $\theta_2 - (\pi - \eta)$ semicircle (to satisfy Parto), with a similar description on the $\theta_2 + (\pi - \eta) \leq \theta_1 \leq \theta_1 + \pi$ side except the values run through the other semicircle. Thus, whatever the mother's choice, the child can affect the outcome only when his preferred choice is in a small interval of length 2η that is directly opposite the mother's choice. As this interval can be chosen to be arbitrarily small, the mother clearly is the dominant decision maker. Who is the "topological dictator?" The child, not the mother! (Let $\eta \to \pi$, which shrinks to zero where the mother's choice dominates. Another step is necessary (see Section 6), but this gives the intuition.) As we show in Section 6, this example is essentially as extreme as possible.

This example seems to conflict with our earlier comment suggesting that if F is homotopic to a dictator, then a first approximation for F is the dictatorial choice. The explanation comes from the enormous flexibility that is allowed by the definition of homotopy; it permits the "dictatorial" approximation to be surprisingly crude. Just as a statistical average can be a misleading indicator for the properties of a data set, or the first term in a Taylor series approximation does not indicate very much about a function's behavior, the assertion that F is "homotopic to a dictator" is limited and potentially misleading.

3.1. The "topological dictator" is a misleading concept

A dramatic way to prove the misleading nature of this "topological dictator" notion is to use a rule that is, conceptually, diametrically opposite a dictatorship: the average choice. By satisfying anonymity (i.e., the outcome does not depend on the identity of any voter) this rule is far removed from a dictatorial setting, yet the following theorem shows for many natural choices of domains that this averaging rule *is* homotopic to a dictator.

Theorem 2. *Let F_A be the rule, defined on a line segment or in a square, that determines the average position desired by each of n people. This rule is homotopic to a dictator.*

Proof. If agent 1 is a dictator, the outcome always coincides with his choice as the decision rule is

$$F_D(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n) = \mathbf{p}_1. \tag{8}$$

If, on the other hand, the location for the n agents is the average of their choices, then

$$F_A(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n) = \frac{\mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n}{n}. \tag{9}$$

Consider the class of decision rules

$$F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n; t) = \frac{\mathbf{p}_1 + t(\mathbf{p}_2 + \cdots + \mathbf{p}_n)}{1 + t(n-1)} \tag{10}$$

where the t value describes the weight voters 2 through n have in the decision process. When $t = 1$, we have that $F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n; 1) = F_{\mathrm{A}}$, or the averaging technique; when $t = 0$ we have that $F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n; 0) = F_{\mathrm{D}}$, or the dictatorial approach. □

As Theorem 2 demonstrates, asserting that a rule is homotopic to a dictator can be essentially meaningless. In fact, any continuous decision rule on the square or line, where the input of $(n-1)$ agents – or criteria – can be decreased to zero, is homotopic to a dictator. The next example more dramatically dismisses the sense that a topological dictator is endowed with *any* power; it shows that even a rule where voter 1 *has absolutely no influence*, because the outcome selects the average choice of voters 2 to n, can be homotopic to the rule where voter 1 is the dictator! This is shown by

$$F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n; t) = \begin{cases} \dfrac{\mathbf{p}_1 + 2t(\mathbf{p}_2 + \cdots + \mathbf{p}_n)}{1 + 2t(n-1)}, & \text{for } t \in \left[0, \frac{1}{2}\right] \\ \dfrac{2(1-t)\mathbf{p}_1 + (\mathbf{p}_2 + \cdots + \mathbf{p}_n)}{2(1-t) + (n-1)}, & \text{for } t \in \left[\frac{1}{2}, 1\right] \end{cases}$$

Here, $F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n, 0)$ is where agent one is the dictator; $F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n, 1/2)$ is an intermediate rule selecting the average value of all n agents, and $F(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n, 1)$ is where agent one's wishes are totally ignored as the outcome is the averaged choice of all other agents. Indeed, it is easy to show that this rule is homotopic to two different dictators, or that two dictators can be homotopic to each other.

The next assertion should raise serious enough questions to completely dethrone the notion of a topological dictator. To keep the proof simple, rather than showing that any continuous rule is homotopic to a dictator, we show the essentially equivalent assertion that it is homotopic to a function of one variable (e.g., the preferences of a single voter), or, even more bothersome, to a constant function. A side feature of the proof is the connection made between contractibility comments and homotopy.

Theorem 3. *If $\mathcal{V}$ is contractible to a point, then any continuous*

$$F_1 : \mathcal{V} \times \cdots \times \mathcal{V} \to \mathcal{V}$$

is homotopic to a function of a single variable—or even to a constant function.

Proof. For $\mathcal{V}$ to be contractible to $\mathbf{p}_0 \in \mathcal{V}$, there exists a continuous

$$G(\mathbf{p}, t) : \mathcal{V} \times [0, 1] \to \mathcal{V}$$

where $G(\mathbf{p}, 1) = \mathbf{p}$ – the identity map – and $G(\mathbf{p}, 0) = \mathbf{p}_0$. To prove that F_1 is homotopic to a function of a single variable, say $\mathbf{p}_n$, use the homotopy

$$F(\mathbf{p}_1, \ldots, \mathbf{p}_n, t) = F_1(G(\mathbf{p}_1, t), \ldots, G(\mathbf{p}_{n-1}, t), \mathbf{p}_n).$$

By construction, $F(\mathbf{p}_1, \ldots, \mathbf{p}_n, 1) = F_1$ while $F(\mathbf{p}_1, \ldots, \mathbf{p}_n, 0) = F_1(\mathbf{p}_0, \ldots, \mathbf{p}_0, \mathbf{p}_n)$—a function of the single variable $\mathbf{p}_n$ (as all other variables are held fixed at $\mathbf{p}_0$). To prove that F_1 is homotopic to the constant function $F_0 = F_1(\mathbf{p}_0, \ldots, \mathbf{p}_0)$, use $F(\mathbf{p}_1, \ldots, \mathbf{p}_n, t) = F_1(G(\mathbf{p}_1, t), \ldots, G(\mathbf{p}_n, t))$. □

In fact, even if $\mathcal{V}$ is not contractible, F could be homotopic to a dictator. To illustrate with the non-contractible $\mathcal{V} = S^1$ and the earlier $F(\theta_1, \theta_2) = 6\theta_1 - 5\theta_2$, with the homotopic relationship

$$F^*(\theta_1, \theta_2, t) = (1 + 5t)\theta_1 - 5t\theta_2 = \theta_1 + 5t(\theta_1 - \theta_2),$$

$F^*(\theta_1, \theta_2, 1) = F(\theta_1, \theta_2)$ while $F^*(\theta_1, \theta_2, 0) = \theta_1$—a dictatorship. Indeed, appealing to the intuitive description of a homotopy, where the influence of all but one agent is progressively diminished to zero, it is difficult to imagine *any* mapping that is not homotopic to a dictator. In other words, rather than providing useful information, a greater surprise may be if a mapping is *not* homotopic to a dictator. Consequently, the notion of a "topological dictator" may be of minimal interest.

3.2. Replacing the topological dictator with a more meaningful concept

The reason the "topological dictator" concept can be seriously misleading is that the homotopy "reduces the influence of agents." Loosely speaking, treat this contraction as a topological first approximation; compare it to asserting that, when deriving the Taylor series of a function, we can compute the first approximation of the leading constant term. While this information is a start, we want much more. With series, we are more interested in learning about any restrictions on computing higher order terms of the series; with choice functions, we want to learn of any restrictions in enhancing the influence of other agents.

In other words, rather than knowing whether a rule can be *contracted* into a dictatorial one, which provides minimal information about the setting, we are more interested in whether it is possible to *expand* the influence of others. To capture the intent of the "topological dictator," we must change the focus. Rather than exploring how to diminish the influence of individuals, we must determine whether a specified domain allows rules that enhance the influence of individuals. The actual design of a rule depends on particular needs; we want to discover whether there are any limitations on this design. Indeed, the proofs in Sectiton 6 identify limitations of what can happen that go beyond what is described with the next condition.

A natural condition is to require a rule to satisfy anonymity; i.e., as $F(x, y) = F(y, x)$, one person does not has greater influence than the other. But it is easy to prove that this desired condition cannot occur (Section 6) in non-contractible regions, so we impose a weaker, more encompassing condition that includes anonymity as a special case. The motivation for the following condition is to ensure that, in similar settings, both agents have the same ability to effect an outcome.

Definition 1. A decision rule $F(x, y)$ satisfies the "shared effect" property if the following is true. If one agent selects $b \in \mathcal{V}$, and the second agent can select an $a \in \mathcal{V}$ to achieve the outcome $d \in \mathcal{V}$, then it also is possible that when the second agent selects the same $b \in \mathcal{V}$, the first agent can select $c \in \mathcal{V}$ to obtain the same d outcome.

While an anonymous rule satisfies this condition with $c = a$, there are many non-anonymous rules that also satisfy this condition.

Although knowing that a decision rule is homotopic to a dictator can be of limited interest, knowing that rules *cannot* have the shared effect property signals the rule's restrictive nature; perhaps one agent can accomplish more than the other. This sense of one agent being more effective because, "anything the other agent can do, the effective one can do better," must be made precise. To avoid technical details, the definition emphasizes not $\mathcal{V}$, but $\mathcal{V}_1$—the contracted domain and range that is a circle, or circles linked by line intervals.

Definition 2. Let continuous $F : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ satisfy unanimity and the generalized Pareto condition. Let $\mathcal{V}_1$ be the contracted version of $\mathcal{V}$ that is either a circle, or circles linked by intervals, and let $\tilde{F}$ be the associated version of F from $\mathcal{V}_1 \times \mathcal{V}_1 \to \mathcal{V}_1$. Agent j is the *more effective* agent if the following is true for $\tilde{F}$. If agent j selects any $b \in \mathcal{V}_1$, and agent k can select an $a \in \mathcal{V}_1$ to achieve the outcome $d \in \mathcal{V}_1$, then it also is possible when agent k selects the same $b \in \mathcal{V}_1$ for agent j to select $c \in \mathcal{V}_1$ yielding the same d outcome. However, the converse is false; there are outcomes agent j can force that, in the reversed situation, agent k cannot achieve.

The "upset child" example indicates why the word "effective" more accurately captures the sense of this definition than, say, "influential." After all, the mother is more influential, but the child is "more effective" in forcing a broader range of outcomes. To see this, let $\theta_1 = \theta^*$, and let the mother select θ_2 so that $F(\theta^*, \theta_2) = d$. The child can force this same d outcome because the range of outcomes for $F(\theta_1, \theta^*)$ with θ_1 in the interval in 2η distance on either side of $\theta^* + \pi$ covers the circle. So, if the mother selects θ^*, then the child can find a θ_1 so that $F(\theta_1, \theta^*) = d$. Thus, in the symmetric setting of who starts with θ^*, the child can obtain any outcome that the mother can attain. The converse, however, is not true. With the mother's choice of $\theta_2 = \theta^*$, the child can obtain the $\theta^* + \pi$ outcome as $F(\theta^* + \pi, \theta^*) = \theta^* + \pi$. If effectiveness were symmetric, then, when the child selects θ^*, the mother could select some θ_2 so that $F(\theta^*, \theta_2) = \theta^* + \pi$. This is impossible because the mother's choice of θ_2 is the outcome *until her choice approaches being diametrically opposite that of the child*; i.e., until θ_2 approaches $\theta^* + \pi$. In this region the outcome becomes progressively closer to the child's choice of θ^*; i.e., the child can force more outcomes than the mother.

Our basic result asserts that with holes in the domain and range, expect that the admissible functions do not allow all of the variables (agent's preferences, criteria, etc.) to have the same effect. Locally they may; globally they do not.

Theorem 4. *Let $\mathcal{V}$ be contractible to a circle, or circles linked by intervals. A continuous F satisfying unanimity and the generalized Pareto condition does not have the shared effectiveness property; instead, one agent is more effective than the other. If $\mathcal{V}$ is contractible to a point, then there exist functions with the shared effectiveness property.*

The proof for the first part of the theorem comes from the nature of our proof for Theorem 1. The second part (Section 6) uses the following nice theorem proved by Chichilnisky and Heal (1983) asserting that if $\mathcal{V}$ is contractible to a point, then there are rules that satisfy anonymity.

Theorem 5 (*Chichilnisky and Heal*). *Let F, given by* Eq. (1), *be an anonymous, continuous social choice function that satisfies unanimity. A necessary and sufficient condition for F to exist is that $\mathcal{V}$ is retractible.*

4. Source of the problems

Although Theorem 4 carries a negative and discouraging message, once we understand *why* (rather than *that*) these difficulties arise, we discover natural ways to sidestep them. To identify what causes these problems, slightly modify the beach party example by assuming that because a private home is located at the north end of the lake, the group cannot meet there. By removing this one point, the region becomes contractible, thus (Theorem 5) it is possible to design an anonymous, continuous choice function. Indeed, because the circle with the missing point can be opened and flattened into a line segment, one of these rules can be equated with the averaging method.

This averaging method has unfortunate properties. Suppose, for instance, there are only two people where both wish to be near the private home but one wishes to be to the east of the home while the other to the west. For this configuration of preferences, the averaging approach locates them on the south end of the lake—far from what either wants. Fortunately, as shown later, there are other rules that not only sidestep the limitations of Theorem 4, but also provide more appropriate conclusions. For now, the main point is that rules exist that enjoy shared effectiveness.

Now consider the beach problem involving both the location and the time. In addition to the private home, suppose that local regulations prohibit gatherings at 6 a.m. By removing this one 6 a.m. instant, time no longer can be modeled by points on a circle. Consequently, the combination of the private home and the time restriction changes the choice domain from the T^2 torus to a contractible region that is equivalent to a square. According to Theorem 5, there now are rules satisfying anonymity.

These examples illustrate that, somehow, the basic problem is caused by the geometry (actually, the topological structure) of the domain *and range* $\mathcal{V}$. To understand why the structure of outcomes plays a crucial role, notice that the averaging technique could locate the beach party in the middle of the lake.

We will explain these problems in terms of Arrow's Theorem. To do so, we first review why Arrow's Theorem occurs and then show how knowing the source of the problem allows us to find ways to avoid Arrow's conclusion. (For a more complete description about Arrow's and Sen's Theorems from this perspective, see (Saari, 1997, 2001).) Then we show why similar kinds of difficulties arise with continuous problems.

4.1. Arrow's Theorem

It is convenient to divide the assumptions for Arrow's Theorem (Arrow, 1951; Saari, 2001) into three parts. The first describes the voter preferences: each individual has a complete, transitive ranking of the $n \geq 3$ alternatives with no restrictions on how a voter can rank them. The second describes the societal outcome: it must be a complete transitive ranking. The final conditions describe the properties of the welfare function. The first property is Pareto: if everyone ranks a pair of alternatives in the same way, that is the pair's society ranking. The second property is binary independence (denoted by BI in what follows): the societal ranking of any pair is strictly determined by how all voters rank that pair. The conclusion for $n \geq 3$ alternatives and two or more voters is that the rule must be a dictatorship; i.e., the societal ranking always agrees with one particular voter's preference ranking.

Start with the three alternatives $\{$Austrian wine, French wine, Italian wine$\}$. If a person prefers French to Italian wine, does that person have transitive preferences? Without more information, this question cannot be answered because "transitivity" is a condition connecting all three binary rankings: without knowing the other two binary rankings, it is impossible to determine whether transitivity is, or is not, satisfied. In other words, while a binary ranking is a "local" construct, transitivity is a "global" constraint describing how the local conditions must connect with one another.

To understand the source of Arrow's Theorem, consider what happens when a decision rule satisfying BI determines the societal ranking of these three alternatives. To determine the societal ranking of French and Italian wines, for instance, BI prohibits the rule from using any information about how the voters rank French and Austrian, or Italian and Austrian wines. By preventing the rule from using anything other than the local information of binary rankings, *BI prohibits the rule from using the crucial assumption that the voters have transitive preferences*. The situation is not unlike an Escher print: locally everything seems reasonable, but, without invoking global constraints to connect the local parts, the final assembly can be unanticipated.

By using only local information, but not the global constraint, the only way we can expect to always obtain transitive outcomes is by restricting the data so severly that only transitive outcomes can occur. This happens with Black's single peaked condition (Black, 1958), which, for three alternatives, means that no voter has a particular alternative bottom ranked. This severe constraint ensures that whatever the societal binary rankings, they assemble into a transitive ranking. Another severe profile restriction is to use the preferences only of a specified individual—this is Arrow's dictator; i.e., rather than a "rule," Arrow's dictator is a profile restriction. If this individual has transitive preferences, the societal ranking is, necessarily, transitive.

This informational interpretation of Arrow's Theorem asserts that, inadvertently, BI forces the rule to ignore the important and available information how local information (individual binary preferences) is connected to create transitive preferences. To avoid the problems identified by Arrow's Theorem, just find ways to allow the decision rules to use this crucial "transitivity" information. Namely, as BI forbids using the information about transitive preferences, modify BI so that this transitivity information can be used. To indicate how to do this, observe that a feature distinguishing transitive rankings from arbitrary binary rankings is that we can count the number of alternatives that separate a particular pair of alternatives in a transitive ranking.

Definition 3. (Saari, 2001) A decision rule satisfies "intensity of binary independence" (IBI) if the societal relative ranking of any two alternatives is determined only by

1. each voter's relative ranking of the pair, and
2. the number of alternatives between the two alternatives in the voter's transitive ranking.

To illustrate, suppose a voter has preferences $A \succ B \succ C$. When determining the societal $\{A, C\}$ outcome, the only information BI allows to be used about this voter is her $A \succ C$ ranking. With IBI, however, the rule can use the $A \succ C$ ranking *and* that A and C are separated by one candidate. As we should expect, by including information about the global constraint of transitivity, Arrow's dictator is replaced by many other methods; one is the Borda Count. This is where each ballot is tallied by assigning $n - j$ points to a voter's jth ranked alternative, and the alternatives are ranked according to the sum of the assigned points.

Theorem 6. (Saari, 2001) *By replacing the BI condition in Arrow's Theorem with the IBI condition, an admissible rule is the Borda Count.*

Even stronger, by including anonymity and a couple of technical conditions, the Borda Count becomes the unique rule.

4.2. Topological dictators

The explanation for the "homotopic to a dictator" assertions and the more informative "more effective agent," is essentially the same as for Arrow's Theorem. Both include conditions that emphasize local behavior; this prevents the rule from using the specified global information about the preferences.

The local condition causing Arrow's Theorem is BI; surprisingly, the local condition causing the Eq. 1 problems is *continuity.* Recall, continuity of a function $f : R \to R$ at point x_0 requires what happens *near* x_0 to occur for f *near* the image $f(x_0)$; i.e., if a sequence of points $\{x_n\}$ approaches x_0 (i.e., $x_n \to x_0$ as $n \to \infty$), their images must approach $f(x_0)$ (i.e., $f(x_n) \to f(x_0)$

as $n \to \infty$). This definition allows continuity to be described in the following intuitive manner: "Suppose for any sequence $\{x_n\}$ and any specified degree of accuracy that, with the exception of some finite number of terms, it is impossible to distinguish the x_n's from x_0. Then f is continuous at x_0 if, with the exception of a finite number of terms, it is impossible to distinguish the $f(x_n)$'s from $f(x_0)$." This "inability to distinguish points" carries the distinct sense of being able to shrink a neighborhood of points about x_0 into x_0. In other words, treat continuity as requiring that, if this sense of "in the small, points can be contracted to x_0" occurs in the domain, then it also happens for image points about $f(x_0)$.

The problem with local behavior is that, alone, it cannot identify the global structure; e.g., just as local information about the shape of the Earth does not indicate whether the Earth is flat or spherical, the continuity assumptions alone when applied to the Eq. (1) choice of F cannot capture whether $\mathcal{V}$ is, or is not, contractible. If the global structure of $\mathcal{V}$ is compatible with the local behavior of continuity, where everything can be collapsed to a point, then we must expect that many admissible decision rules exist. For contractible regions, then, expect that many "shared effective" rules satisfy the conditions: this is the assertion of Theorem 5. On the other hand, with a non-contractible region, the local condition of continuity alone cannot capture adequate information about the structure of the domain for the rule: this causes the Theorem 4 type problems. The resolution is clear; as true with BI, we must supplement the local condition of continuity by including information about the global structure of preferences.

Following the lead of our discussion of Arrow's Theorem, there are several ways to address this problem. One is to severely restrict what data can be used, the other is to modify the conditions the rules must satisfy so that the information about the global properties can be used. To impose constraints on data, an approach that mimics Black's condition is to restrict the allowable voters' preferences; e.g., restrict voter preferences so that the de facto space of preferences is contractible. With the beach problem, for instance, this occurs if no voter can choose the north end of the beach. But without restricting what portions of the available data the rule can use, we end up with the Theorem 4 assertion about one agent being more effective.

Another approach is to modify the assumptions so the rules can use information about the global structure of $\mathcal{V}$. With Arrow's Theorem, this change was accomplished by replacing BI with IBI. The value of IBI is to indicate how to assemble the local structure of binary rankings to create the global transitivity structure. Similarly, for our Eq. (1) problems, we need to identify how to assemble local contractible regions into the non-contractible $\mathcal{V}$. We provide one approach, but most surely there are better ones: the nature of particular problems may dictate how to circumvent the difficulties.

4.3. A resolution

In all of the above examples, $\mathcal{V}$ is a smooth k-dimensional manifold; view this as a smooth k-dimensional surface that may reside in a higher dimensional space. Local portions of a k-dimensional manifold resemble open sets of R^k; e.g., locally the Earth, a sphere, resembles a portion of a plane. In fact, formal definitions of a k-dimensional manifold (e.g., see Spivak, 1999) describe a manifold in terms of "charts;" these are overlapping images of portions of R^k. An important part of the description describes how to connect the charts. As an example involving the Earth, a local chart is represented by a road map describing a particular region where, to use them, we understand how the overlaps of the different maps connect.

If a manifold is compact (if the domain is in n-dimensional spaces, this means "closed and bounded"), standard results from analysis assert that it can be covered with a finite number of charts. Each chart is contractible, so this description suggests how to sidestep the problem: let F also use information about the charts.

To see what can be done, cover $\mathcal{V}$ with $n \geq 3$ charts $\mathcal{C} = \{A_1, \ldots, A_n\}$ and let $\mathcal{R}$ be the space of complete, transitive rankings of the charts. Thus, $\mathbf{p} \in \mathcal{C} \times \mathcal{R}$ specifies a point in each chart and a ranking of the n charts. To illustrate with the beach party, divide the beach into three regions: a point $\mathbf{p}$ ranks the three regions, presumably reflecting the person's preferences to have the party in the different regions, and then specifies the voter's preferred location in each of the three regions.

This change in the domain requires minor modifications. Replace Eq. (1) with[3]

$$F : [\mathcal{C} \times \mathcal{R}] \times [\mathcal{C} \times \mathcal{R}] \times \cdots \times [\mathcal{C} \times \mathcal{R}] \to \mathcal{V}. \tag{11}$$

Namely, information about each agent's ranking of the regions and preferred choice within each region determines the location of the picnic. Continuity of F, then, is a local condition about points in each chart.

Replace the unanimity condition, Eq. 2, with the condition that for each $\mathbf{p} \in [\mathcal{C} \times \mathcal{R}]$,

$$F(\mathbf{p}, \mathbf{p}, \ldots, \mathbf{p}) = \mathbf{p}^* \in \mathcal{V} \tag{12}$$

where $\mathbf{p}^*$ is the common position specified in each voter's top-ranked chart. Just as Arrow's Theorem can be replaced with the positive Theorem 6 by permitting rules to use global information, by using information about the global structure of $\mathcal{V}$, all of the earlier negative theorems are replaced by the following positive assertion.

Theorem 7. *Let $\mathcal{V}$ be a compact manifold. For any number of agents there exist continuous, anonymous mappings F satisfying* Eq. (11), *the generalized Pareto condition, and the unanimity condition* Eq. (12).

Proof. Since each chart is contractible, a continuous distortion of the averaging approach can identify a point in that chart. It remains to select and rank the charts. Do this by using the Borda Count over the rankings. Break ties with some random approach. The outcome is the selected point in the top-ranked chart. □

Theorem 7 corresponds to common sense: first select the region, and then the location. Saari and Williams (1986) used a similar approach when analyzing economic message mechanisms.

There remains the undesirable properties exhibited by the "averaging" approach when a private home prevented having a beach party at that position. The reason for the undesired outcome is that, while removing a point avoids the dictatorial conclusion, the averaging rule still neglects the problem's global structure; it fails to utilize the information about the circular shape of the beach. But by using charts, a more reasonable outcome is obtained. (The optimal number and size of charts varies with the problem.)

5. Conclusion

When considering multivariable mappings from a space to itself (a decision, statistical, or aggregation function), it is surprising how seemingly minor restrictions of "holes" in the space can severely restrict the choice of mappings that satisfy minimal conditions. But by recognizing that the associated negative conditions occur because intended information is not being used, it

[3] With minor and obvious changes, the image $\mathcal{V}$ could be replaced with $\mathcal{C} \times \mathcal{R}$.

is possible to craft all sorts of new approaches to avoid these complexities. This assertion holds for Arrow's and Sen's Theorems (Saari, 2001; Saari and Petron, 2006) and, as shown here, for topological dictators and the more meaningful "effectiveness" problems.

6. Proofs

It remains to prove Theorem 1 and the first part of Theorem 4. In doing so, our goal is to develop intuition about these issues by making it clear why the conclusions hold. As in (Saari, 1997), we analyze the level sets of $F : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$; thus our proofs appeal to our understanding of indifference curves from economics. For instance, as true with indifference curves, the level sets define a *foliation* (see, for example, Spivak, 1999); e.g., different level sets cannot cross.

As examples, if $\mathcal{V}$ is the line segment $[0, 2\pi]$, then, with two agents, $\mathcal{V} \times \mathcal{V}$ is a square as in Fig. 2. The Fig. 2a level sets are for the averaging function $F(x, y) = (x + y)/2$. The diagonal $y = x$ describes unanimity points; each level set passes through this diagonal. The slanted nature of these level sets means that, to identify a particular level set, both x and y values are needed. In contrast, with the Fig. 2b horizontal level sets, each set is completely identified with only the y value. As F depends only on the y value – the x value is irrelevant – y is the dictator. Similarly, with vertical level sets, x is the dictator. The actual value of F is irrelevant, so $F(x, x) = x$ could be replaced with $F(x, x) = g(x)$.

The effect of holes; $k = 1$. To understand why the choices of F are constrained when $\mathcal{V}$ is a circle S^1, identify the torus $S^1 \times S^1$ with a square. To do so, remove the top point of the circle and flatten it into a line segment; i.e., each circle is identified with the segment $[0, 2\pi]$ where the endpoints agree. In this manner, $S^1 \times S^1$ can be represented as a square where the top and bottom edges agree (are the same) and the side edges agree. Indeed, taping together the top and bottom edges of the Fig. 2b square creates a cylinder. Next, taping the ends of the cylinder (the side edges) creates a torus.

When using a square to represent a torus, if a level set meets a top edge, then it also meets the bottom edge at the same x-distance; if a level set meets a side edge, it meets the other side at the same y-height. (Recall, opposite edges are the same.) These comments already indicate why "holes" limit the admissible functions. For instance, the Fig. 2a level sets *cannot* be level sets for a function on a torus because where a level set hits the left edge is not duplicated on the right edge. But the Fig. 2c curved line does meet both side edges at the same height, so it satisfies this condition. (This geometry makes it difficult to have anonymous mappings. For instance, anonymity requires a point (x, y) that is b height on the left edge to be on the same level set as (y, x), which is b distance from the right corner on the bottom edge. Thus any level set passing through an edge must pass through four points; two on the side edges and on the top and bottom. As level sets cannot cross, with a little experimentation, one sees why it is impossible to do this and satisfy the unanimity condition.)

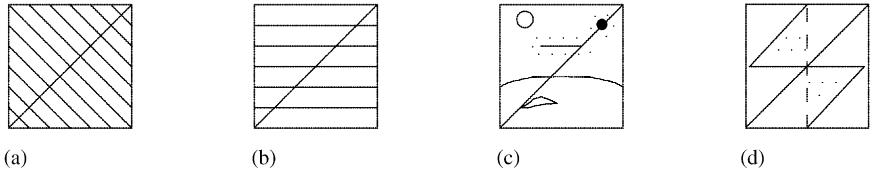


Fig. 2. Level sets for two methods. (a) Average; (b) y-dictator; (c) examples; (d) upset child.

The properties we need are

1. Level sets cannot cross.
2. Each level set meets the diagonal $y = x$ in precisely one point. (Unanimity)
3. If a level set meets the left edge, it meets the right edge at the same y height; if it meets the top edge, it meets the bottom edge at the same x distance.
4. As F is continuous, the level sets respect continuity conditions.

If $F_1(x, y)$ is homotopic to a $F_0(x, y) = y$ dictator, there is a continuous function $F(x, y, t) : S^1 \times S^1 \to S^1$ where $F(x, y, 1) = F_1$ and $F(x, y, 0) = F_0$. Varying parameter t, the $F(x, y, t)$ level sets continuously change from the F_1 level sets to horizontal lines. In Fig. 2d, for example, a levels set for the "upset child rule" are tilted and reversed "Z's;" all other level sets are direct copies translated along the unanimity line in the obvious manner. (Portions of the other level sets leave one edge and emerge on the other side.) The dashed line is where the child is the dictator; the horizontal dotted lines show how to diminish the mother's influence (with a homotopy) to change the rule from where the mother actually dominates into one where the child is the dictator.

Thus a geometric proof involves showing that if F_1 satisfies the Theorem 1 conditions, its level sets can be transformed either into horizontal or vertical lines. But we want more from our proof than what is stated in the theorems; we want to establish the precise constraints on F that are generated by these conditions. The way we do so is to determine the constraints on the level sets.

Consequences of the unanimity condition: Generically, level sets are closed lines. By continuity, each level set is a closed set: it could be points, or, in general, either closed curves (i.e., the image of a circle) or have endpoints (i.e., the image of a closed interval); the full level set could be a union of several of these objects. (See Fig. 2c. A level set can have "thick" regions, but this creates no problems as we can use a contained line.)

The unanimity assumption requires a level set to pass through the unanimity line only once. To appreciate the strong consequences of this condition, notice that if the level set is a point on the unanimity line, then, with continuity, neighboring level sets (the dotted circle in Fig. 2c) enclose the point. But this is not permitted as a neighboring level set would pass through the unanimity line more than once. A topological equivalent (as intervals can be collapsed into points) is if a level set is the image of a closed line segment that passes through the unanimity line, a loop touching (but not crossing the unanimity line), or a loop passing through the unanimity line creating a figure-eight. In all cases a neighboring level set (e.g., the Fig. 2c dotted ellipse) traces the figure and crosses the unanimity line in several locations, which is not allowed.

Indeed, suppose the x_0 level set passes through the unanimity line but does not divide the domain into two parts. This allows a curve to be drawn, close to but not meeting the x_0 level set, that connects points on the unanimity line slightly below and above (x_0, x_0). As the value of F on the two endpoints of this curve are above and below x_0, it follows from the intermediate value theorem, the Pareto condition (to avoid values of F going through $x_0 + \pi$), and the continuity of F that some point of this curve equals x_0, which is a contradiction. *Thus any level set passing through the unanimity line separates the domain.*[4] Therefore, for any x_0, the x_0-level set passing through (x_0, x_0) must continue to the left of the unanimity line, eventually pass through an edge

[4] The unanimity line divides the square into two parts and requires the level set for each outcome to meet this line in an *unique* point. Consequently the unanimity assumption can be replaced with any assumption satisfying these conditions (with a modified generalized Pareto condition).

of the square to emerge on the opposite edge and meet the (x_0, x_0) from the right. The following statements impose restrictions on how this can be done.

Claim 1. *For any x_0, $F(x_0, x_0 + \pi)$ equals either x_0 or $x_0 + \pi$.*

Proof. On the circle, x_0 and $x_0 + \pi$ are opposite points. For non-zero, arbitrarily small values of η, the Pareto condition places $F(x_0, x_0 + \pi + \eta)$ between x_0 and $x_0 + \pi + \eta$; i.e., $\eta > 0$ puts the image in one semicircle while $\eta < 0$ puts it in the other. If the claim is false, then $F(x_0, x_0 + \pi)$ is strictly in the interior of one semicircle. By choosing the sign of η so that $x_0 + \pi + \eta$ is in the opposite semicircle, the obvious contradiction (obtained by letting $\eta \to 0$) to the continuity of F proves the claim. □

Without loss of generality, assume that $F(x_0, x_0 + \pi) = x_0$. This is because if the contrary $F(x_0, x_0 + \pi) = x_0 + \pi$ held, we would analyze the $x' = x_0 + \pi$ level set where the second agent's choice is observed; i.e., $F(x' + \pi, x') = x'$.

Claim 2. *If for some x_0, we have that $F(x_0, x_0 + \pi) = x_0$, then for all $x \in S^1$, we have that $F(x, x + \pi) = x$. In other words, with diametrically opposed inputs, the first variable determines the outcome.*

Proof. The proof follows immediately from the continuity of F. Namely, for x arbitrarily close to x_0, continuity requires $F(x, x + \pi) = x$. If an x_1 exists where $F(x_1, x_1 + \pi) = x_1 + \pi$, it is trivial to show that the jump in the image to the "opposite side" violates continuity. □

Claim 3. *The x_0-level set meets the line $y = x + \pi$ only where $x = x_0$. In Fig. 3b, the x_0-level set can meet the slanted line only at the bullets at the top and bottom of the square. Moreover, the level set cannot be in either shaded triangular region.*

Proof. As $F(x, x + \pi) = x$ (Claim 2), the x_0 level set meets this slanted line iff $x = x_0$. To explain the upper left shaded triangular region, let x^* be to left of x_0; i.e., $x_0 - \pi < x^* < x_0$. The values (x^*, y) in this triangular region have $x^* + \pi < y \leq x_0 + \pi$; i.e., the shortest arc on the circle connecting x and y excludes x_0. Thus, the Pareto condition requires $F(x, y) \neq x_0$, so, the level set cannot enter this region. A similar explanation holds for the lower triangular region. □

Claim 4. *The x_0 level set cannot enter the two shaded squares given by $\{(x, y) | x_0 - \pi \leq x, y < x_0$ or $x_0 < x, y \leq x_0 + \pi\}$. Also, it cannot meet the points $(x_0 \pm \pi, x_0)$.*

Proof. The level set cannot meet $(x_0 \pm \pi, x_0)$ because $F(x_0 + \pi, x_0) = x_0 + \pi$. The rest of the conclusion follows from the Pareto condition; e.g., the lower square corresponds to where both

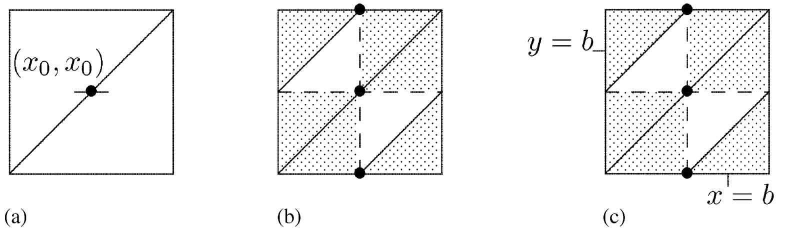


Fig. 3. Geometric arguments. (a) Segment of x_0-level set; (b) forbidden regions; (c) more effective.

x and y are to the left of x_0, so the outcome must also be in this semicircle. The upper square is where both are to the right. □

The severe limitations on the x_0-level set are apparent; it must stay in the open triangular regions (or on the dashed lines) and it must pass through the top and bottom center points. Thus the upset child rule is an extreme example; its x_0 level set closely traces the boundary by being on the horizontal dashed line until near the two end points, and then closely tracing the $y = x + \pi$ lines to the points on the top and bottom.

Any such level set can be contracted to the horizontal line, so the proof of Theorem 1 is proved for a circle (or any domain that can be contracted to a circle). To prove the first part of Theorem 4, we must show that the first agent is more effective than the second. To determine what the second agent can accomplish, let $x = b$, and assume, without loss of generality, that the second agent selected a so that $F(b, a) = x_0$. With Fig. 3c, (b, a) must be in one of the open triangular regions and the x_0 level set passes through this point. We now show that if the second agent assumes the value $y = b$, the first agent can find a value $x = c$ that also yields x_0. This follows from the construction. For instance, if $b > x_0$ (as in Fig. 3), then the point (b, a) that meets the x_0 level set is on the vertical line passing through the $x = b$ point on the bottom. For the first agent to obtain the same outcome when $y = b$, we only need to show that the horizontal line at height $y = b$ passes through the x_0 level set. But as the x_0 level set passes through the center (the unanimity line) and meets points at the bottom and top edges, it passes through any horizontal line.

The geometric constraints on the level set (where the horizontal span is less than the vertical one) prove that the first agent can cause outcomes that the second cannot. The extreme setting is if $y = b$ where the first agent can force the $b + \pi$ outcome because $F(b + \pi, b) = b + \pi$. (Claim 2). However, if $x = b$ and if $y = b + \pi$, then the outcome is b. If y is on either side of $b + \pi$, Pareto forces the outcome to be in the shortest semicircle between these values, which excludes $b + \pi$.

To generalize the Pareto condition, notice that the slanted $y = x + \pi$ constraint is due to continuity, so the only constraints caused by Pareto are the shaded squares. The Theorems 1 and 4 conclusions only require that the level set cannot meet the exterior edges of these squares, so the Pareto condition can be significantly relaxed.

More circles: It remains to handle settings where $\mathcal{V}_1$ consists of circles connected by an interval as in Fig. 4a. As it will be clear, the proof for two circles extends to any number, so this is the case considered here. Instead of a square, the unfolded region consists of four squares, labeled in the usual counterclockwise direction, from 1 to 4, connected by rectangles labeled E, N, W, S according to their compass direction, and a small central square C. Squares 1 and 3 represent where both agent's choices are, respectively, in the top and bottom circles. (Each circle is cut

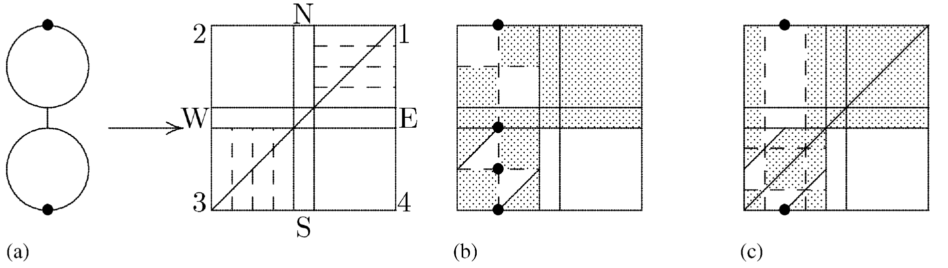


Fig. 4. Final arguments. (a) Analyzing two circles; (b) forbidden regions; (c) other choices.

open where it meets the connecting line.) Square 2 is where the first and second agents' choices are, respectively, in the lower and upper circle, while square 4 is the reverse. The rectangles are where one choice is in a circle and the other in the connecting interval; the small square C is where both choices are in the interval. In each circle, the earlier argument proves that one agent is more effective than the other; the problem is if the identity of who is more effective changes with the circles; e.g., as indicated with the Fig. 4a dashed lines representing level sets, the first agent is more effective in the lower circle while the second agent is in the upper circle.

To show that this cannot occur, consider the level set for the extreme point (diametrically opposite the connecting interval) in the lower circle indicated by the bullet; let it be x_0. The forbidden regions for its level set when both variables are in the lower circle are as in Fig. 3b. To extend these regions, consider a vertical line in Fig. 4b that is to the left of the bullet indicating x_0: any point on this line above square 3 represents where x is to the left of x_0 and y is either on the connecting interval or the other circle. As long as y is to the right of the bullet on the top circle, the Pareto condition prohibits a x_0 outcome. For points on the upper circle to the right of the top bullet, the shortest distance to x includes x_0, so that open square in square 2 is an admissible region for the level set. A similar argument holds for any vertical line to the right of the bullet, with an exchange of which half circle includes or excludes x_0, so the forbidden region excludes the two open small squares in square 2. Square 1 and regions C, N, E are where both variables are in the interval or upper circle, so Pareto excludes a x_0 outcome. Thus the forbidden region includes the shaded region of Fig. 4b.

Claim 5. *The x_0 level set connects the bullets on the top and bottom edges of* Fig. 4b.

Proof. The two vertices on the upper edge of square 3 belong to (by unanimity) the $x_0 \pm \pi$ level set while the midpoint is on the x_0 level set. Thus, from continuity and the Pareto condition, points on the each side of the midpoint are on level sets for values on different sides of the lower circle. This property is used in a manner mimicking the use of the unanimity line.

If the x_0 level set ends at this point, or if it extends into rectangle W but does not separate it, then there is a curve connecting points slightly to the left and right of the top bullet and missing the x_0 level set. Because the values of F on each endpoint are on opposite sides of x_0, a contradiction is obtained from the intermediate value theorem. Thus the x_0 level set must continue to meet square 2. A similar argument shows that if this level set does not separate square 2, then a curve can be found connecting opposite sides of where the x_0 set meets the bottom of square 2. The same intermediate value theorem leads to a contradiction.

To separate square 2, the level set cannot go through a side edge, as this would force it to enter a forbidden region on the other side. Similarly it cannot exit from the top or bottom edges except at the midpoint. This completes the proof. □

Claim 6. *If an agent is more effective in one circle, she is in all circles.*

Proof. Suppose the first agent is more effective in the lower circle and the second agent is in the upper circle; a situation as depicted in Fig. 4a. According to Claim 5, the level set for the most extreme bottom circle connects the top and bottom edges of the Fig. 4b square. By symmetry, the level set for the most extreme upper circle connects the left and right edges of the same square. Level sets cannot meet, so the contradiction proves the claim. □

A similar geometric argument shows that all levels sets for values in the upper or lower circle either meet the top and bottom edges, or the left and right edges. Indeed, in addition to the forbidden regions in Fig. 4c, the x_0 level set connecting the bullets on the top and bottom of the square further restricts other level sets. Assume that they are all horizontal; it remains to consider regions N, C, S. But, again, the same continuity arguments show that the level sets must connect the top and bottom edges of the large square. With this structure of the level sets, it is immediate that F is homotopic to a dictator and that one agent is more effective than the other.

The conclusion follows even after relaxing the generalized Pareto condition to restrict only certain boundaries of the shaded regions. Thus, these conclusions are general.

References

Arrow, K., 1951. Individual Values and Social Choice. Wiley, New York (2nd ed. 1962).

Baigent, N., 2006. in: Arrow K., Sen A., Suzumura K. (Eds.), Topological theories of social choice, Handbook of Social Choice and Welfare, vol. II, in press.

Black, D., 1958. The Theory of Committees and Elections. Cambridge University Press, London, New York.

Chichilnisky, G., 1982. The topological equivalence of the Pareto conditions and the existence of a dictator. J. Math. Econ. 9, 223–233.

Chichilnisky, G., Heal, G., 1983. Necessary and sufficient conditions for a resolution of the social choice paradox. J. Econ. Theory 31, 68–87.

Saari D.G., 1993. Comments at a conference “Analytic Models of Social Justice” at Columbia University organized by G. Chichilnisky and G. Heal.

Saari, D.G., 1997. Informational geometry of social choice. Soc. Choice Welfare 14, 211–232.

Saari, D.G., 2001. Decisions and Elections; Explaining the Unexpected. Cambridge University Press.

Saari D.G., Petron A., 2006. Negative externalities and Sen’s liberalism theorem. Econ. Theory 28, 265–281.

Saari, D.G., Williams, S., 1986. On the local convergence of economic mechanisms. J. Econ. Theory 40, 152–167 (Reprinted in J.M. Grandmont (Ed.), 1987. Nonlinear Economic Dynamics, Academic Press, Orlando, Florida).

Spivak, M., 1999. A Comprehensive Introduction to Differential Geometry, vol. 1-V, third ed. Publish or Perish, Inc., Houston, Texas.