Truthful germs are contagious: A local to global characterization of truthfulness

Aaron Archer
AT&T Labs – Research
180 Park Avenue, Florham Park, NJ 07932.
aarcher@research.att.com

Robert Kleinberg*
Computer Science Department, Cornell
University Ithaca, NY 14853
rdk@cs.cornell.edu

ABSTRACT

We study the question of how to easily recognize whether a social choice function f from an abstract type space to a set of outcomes is truthful, i.e. implementable by a truthful mechanism. In particular, if the restriction of f to every "simple" subset of the type space is truthful, does it imply that f is truthful? Saks and Yu proved one such theorem: when the set of outcomes is finite and the type space is convex, a function f is truthful if its restriction to every 2-element subset of the type space is truthful, a condition called weak monotonicity. This characterization fails for infinite outcome sets.

We provide a local-to-global characterization theorem for any set of outcomes (including infinite sets) and any convex space of types (including infinite-dimensional ones): a function f is truthful if its restriction to every sufficiently small 2-D neighborhood about each point is truthful. More precisely, f is truthful if and only if it satisfies local weak monotonicity and is *vortex-free,* meaning that the loop integral of f over every sufficiently small triangle vanishes. Our results apply equally well to multiple solution concepts, including dominant strategies, Nash and Bayes-Nash equilibrium, and to both deterministic and randomized mechanisms. When the type space is not convex, we show that f is truthful if and only if it extends to a truthful function on the convex hull of the original type space.

We use our characterization theorem to give a simple alternate derivation of the Saks-Yu theorem. Generalizing this, we give a sufficient condition for constructing a truthful function by "stitching together" truthful subfunctions on different subsets of the domain.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*; K.4.4 [**Computers and Society**]: Electronic Commerce

General Terms

Algorithms, Economics, Theory

*Supported in part by NSF awards CCF-0643934 and CCF-0729102.

Keywords

Mechanism design, truthful, weak monotonicity, vortex-free

1. INTRODUCTION

Mechanism design is a branch of social choice theory that seeks to implement social choice functions by pairing them with payments that induce players to report their preferences truthfully. An important question is how to easily recognize whether a particular social choice function f is truthful, i.e., whether or not there exists a payment scheme that can be paired with it to produce a truthful mechanism, without actually specifying the payments explicitly. The main goal of this paper is to give a broadly applicable characterization result that we hope will make it much easier for researchers to design truthful mechanisms for players with multi-dimensional types.

What should we look for in such a characterization? We want necessary and sufficient conditions for f to be truthful, and we want these conditions to be easy to check. In many cases, especially those of most interest to computer scientists, f is defined implicitly via an algorithm used to compute it. Understanding the behavior of f then entails understanding how the decisions in the algorithm are affected by changing the types reported by the players. Typically, it is much easier to analyze how an algorithm reacts to small, local perturbations of an input than to arbitrary changes. For instance, if the algorithm involves solving a linear program, then restricting ourselves to small perturbations allows us to do sensitivity analysis based on dual variables. Therefore, it would be highly desirable to have a local-to-global characterization result that allows us to check some conditions on arbitrarily small neighborhoods about each point in the type space, and conclude that f is truthful on the entire type space.[1] Moreover, it would be nice if these neighborhoods had further structure that made them even simpler, such as being low-dimensional.

Saks and Yu [15] make progress in this direction. Their result applies if there are finitely many outcomes and the type space is convex, in which case f is truthful as long as its restriction to every 2-element subset of the type space is truthful. This can be restated as a condition called *weak monotonicity* (WMON). It is easy to extend the Saks-Yu characterization to require only *local* WMON, i.e., it is sufficient to verify that WMON holds merely in some neighborhood of each point (Theorem 4.1). This gives the desired local-to-global characterization. This characterization can

[1] To put the question in more precise terms, let us say that the *germ* of f at $\mathbf{x}$ is truthful if there exists a truthful social choice function $\overline{f}$ such that $\overline{f} = f$ on an open neighborhood of $\mathbf{x}$. If the germ of f is truthful at every point, does it follow that f itself is truthful? We use the word "germ" here as it is used in sheaf theory. This also explains the pun in the title of this paper.

also be viewed as saying that f is truthful iff its restriction to every 1-D affine subspace of types is truthful. However, their result fails if there are infinitely many outcomes — a situation which is quite prevalent in mechanism design, e.g. problems that involve fractionally allocating divisible resources, or designing truthful-in-expectation mechanisms by extending a finite outcome set to the simplex of lotteries over those outcomes.

Statement of main results. Our main result (Theorem 3.1) applies to any arbitrary set $\mathcal{O}$ of outcomes (including infinite sets) provided the type space $\mathcal{T}$ is still convex (although it could be infinite-dimensional). In this case, truthfulness of f is equivalent to local WMON plus an additional property we call *vortex-freeness*. This condition says that for every point $\mathbf{x}$ in the type space, the loop integral of f is zero around every sufficiently small triangle with one corner at $\mathbf{x}$. Thus, while truthfulness in all 1-D affine subspaces is sufficient when the number of outcomes is finite, truthfulness in 2-D affine planes is sufficient when the number of outcomes is infinite. To demonstrate the power of our characterization theorem, we use it in Section 4 to easily derive the Saks-Yu theorem.

One potentially powerful tool that has been largely missing from mechanism design is an understanding of how one can combine multiple truthful allocations together such that the result is truthful. For example when an allocation function is computed by a program containing branch points, it may be the case that the for any particular sequence of branch outcomes, the resulting function is truthful; in other words, the type space can be decomposed into finitely many subsets (defined by the outcomes of the branch points) with a truthful allocation function on each. One then wishes to know whether the allocation function obtained by "stitching together" these subfunctions is truthful. In Section 5 we apply our characterization theorem to show this happens whenever WMON holds along the boundaries between pieces of the decomposition.

Since our characterization theorem holds only if $\mathcal{T}$ is convex, we complete the picture by considering what happens if it is not. In this case, Theorem 6.1 shows that f is truthful on $\mathcal{T}$ iff it can be extended to a truthful function on $\mathcal{T}^\sharp$, the convex hull of $\mathcal{T}$.

Extensions. Since WMON can be phrased as the absence of negative 2-cycles in a certain graph associated to f, we define MON(k) to be the absence of negative cycles of length k or less in the same graph, and *cycle monotonicity* (CMON) to be the absence of all negative cycles. In the Saks-Yu setting, truthfulness is equivalent to WMON, while in general, it is equivalent to CMON, by a theorem of Rochet (Theorem [14]). Thus, it is tempting to conjecture that, either in general or in some intermediate setting, there exists some k such that MON(k) is equivalent to truthfulness. Since Theorem 3.1 shows that loop integrals are important, it seems most logical to look at type spaces that are path-connected. In Section 7 we give examples of functions on path-connected domains that refute this conjecture: they satisfy MON(k) but not MON($k+1$). These examples are based on rotations in $\mathbb{R}^2$.

Solution concepts. So far, we have been deliberately vague about the game-theoretic solution concept we are using. This is because our characterization theorem applies equally well to each of the most common solution concepts in mechanism design: dominant strategies, Nash and Bayes-Nash equilibrium (with interdependent but independently distributed types). In all these cases, determining whether a mechanism is truthful boils down to analyzing mechanisms involving just a single player. For dominant strategies, bidding his true type must be each player's best strategy regardless of the types declared by the other players. Thus, every vector of types declared by the other players presents a single-player mechanism to this player. The social choice function f is truthful in dominant strategies iff each such single-player mechanism it induces is truthful.

For Nash equilibrium, we must consider the single-player mechanism induced when all other players bid truthfully. But since each player's true type can be anything in the type space, we must analyze the same induced single-player mechanisms as in dominant strategies.

For Bayes-Nash equilibrium, a player's bid and the distribution of the other players' types induces randomizations over the pure outcomes. The outcome space faced in the single-player mechanism is precisely this set of lotteries. If the players' valuations are interdependent (i.e., one player's valuation for a given outcome depends on the actual types of other players), then from the viewpoint of any one player, the mechanism combined with the random process of sampling the other players' types, can be regarded together as a single-player mechanism in which the outcomes are tuples consisting of an outcome of the original mechanism design problem together with a vector of the other players' types. This expanded outcome space will typically be infinite, corresponding to an expanded type space of infinite dimension. Similarly, we handle randomized mechanisms by considering the outcome space to include lotteries over the pure outcomes.

In light of this discussion, from now on we focus only on single-player mechanisms.

Conventions. As is common in mechanism design, we work in the setting of *quasi-linear* utility functions, which means that the player's utility is the sum of his intrinsic *valuation* for the outcome and the monetary payment he receives. The player's *type* is simply his valuation function from outcomes to the reals. Thus, $\mathcal{T} \subseteq \mathbb{R}^{\mathcal{O}}$. When we assume convexity of the type space $\mathcal{T}$, we mean convexity in the linear space of all functions over $\mathcal{O}$. That is, for every pair of types $\mathbf{x}_0, \mathbf{x}_1 \in \mathcal{T}$ and every $\lambda \in [0,1]$ there exists a type $\mathbf{x}_\lambda \in \mathcal{T}$ satisfying $\mathbf{x}_\lambda(\mathbf{a}) = (1-\lambda)\mathbf{x}_0(\mathbf{a}) + \lambda\mathbf{x}_1(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{O}$.

To make our exposition more accessible and intuitive, we have written the majority of the paper such that the reader may assume a more familiar model — that $\mathcal{T}$ and $\mathcal{O}$ are both subsets of $\mathbb{R}^n$, and $\mathbf{x}(\mathbf{a}) = \mathbf{x} \cdot \mathbf{a}$, where $\mathbf{x} \in \mathcal{T}$ and $\mathbf{a} \in \mathcal{O}$. In this case, convexity of the type space is the same as geometric convexity in $\mathbb{R}^n$. If $\mathcal{O}$ is finite, then we can take $n = |\mathcal{O}|$, $\mathcal{O} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, and $\mathbf{x}_i$ is the valuation for outcome i. If the outcome function is randomized and $\mathbf{a}_i$ is the probability of selecting outcome i, then $\mathbf{x} \cdot \mathbf{a}$ is the player's expected valuation of the outcome. Alternatively, if the n dimensions correspond n different goods, $\mathbf{a}_i$ is the amount of good i that the player receives, $\mathbf{x}_i$ is his value per unit of that good, and the values are additive, then $\mathbf{x} \cdot \mathbf{a}$ is his value for the bundle $\mathbf{a}$. While we focus on this model for its comfort and familiarity, our characterization theorem holds for the more abstract (possibly infinite-dimensional) setting described above. In Remark 3.1 we explain the syntactic changes in our proof that make it valid for the more general setting.

1.1 Related work

The problem of characterizing truthful social choice functions has been studied for many years for various domains of types. The most universal positive result, due to Vickrey [17], Clarke [4] and Groves [5], says that for any $\mathcal{T}$ and $\mathcal{O}$, the function f that maximizes the sum of the players' valuations is truthful. This is the celebrated VCG mechanism. Roberts [13] showed that if $3 \leq |\mathcal{O}| = n < \infty$ and $\mathcal{T} = \mathbb{R}^n$ for each player, then weighted affine maximizers (i.e., simple variations of the VCG mechanism) are the only truthful functions. Thus, for finite sets of more than two outcomes and unrestricted types, a mechanism is truthful iff it is weighted VCG. Thus, the interesting cases are when either $|\mathcal{O}| = \infty$ or $\mathcal{T}$ is restricted in some way.

When the types and outcomes are both 1-dimensional and satisfy the Spence-Mirrlees single-crossing condition, f is truthful iff it is monotone. This result is implicit in the work of Mirrlees [8] and Spence [16]. Note that such type spaces are not necessarily convex in function space. For the special case of dot-product valuations, this result was made explicit by Myerson in the setting of auction design [12], then later rediscovered by Archer and Tardos in the context of discrete optimization [2].

In multi-dimensional type spaces, WMON is essentially the 1-D monotoncity condition applied to 1-D affine subspaces. Bikchandani et al. prove that WMON is equivalent to truthfulness on a variety of domains [3], and Gui, Müller, and Vohra extend this result to more domains [6]. All of these type spaces can be described as polyhedra in $\mathbb{R}^n$, where $n = |\mathcal{O}|$. Saks and Yu [15] generalize this result to cover all convex domains in $\mathbb{R}^n$. Monderer gives an alternate inductive proof of the Saks-Yu theorem [9]. Yet another simple proof of the Saks-Yu theorem is given by Vohra in [18]; this proof generalizes to show that even in the infinite-outcome case, an allocation function on a convex domain is truthful if its restriction to every subset of the type space with at most three extreme points is truthful.

McAfee and McMillan studied the case in which the type space and outcome space lie in $\mathbb{R}^n$, and the utilities are not necessarily quasi-linear but rather any twice-differentiable function of the type, allocation, and payment [7]. When specialized to the quasi-linear case with dot-product valuations, their result says that a mechanism (f, p) is truthful iff the matrix f' (the derivative of f) is positive semidefinite and the consumer surplus function $\mathbf{x} \cdot f(\mathbf{x}) + p(\mathbf{x})$ is given by a line integral of f starting from a fixed basepoint $\mathbf{x}_0$, plus any fixed additive constant.[2] In particular, a payment function making (f, p) truthful exists only if the line integral is path-independent, but they do not mention the conditions under which this is the case. It turns out that this is the case iff f' is symmetric. Being symmetric is equivalent to f having zero exterior derivative, which for differentiable functions is equivalent to vortex-freeness (by Stokes's theorem). It is easy to see that for differentiable f, f' is positive semidefinite iff f satisfies WMON. Thus, the two conditions of our Theorem 3.1 map directly to the two conditions in (our modified version of) the McAfee-McMillan theorem. Indeed, trying to understand the relationship between this theorem and Rochet's CMON characterization was a jumping-off point for our research.

Any allocation function computed via an algorithm containing a branching point is almost guaranteed to be non-smooth. Hence, the McAfee-McMillan theorem is inadequate for most problems of interest to computer scientists. Our characterization theorem fills that void since it applies even to non-continuous f. In fact, our "truthful stitching" theorem (Theorem 5.1) directly applies to the case of algorithms containing branching points which result in non-smooth but piecewise-truthful allocation functions.

Müller, Perea, and Wolf study truthfulness in finite-dimensional type spaces under Bayes-Nash equilibrium [10]. They allow interdependent valuations but assume independently distributed types (as we do). In this setting, they prove some of the results we needed along the way to our characterization theorem (although we developed our results before learning of theirs). Translating their results to our terminology, they show that an allocation function f is truthful iff it satisfies WMON and its line integrals are path-independent, provided the type space is convex. However, their result does not feature the main selling point of our chacterization, which is that ours depends only on local properties of f, whereas the path-independence condition is a global property of f that is usually much more difficult to check. Our approach offers further benefits. Namely, by abstracting to infinite-dimensional type spaces, we can treat the Bayes-Nash and dominant strategy solution concepts using a unified framework. This crystalizes the geometric ideas underlying the argument, while substantially generalizing it, simplifying notation, and making it (we hope) more accessible.

[2]For readers trying to replicate our translation of this result, we note that it involves integration by parts, and reconstructing potentials using line integrals of their gradients.

2. PRELIMINARIES

We specified our conventions regarding outcome sets and type spaces in Section 1. We will sometimes write the agent's valuation $\mathbf{x}(\mathbf{a})$ in the alternate form $v(\mathbf{x}, \mathbf{a})$.

An *allocation function* is a function $f : \mathcal{T} \to \mathcal{O}$. A *single-player mechanism* is a pair consisting of an allocation function f and a payment function $p : \mathcal{T} \to \mathbb{R}$. The semantics behind this notation are that the player knows his type $\mathbf{x}$ but reports a (possibly different) type $\mathbf{y}$ to the mechanism, whereupon the mechanism selects the outcome $f(\mathbf{y})$ and gives the player a monetary payment of $p(\mathbf{y})$. The mechanism is *truthful* if the agent can never improve his utility (valuation plus payment) by lying about his type. In other words, truthfulness is equivalent to the statement that for all $\mathbf{x}, \mathbf{y} \in \mathcal{T}$,

$$v(\mathbf{x}, f(\mathbf{x})) + p(\mathbf{x}) \geq v(\mathbf{x}, f(\mathbf{y})) + p(\mathbf{y}). \tag{1}$$

An allocation function f is truthful (sometimes called *rationalizable* or *incentive-compatible*) if there exists a payment function p such that the mechanism (f, p) is truthful. When a player of type $\mathbf{x}$ bids truthfully, his resulting utility is called the *consumer surplus*, $\sigma(\mathbf{x}) = v(\mathbf{x}, f(\mathbf{x})) + p(\mathbf{x})$. The following well-known characterization of truthfulness follows directly from (1).

Proposition 2.1. *If $\mathcal{T}$ is convex, then the mechanism (f, p) is truthful iff the consumer surplus function σ is convex and $f(\mathbf{x})$ is a subgradient of σ at $\mathbf{x}$, for each $\mathbf{x} \in \mathcal{T}$.*

The following characterization of truthful allocation functions f is due to Rochet [14].

Definition 2.1. An allocation function $f : \mathcal{T} \to \mathcal{O}$ satisfies *cyclic monotonicity* (CMON) if for every sequence of types $\mathbf{x}_1, \ldots, \mathbf{x}_k$ (with indices taken mod k) it holds that

$$\sum_{i=0}^{k} v(\mathbf{x}_{i+1}, f(\mathbf{x}_{i+1})) - v(\mathbf{x}_i, f(\mathbf{x}_{i+1})) \geq 0. \tag{2}$$

A sequence $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k = \mathbf{x}_0$ which violates (2) is called a *negative k-cycle*.

Theorem 2.2 (Rochet). *An allocation function $f : \mathcal{T} \to \mathcal{O}$ is truthful if and only if it satisfies CMON.*

For completeness, and to give the reader an appreciation for how it relates to the ideas behind our results, we include what is essentially Rochet's original proof of Theorem 2.2, but using some of the network language of Gui, Müller and Vohra [6].

Proof of Theorem 2.2: If the mechanism (f, p) is truthful, then by (1), $v(\mathbf{x}_i, f(\mathbf{x}_i)) + p(\mathbf{x}_i) \geq v(\mathbf{x}_i, f(\mathbf{x}_{i+1})) + p(\mathbf{x}_{i+1})$ for each i. Summing over i yields (2), so f satisfies CMON.

Conversely, suppose f satisfies CMON. For any two types $\mathbf{x}, \mathbf{y} \in \mathcal{T}$ define $\ell(\mathbf{x}, \mathbf{y})$ to be the infimum of the lengths of all finite paths from $\mathbf{x}$ to $\mathbf{y}$ in $G_P(f)$. Note that the set of all such path lengths in bounded below by $-w_P(\mathbf{y}, \mathbf{x})$, as otherwise there would be a negative-weight cycle. Hence $\ell(\mathbf{x}, \mathbf{y})$ is a well-defined real number. Now fix any type $\mathbf{x}_0$ and define a payment function by $p(\mathbf{x}) = \ell(\mathbf{x}_0, \mathbf{x})$. Observe that

$$p(\mathbf{x}) \leq p(\mathbf{y}) + w_P(\mathbf{y}, \mathbf{x}) = p(\mathbf{y}) + v(\mathbf{y}, f(\mathbf{y})) - v(\mathbf{y}, f(\mathbf{x})),$$

and the assertion that (f, p) is truthful follows by rearranging terms. ■

Note that after rearranging terms, we may rewrite (2) as

$$\sum_{i=1}^{k} v(\mathbf{x}_i, f(\mathbf{x}_i)) - v(\mathbf{x}_i, f(\mathbf{x}_{i+1})) \geq 0. \tag{3}$$

One may define a directed graph with (possibly infinite) vertex set $\mathcal{T}$ and with edge weights defined using the summands in either (2) or (3). When the weight of edge $\mathbf{x} \to \mathbf{y}$ is $v(\mathbf{y}, f(\mathbf{y})) - v(\mathbf{x}, f(\mathbf{y}))$ we refer to this as the S-weight $w_S(\mathbf{x}, \mathbf{y})$ and denote the graph by $G_S(f)$. When the weight of edge $\mathbf{x} \to \mathbf{y}$ is $v(\mathbf{x}, f(\mathbf{x})) - v(\mathbf{x}, f(\mathbf{y}))$ we refer to this as the P-weight $w_P(\mathbf{x}, \mathbf{y})$ and denote the graph by $G_P(f)$. The S and P are mnemonics that we motivate below. Note that CMON is equivalent to the assertion that $G_P(f)$ contains no negative-weight cycles or, equivalently, that $G_S(f)$ contains no negative-weight cycles. Also note that when $\mathcal{T}, \mathcal{O} \subseteq \mathbb{R}^n$ then the P-weights and S-weights are given by the formulas

$$w_S(\mathbf{x}, \mathbf{y}) = (\mathbf{y} - \mathbf{x}) \cdot f(\mathbf{y}) \tag{4}$$

$$w_P(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot (f(\mathbf{x}) - f(\mathbf{y})). \tag{5}$$

The description of CMON in terms of negative cycles in $G_P(f)$ is due to Gui, Müller and Vohra. It allows for a very nice interpretation of the payments as being dual weights in a shortest path problem, since the dual feasibiliy constraint on edge $\mathbf{x} \to \mathbf{y}$ boils down to (1) [6]. We term (4) the P-weight precisely because of this connection, where P stands for ***P***ayment. We named (5) the S-weight because shortest path labels in $G_S(f)$ correspond to the consumer ***S***urplus σ.

3. THE CHARACTERIZATION THEOREM

We now embark on the proof of our main characterization theorem. To understand both where our proof is going and the creative process that led to it, it is helpful to draw a connection between Rochet's CMON condition and loop integrals. There are two ways to write CMON, in terms of S-weights (5), or in terms of P-weights (4). In recent papers, the latter has been more popular because it ties more directly into the beautfiul Gui-Müller-Vohra network interpretation [6], allowing us to view the payment function in terms of linear programming duality. However, we prefer the former, because S-weights are more obviously related to line integrals. In particular, a cycle C in $G_S(f)$ corresponds to a polygonal loop Γ in $\mathcal{T}$, connecting each type in the cycle to the next. If we were to take the right-hand Riemann sum to approximate the loop integral of f over Γ, using the very rough partition consisting of just the types in C, the terms would correspond exactly to the S-weights in C. But notice that we can refine this partition of Γ as much as we like, so CMON implies that every loop integral is non-negative, which in turn implies that every loop integral is zero (since reversing the loop negates its integral). Conversely, if we can show that every polygonal loop integral of f is zero and the Riemann sums converge from above, then f satisfies CMON. The first condition will follow from vortex-freeness, whereas the second condition follows from WMON.

Remark 3.1. All of the statements and proofs in this section hold in the setting of abstract outcome sets and infinite-dimensional type spaces, provided the notation is interpreted correctly. However, it is conceptually helpful to assume for this discussion that $\mathcal{T}, \mathcal{O} \subseteq \mathbb{R}^n$. For example, this ensures that dot products have a well-defined meaning. The next two paragraphs explain how to re-interpret our notation, definitions, and proofs in the infinite-dimensional setting. The reader is advised to skip these and assume $\mathcal{T}, \mathcal{O} \subseteq \mathbb{R}^n$ on a first pass through this section.

A dot product such as $\mathbf{x} \cdot \mathbf{a}$ should be interpreted to mean the valuation $\mathbf{x}(\mathbf{a})$ (equivalently, $v(\mathbf{x}, \mathbf{a})$). This also applies to expressions that need to be expanded using the distributive law, e.g., $(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{a} - \mathbf{b})$ denotes $\mathbf{x}(\mathbf{a}) - \mathbf{x}(\mathbf{b}) - \mathbf{y}(\mathbf{a}) + \mathbf{y}(\mathbf{b})$.

The interpretation of line integrals in the infinite-dimensional case is as follows. Since every line integral considered in this section is defined over a polygonal path consisting of one or more line segments in $\mathcal{T}$, it suffices to define the line integral over a single line segment. If f is a function from $\mathcal{T}$ to $\mathcal{O}$, $\mathbf{x}_0, \mathbf{x}_1$ are any two types, L is the line segment from $\mathbf{x}_0$ to $\mathbf{x}_1$, and $\mathbf{x}_t = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1$, then $\int_L f(\mathbf{x}) \cdot d\mathbf{x}$ denotes the integral $\int_0^1 \mathbf{x}_1(f(\mathbf{x}_t)) - \mathbf{x}_0(f(\mathbf{x}_t))\, dt$.

With this in mind, we present the following definitions.

Definition 3.1. An allocation function $f : \mathcal{T} \to \mathcal{O}$ satisfies *local weak monotonicity (local WMON)* if for every $\mathbf{x} \in \mathcal{T}$ and every line L through $\mathbf{x}$, there exists an open neighborhood U about $\mathbf{x}$ such that

$$(\mathbf{x} - \mathbf{y}) \cdot (f(\mathbf{x}) - f(\mathbf{y})) \geq 0$$

for all $\mathbf{y} \in L \cap U$.

Definition 3.2. An allocation function $f : \mathcal{T} \to \mathcal{O}$ is *locally path-integrable* if $\mathcal{T}$ has an open covering such that for every line segment L lying entirely in a single piece of the open covering, the line integral $\int_L f(\mathbf{x}) \cdot d\mathbf{x}$ is well-defined and finite.

Definition 3.3. A locally path-integrable allocation function $f : \mathcal{T} \to \mathcal{O}$ is *vortex-free* if for every $\mathbf{x} \in \mathcal{T}$ and every 2-dimensional plane Π through $\mathbf{x}$, there exists an open neighborhood U about $\mathbf{x}$ such that the path integral $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x}$ vanishes for every triangle $\triangle$ in $\Pi \cap U$ with one corner at $\mathbf{x}$.

Remark 3.2. The definition of vortex-free implies a seemingly stronger condition: for every $\mathbf{x} \in \mathcal{T}$ and every 2-dimensional plane Π through it, there exists an open neighborhood U about $\mathbf{x}$ such that the path integral $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x}$ vanishes for every triangle $\triangle$ in $\Pi \cap U$. To see this, take U as in Definition 3.3, let $\mathbf{x}_0 = \mathbf{x}$ and let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ be the corners of $\triangle$. Now for $0 \leq i, j \leq 3$, define L_{ij} to be a line segment directed from $\mathbf{x}_i$ to $\mathbf{x}_j$ and let $W_{ij} = \int_{L_{ij}} f(\mathbf{x}) \cdot d\mathbf{x}$. From the definition of vortex-free, we know that the loop integral $\oint_{\triangle'} f(\mathbf{x}) \cdot d\mathbf{x}$ vanishes when $\triangle'$ is a triangle contained in $\Pi \cap U$ with one corner at $\mathbf{x}_0$. Thus,

$$W_{01} + W_{12} - W_{02} = 0$$
$$W_{02} + W_{23} - W_{03} = 0$$
$$W_{03} + W_{31} - W_{01} = 0,$$

where we used the fact that reversing a path negates its line integral. Summing these three equations, we obtain $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x} = 0$, as desired.

We can now state our characterization theorem.

Theorem 3.1. *Let $\mathcal{T}$ be a convex type space and let $f : \mathcal{T} \to \mathcal{O}$ be a locally path-integrable allocation function. Then f is truthful if and only if it is vortex-free and satisfies local WMON.*

Let us compare this theorem to the McAfee-McMillan result we discussed in Section 1.1. There, in the differentiable case, f is truthful if and only if its derivative matrix f' is symmetric and positive semidefinite. The symmetry condition is equivalent to f being the gradient of some function, which turns out to be the consumer surplus function σ.[3] Given this, the positive-semidefiniteness of f' is equivalent to convexity of σ. Hence, these conditions together are necessary and sufficient for truthfulness by Proposition 2.1. It turns out that the important thing about f' being symmetric is that it guarantees the path integral $\oint_\Gamma f(\mathbf{x}) \cdot d\mathbf{x}$ is equal to zero. Thus, it should be clear that the positive semidefiniteness of f' plays an analogous role to local WMON in Theorem 3.1, while the symmetry of f' is analogous to vortex-freeness.

Before proving Theorem 3.1, it will be useful to establish a few properties that follow from Definitions 3.1 and 3.3.

Lemma 3.2. *An allocation function $f : \mathcal{T} \to \mathcal{O}$ satisfies WMON iff for every $\mathbf{x} \in \mathcal{T}$ and every vector $\mathbf{h}$, the function $g(t) = f(\mathbf{x} + t\mathbf{h}) \cdot \mathbf{h}$ is increasing on the subset of $\mathbb{R}$ on which it is defined. In particular, if $\mathcal{T}$ is convex and f satisfies local WMON, then it satisfies WMON.*

Proof: Let $\mathbf{x}, \mathbf{h}, g$ be as in the lemma, and let $s < t$ be any two real numbers such that the vectors $\mathbf{y} = \mathbf{x} + s\mathbf{h}, \mathbf{z} = \mathbf{x} + t\mathbf{h}$ belong to $\mathcal{T}$. From the equation

$$g(t) - g(s) = (f(\mathbf{z}) - f(\mathbf{y})) \cdot \mathbf{h} = (f(\mathbf{z}) - f(\mathbf{y})) \cdot \frac{(\mathbf{z} - \mathbf{y})}{t - s}$$

we see that f satisfies WMON if and only if $g(t) > g(s)$ for all such g, s, t. The final statement in the lemma follows because any function which is defined on an interval and is locally increasing at every point is increasing on the whole interval. ∎

Lemma 3.3. *Let $g : [0, 1] \to \mathbb{R}$ be an increasing function, and let $0 = x_0 < x_1 < \cdots < x_N = 1$ and $0 = y_0 < y_1 < \cdots < y_M = 1$ be two increasing sequences such that $(y_j)_{j=0}^M$ refines $(x_i)_{i=0}^N$, i.e., (x_i) is a subsequence of (y_j). Then the Riemann sums of g with respect to $(x_i), (y_j)$ satisfy the inequality*

$$\sum_{i=1}^{N}(x_i - x_{i-1})g(x_i) \geq \sum_{j=1}^{M}(y_j - y_{j-1})g(y_j). \qquad (6)$$

Proof: It suffices to prove the lemma in the case $M = N + 1$; the general case then follows by induction. So assume that for some r we have $x_{r-2} = y_{r-1}$ and $x_r = y_r$. We will use the notation Δ_i^x (resp. Δ_i^y) to denote $x_i - x_{i-1}$ (resp. $y_i - y_{i-1}$). In the sum on the left side of (6) the $i = q$ term on the left side matches the $j = q$ term on the right side for $q \leq r$ and it matches the $j = q + 1$ term on the right side for $q \geq r + 2$. Hence

$$\begin{aligned}
&\sum_{i=1}^{N} \Delta_i^x g(x_i) - \sum_{j=1}^{M} \Delta_j^y g(y_j) \\
&\quad = \Delta_r^x g(x_r) - \Delta_r^y g(y_r) - \Delta_{r-1}^y g(y_{r-1}) \\
&\quad = (y_r - y_{r-2})g(y_r) - \Delta_r^y g(y_r) - \Delta_{r-1}^y g(y_{r-1}) \\
&\quad = \Delta_{r-1}^y (g(y_r) - g(y_{r-1})) \geq 0.
\end{aligned}$$

∎

Corollary 3.4. *If $f : \mathcal{T} \to \mathcal{O}$ is an allocation function that satisfies WMON, and L is a line segment in $\mathcal{T}$ with endpoints $\mathbf{x}, \mathbf{y}$, then*

$$(\mathbf{y} - \mathbf{x}) \cdot f(\mathbf{y}) \geq \int_L f(\mathbf{z}) \cdot d\mathbf{z} \qquad (7)$$

[3]The differentiability assumptions rule out σ having multiple subgradients.

Proof: We will apply Lemma 3.3 to the function $g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \cdot (\mathbf{y} - \mathbf{x})$, which is increasing by Lemma 3.2. For non-negative integers k and $i \leq 2^k$, let $x_i^k = i/2^k$. Note that for each k, $0 = x_0^k < x_1^k < \cdots < x_{2^k}^k = 1$, and that $(x_j^{k+1})_{j=0}^{2^{k+1}}$ refines $(x_i^k)_{i=0}^{2^k}$. By Lemma 3.3, the sequence of Riemann sums

$$S_k = \sum_{i=1}^{2^k}(x_i^k - x_{i-1}^k)g(x_i^k)$$

is decreasing. Moreover, by the definition of the Riemann integral, $\int_L f(\mathbf{z}) \cdot d\mathbf{z} = \lim_{k\to\infty} S_k$. Hence

$$(\mathbf{y} - \mathbf{x}) \cdot f(\mathbf{y}) = S_0 \geq \lim_{k\to\infty} S_k = \int_L f(\mathbf{z}) \cdot d\mathbf{z}.$$

∎

Definition 3.4. Suppose oriented triangles $\triangle_1 = \mathbf{x}_0 \to \mathbf{x}_1 \to \mathbf{x}_2 \hookleftarrow$ and $\triangle_2 = \mathbf{y}_0 \to \mathbf{y}_1 \to \mathbf{y}_2 \hookleftarrow$ are co-planar. Then we say $\triangle_1$ and $\triangle_2$ are *oriented consistently* if there exists a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with positive determinant such that $[(\mathbf{y}_1 - \mathbf{y}_0); (\mathbf{y}_0 - \mathbf{y}_2)]A = [(\mathbf{x}_1 - \mathbf{x}_0); (\mathbf{x}_0 - \mathbf{x}_2)]$.

This definition captures the intuitive notion that $\triangle_1$ and $\triangle_2$ are both oriented clockwise or both counterclockwise.

Lemma 3.5. *If $f : \mathcal{T} \to \mathcal{O}$ is vortex-free, then for every triangle $\triangle$ contained in $\mathcal{T}$, the path integral $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x}$ vanishes.*

Proof: For clarity, in this proof we will distinguish between *triangles* (sets consisting of three points and the three line segments joining them) and 2-*simplices* (the convex hull of three points). We will use the following geometric fact: if σ_1, σ_2 are 2-simplices with disjoint interiors which share a side in common, and the boundaries of σ_1, σ_2 are triangles $\triangle_1, \triangle_2$ oriented consistently. then $\triangle_1$ and $\triangle_2$ traverse the common side of σ_1, σ_2 in opposite directions.

Let V be the 2-simplex consisting of $\triangle$ and its interior. The definition of vortex-free (combined with Remark 3.2) implies that V has an open covering $\{U_i \,|\, i \in \mathcal{I}\}$ such that for every i and every triangle $\triangle'$ contained in U_i, the integral $\oint_{\triangle'} f(\mathbf{x}) \cdot d\mathbf{x}$ vanishes. Because V is compact, we can apply the Lebesgue number lemma [11] to deduce that there is a $\delta > 0$ such that every set of diameter less than δ is contained in one of the sets U_i. We can subdivide V into 2-simplices $\sigma_1, \sigma_2, \ldots, \sigma_N$ of diameter less than δ, and let $\triangle_i$ be a closed curve tracing out the boundary of σ_i; assume every $\triangle_i$ is oriented consistently with a single, fixed orientation of T. If we write

$$0 = \sum_{i=1}^{N} \oint_{\triangle_i} f(\mathbf{x}) \cdot d\mathbf{x}$$

and break each loop integral on the right side into a sum of three integrals along line segments forming the boundary of σ_i, then each such line segment appears either

1. twice with opposite orientations, if it is on the common boundary between two simplices σ_i, σ_j,
2. once, if it is a subset of T.

Terms of the first type cancel each other out, while those of the second type sum up to $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x}$. Thus $\oint_\triangle f(\mathbf{x}) \cdot d\mathbf{x} = 0$, as claimed. ∎

Now, for the main proof.

Proof of Theorem 3.1: First assume f is truthful. Hence it satisfies CMON. This immediately implies WMON and therefore, *a* *fortiori*, local WMON. If L is any line segment contained in $\mathcal{T}$ and $\mathbf{x}, \mathbf{y}$ are its endpoints, then the function $g(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \cdot (\mathbf{y} - \mathbf{x})$ is increasing on $[0, 1]$, hence it is Riemann integrable on that interval. This implies that f is locally path-integrable. To see that f is vortex-free, we argue by contradiction, i.e., we will show that if f is not vortex-free then it fails to satisfy CMON. Assuming f is not vortex-free, there is a triangle $\triangle$ such that $\oint_\triangle f \cdot d\mathbf{x} \neq 0$. Reversing the orientation of $\triangle$ if necessary, we may assume that $\oint_\triangle f \cdot d\mathbf{x} < 0$. Since the integral is the limit of Riemann sums, there must be a negative Riemann sum, i.e., a sequence of points $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in $\triangle$ such that

$$\sum_{i=i}^{N} f(\mathbf{x}_i) \cdot (\mathbf{x}_i - \mathbf{x}_{i-1}) < 0.$$

This sequence constitutes a negative N-cycle.

Conversely, suppose f is vortex-free and satisfies local WMON. We will prove that f satisfies CMON, from which it follows immediately that f is truthful. For any sequence of type vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N = \mathbf{x}_0$, let $\{L_{ij} : 0 \leq i < j \leq N\}$ denote the set of paths $L_{ij}(t) = (1-t)\mathbf{x}_i + t\mathbf{x}_j$, i.e., L_{ij} traces out a line segment from $\mathbf{x}_i$ to $\mathbf{x}_j$. If P is the polygonal closed curve formed by concatenating $L_{01}, L_{12}, \ldots, L_{(N-1)N}$, then Corollary 3.4 implies

$$\oint_P f(\mathbf{x}) \cdot d\mathbf{x} = \sum_{i=1}^{N} \int_{L_{(i-1)i}} f(\mathbf{x}) \cdot d\mathbf{x} \leq \sum_{i=1}^{N} f(\mathbf{x}_i) \cdot (\mathbf{x}_i - \mathbf{x}_{i-1}), \tag{8}$$

so to prove CMON (i.e., that the sum on the right side of (8) is non-negative) it suffices to prove that $\oint_P f(\mathbf{x}) \cdot d\mathbf{x} = 0$. For $0 \leq i < j \leq N$ let

$$W_{ij} = \int_{L_{ij}} f(\mathbf{x}) \cdot d\mathbf{x}, \quad W_{ji} = -W_{ij}.$$

For $i = 1, 2, \ldots, N-2$ let T_i denote the triangle formed from $L_{0i}, L_{i(i+1)}, L_{(i+1)0}$. Lemma 3.5 implies that

$$0 = \oint_{T_i} f(\mathbf{x}) \cdot d\mathbf{x} = W_{0i} + W_{i(i+1)} + W_{(i+1)0}. \tag{9}$$

Interpreting the subscripts mod N and summing (9) as i runs from 1 to $N-2$ yields

$$\begin{aligned} 0 &= \sum_{i=1}^{N-2} (W_{0i} + W_{i(i+1)} + W_{(i+1)0}) \\ &= \sum_{i=1}^{N} W_{i(i+1)} + \sum_{i=2}^{N-2} (W_{0i} + W_{i0}) \\ &= \sum_{i=1}^{N} W_{i(i+1)} = \oint_P f(\mathbf{x}) \cdot d\mathbf{x}. \end{aligned}$$

■

The following two corollaries are now trivial.

Corollary 3.6. *If $\mathcal{T}$ is convex and the restriction of f to $\Pi \cap \mathcal{T}$ is truthful for every 2-dimensional affine subspace Π, then f is truthful.*

Proof: The definitions of local WMON and vortex-free depend only on the restrictions of f to sets of the form $\Pi \cap \mathcal{T}$ where Π is an affine subspace of dimension at most 2. (In the case of local WMON, in fact, it suffices to consider 1-dimensional affine subspaces Π.) ■

Corollary 3.7. *If $\mathcal{T}$ is convex and every $\mathbf{x} \in \mathcal{T}$ has an open neighborhood U such that the restriction to f to U is truthful, then f is truthful.*

Proof: The definitions of local WMON and vortex-free depend only on the restrictions of f to sufficiently small neighborhoods of every point in $\mathcal{T}$. ■

4. APPLICATION TO FINITE $|\mathcal{O}|$

In this section, we use Theorem 3.1 to give an easy new proof of the Saks-Yu theorem [15]. *As such, each of our lemmas will include the implicit blanket hypothesis that $|\mathcal{O}| < \infty$, f satisfies WMON, and $\mathcal{T}$ is convex.* Their paper states the theorem in terms of WMON rather than local WMON; the equivalence of the version stated here follows easily, e.g., using Lemma 3.2.

Theorem 4.1 (Saks-Yu [15])**.** *If $|\mathcal{O}|$ is finite, $\mathcal{T}$ is convex, and f satisfies local WMON, then f is truthful.*

We begin by exploring the geometric structure of f and some simple properties of its line integrals. For $\mathbf{a} \in \mathcal{O}$, let $D_\mathbf{a} = f^{-1}(\mathbf{a}) = \{\mathbf{x} \in \mathcal{T} : f(\mathbf{x}) = \mathbf{a}\}$. For any set S, let $\bar{S}$ and S° denote its topological closure and interior, respectively.

Definition 4.1. Let $I \subseteq \mathcal{O}$. Define $B_I = \cap_{\mathbf{a} \in I} \bar{D}_\mathbf{a}$ to be the *common boundary* of I, and $E_I = B_I \cap f^{-1}(I)$ to be the *exclusive common boundary* of I. We say that outcomes $\mathbf{a}, \mathbf{b}$ are *adjacent* if $B_{\{\mathbf{a},\mathbf{b}\}} \neq \emptyset$, and each type in $B_{\{\mathbf{a},\mathbf{b}\}}$ is called a *boundary type* for $\mathbf{a}$ and $\mathbf{b}$.

By analogy to physical chemistry, we think of the ensemble of sets $\bar{D}_\mathbf{a}, \mathbf{a} \in \mathcal{O}$, as the "phase diagram" of the allocation function f, because it represents the value of f geometrically, and the boundaries correspond to "phase transitions" where the behavior of f changes abruptly. Each $\bar{D}_\mathbf{a}$ is a "cell" of this phase diagram. Lemma 4.2 sums up some key structural properties of the phase diagram that are also used in [6, 15]. Its proof is a straightforward application of WMON.

Lemma 4.2. *For $\mathbf{a} \in \mathcal{O}$, $\bar{D}_\mathbf{a}$ is the intersection of $\bar{\mathcal{T}}$ with the polyhedron $P_\mathbf{a}$ defined by the inequalities $\mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) \geq \sup_{\mathbf{y} \in \bar{D}_\mathbf{b}} \mathbf{y} \cdot (\mathbf{a} - \mathbf{b})$, for each $\mathbf{b} \in \mathcal{O}$. Moreover, if $\mathbf{x}_0 \in \mathcal{T}$ lies in $B_{\{\mathbf{a},\mathbf{b}\}}$, then the entire boundary is contained in the hyperplane defined by $\{\mathbf{x} \in \mathcal{T} : \mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) = \mathbf{x}_0 \cdot (\mathbf{a} - \mathbf{b})\}$.*

Proof: If $\mathbf{a}, \mathbf{b} \in \mathcal{O}$ and $\mathbf{x} \in D_\mathbf{a}$, $\mathbf{y} \in D_\mathbf{b}$, then WMON implies that

$$\mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) \geq \mathbf{y} \cdot (\mathbf{a} - \mathbf{b})$$
$$\inf_{\mathbf{x} \in \bar{D}_\mathbf{a}} \mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) \geq \sup_{\mathbf{y} \in \bar{D}_\mathbf{b}} v \cdot (\mathbf{a} - \mathbf{b}). \tag{10}$$

Note that if $\mathbf{a}$ and $\mathbf{b}$ are adjacent, then the inequality in (10) is tight, with the extremum for each side occuring at every boundary type.

Let the half-space $H_{\mathbf{ab}} = \{\mathbf{x} \in \mathcal{T} : \mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) \geq \sup_{\mathbf{y} \in \bar{D}_\mathbf{b}} \mathbf{y} \cdot (\mathbf{a} - \mathbf{b})\}$, and let $P_\mathbf{a}$ denote the polyhedron $\cap_{\mathbf{b} \in \mathcal{O}} H_{\mathbf{ab}}$. We will show that $\bar{D}_\mathbf{a} = P_\mathbf{a} \cap \bar{\mathcal{T}}$. By (10), we have $\bar{D}_\mathbf{a} \subseteq P_\mathbf{a}$. Moreover, no type $\mathbf{x} \in P_\mathbf{a}^\circ$ can have $f(\mathbf{x}) = \mathbf{b} \neq \mathbf{a}$, since $\mathbf{x} \in H_{\mathbf{ab}}^\circ$, which implies $\mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) > \sup_{\mathbf{y} \in \bar{D}_\mathbf{b}} \mathbf{y} \cdot (\mathbf{a} - \mathbf{b})$. Hence, $P_\mathbf{a}^\circ \cap \mathcal{T} \subseteq D_\mathbf{a}$, so $P_\mathbf{a} \cap \bar{\mathcal{T}} \subseteq \bar{D}_\mathbf{a}$. Finally, if $\mathbf{x}_0$ is on the boundary between $\mathbf{a}$ and $\mathbf{b}$, then we know $\bar{D}_\mathbf{a} \subseteq H_{\mathbf{ab}}$ and $\bar{D}_\mathbf{b} \subseteq H_{\mathbf{ba}}$, so $\bar{D}_\mathbf{a} \cap \bar{D}_\mathbf{b} \subseteq H_{\mathbf{ab}} \cap H_{\mathbf{ba}} = \{\mathbf{x} : \mathbf{x} \cdot (\mathbf{a} - \mathbf{b}) = \mathbf{x}_0 \cdot (\mathbf{a} - \mathbf{b})\}$. ■

Lemma 4.3. *Suppose $I \subseteq \mathcal{O}$, $\mathbf{a} \in I$ and Γ is a polygonal path from $\mathbf{y}_0$ to $\mathbf{y}_1$ that lies entirely in E_I, except perhaps for $\mathbf{y}_0$ and $\mathbf{y}_1$. Then $\int_\Gamma f(\mathbf{x})d\mathbf{x} = \mathbf{a} \cdot (\mathbf{y}_1 - \mathbf{y}_0)$.*

Proof: It suffices to consider the case where Γ is just a line segment, which we parametrize as $\mathbf{y}_t = \mathbf{y}_0 + t(\mathbf{y}_1 - \mathbf{y}_0)$, $t \in [0,1]$. For every $t \in (0,1)$, $\mathbf{y}_t \in E_I$, so $(\mathbf{a} - f(\mathbf{y}_t)) \cdot (\mathbf{y}_1 - \mathbf{y}_0) = 2(\mathbf{a} - f(\mathbf{y}_t)) \cdot (\mathbf{y}_{1/4} - \mathbf{y}_{3/4}) = 0$ by Lemma 4.2. Thus,

$$\int_\Gamma f(\mathbf{x}) \cdot d\mathbf{x} = \int_0^1 f(\mathbf{y}_t) \cdot (\mathbf{y}_1 - \mathbf{y}_0)dt$$
$$= \int_0^1 \mathbf{a} \cdot (\mathbf{y}_1 - \mathbf{y}_0)dt = \mathbf{a} \cdot (\mathbf{y}_1 - \mathbf{y}_0).$$

■

Lemma 4.4. *Suppose $\mathbf{x} \in \mathcal{T}$, Π is a 2-D plane through $\mathbf{x}$ such that $\Pi \cap \mathcal{T}$ is 2-D, and $\mathcal{O}_\mathbf{x} = \{\mathbf{a} : \mathbf{x} \in \bar{D}_\mathbf{a}, (\bar{D}_\mathbf{a} \cap \Pi) - \{\mathbf{x}\} \neq \emptyset\}$. Then there exists a disc $N \subseteq \Pi \cap \mathcal{T}$ centered at $\mathbf{x}$ such that for every $\mathbf{b} \in \mathcal{O}_\mathbf{x}$, $\bar{D}_\mathbf{b} \cap N$ is a convex sector of N.*

In other words, the restriction of each cell of the phase diagram to N is either all of N or looks like a slice of pizza. This follows easily from the cells of the phase diagram being polyhedral.

Proof of Lemma 4.4: Take N to be sufficiently small that it excludes all cells of the phase diagram that do not contain $\mathbf{x}_0$ (which we can do because the cells are closed). Moreover, for every $\mathbf{a} \in \mathcal{O}$ for which $\mathbf{x}_0$ lies in the relative interior of $\bar{D}_\mathbf{a}$ with respect to Π, let us also take N to be so small that it lies inside $\bar{D}_\mathbf{a}$. For all other $\mathbf{a} \in \mathcal{O}$, $\mathbf{x}_0$ lies on the boundary of $\bar{D}_\mathbf{a} \cap \Pi$ relative to Π. By Lemma 4.2, there is some polyhedron $P_\mathbf{a}$ such that $\bar{D}_\mathbf{a} = P_\mathbf{a} \cap \bar{\mathcal{T}}$. Hence, $P_\mathbf{a} \cap \Pi$ is some (possibly unbounded) polygon with $\mathbf{x}_0$ on its boundary. Let us choose N to be small enough that it contains no vertices of $P_\mathbf{a} \cap \Pi$, aside from $\mathbf{x}_0$. Let $\mathcal{O}_N = \{\mathbf{a} \in \mathcal{O} : \mathbf{x}_0 \in \bar{D}_\mathbf{a}, (\bar{D}_\mathbf{a} \cap N) - \{\mathbf{x}_0\} \neq \emptyset\}$, that is, the set of outcomes $\mathbf{a}$ for which $\mathbf{x}_0$ is a boundary type, and for which $\bar{D}_\mathbf{a}$ intersects N at some other point in addition to $\mathbf{x}_0$. By our construction of N, for every $\mathbf{a} \in \mathcal{O}_N$, $\bar{D}_\mathbf{a} \cap N$ is either all of N or a wedge-shaped slice of N, coming to a point at $\mathbf{x}_0$ (including a 180° wedge as one possible case). ■

Now for the main proof.

Proof of Theorem 4.1: In order to apply Theorem 3.1, we need only show that f is vortex-free. Consider any type $\mathbf{x}_0$, and fix any 2-D plane Π through $\mathbf{x}_0$. If $\mathbf{x}_0$ is not a boundary type, then f is constant in some ball around $\mathbf{x}_0$, so loop integrals inside that ball vanish. If $\Pi \cap \mathcal{T}$ is at most 1-D, then there are no non-degenerate triangular loops to check, so vortex-freeness holds trivially. In the case where $\Pi \cap \mathcal{T}$ is 2-D and $\mathbf{x}_0$ is a boundary type, then we choose N as in Lemma 4.4.

Consider how the the phase diagram of f could look, restricted to the small neighborhood N. Because each $\bar{D}_\mathbf{a}$ (for $\mathbf{a} \in \mathcal{O}_{\mathbf{x}_0}$) is either all of N or a wedge radiating from $\mathbf{x}_0$ (by Lemma 4.4), when we overlay all of these wedges to obtain the phase diagram, it looks like a pizza with some finite number of slices, all coming together at $\mathbf{x}_0$. Each slice S_i is the intersection of N with B_{I_i} for some $I_i \subseteq \mathcal{O}$, and the slices are numbered consecutively around $\mathbf{x}_0$. For each slice S_i, select some outcome $\mathbf{a}_i \in I_i$.

We fix an arbitrary[4] polygonal loop Γ within N and will show that $\int_\Gamma f(\mathbf{x}) \cdot d\mathbf{x} = 0$. Let $\mathbf{y}_i$, $i = 1, \ldots, k$ be the consecutive types along Γ where the loop crosses from one pizza slice $S_{j(i-1)}$ to the next $S_{j(i)}$. Let Γ_i be the path from $\mathbf{y}_i$ to $\mathbf{y}_{i+1}$ along Γ. By Lemma 4.3, $\int_{\Gamma_i} f(\mathbf{x}) \cdot d\mathbf{x} = \mathbf{a}_{j(i)} \cdot (\mathbf{y}_{i+1} - \mathbf{y}_i)$. Hence,

$$\int_\Gamma f(\mathbf{x}) \cdot d\mathbf{x} = \sum_{i=1}^k \int_{\Gamma_i} f(\mathbf{x}) \cdot d\mathbf{x}$$
$$= \sum_{i=1}^k \mathbf{a}_{j(i)} \cdot (\mathbf{y}_{i+1} - \mathbf{y}_i)$$
$$= \sum_{i=1}^k \mathbf{y}_i \cdot (\mathbf{a}_{j(i-1)} - \mathbf{a}_{j(i)})$$
$$= \sum_{i=1}^k \mathbf{x}_0 \cdot (\mathbf{a}_{j(i-1)} - \mathbf{a}_{j(i)}).$$

The last line follows from Lemma 4.2 because both $\mathbf{y}_i$ and $\mathbf{x}_0$ are boundary types for outcomes $\mathbf{a}_{j(i-1)}$ and $\mathbf{a}_{j(i)}$. This last sum telescopes to zero. ■

[4]For this proof, it will not matter whether Γ is a triangle or whether $\mathbf{x}_0$ is a vertex of Γ.

5. STITCHING TRUTHFUL FUNCTIONS

This section applies Theorem 3.1 to address the following question: when is a "piecewise-truthful" function guaranteed to be truthful? In other words, when can one construct a truthful function f on a convex type space $\mathcal{T}$ by "stitching together" truthful functions f_i defined on subsets of $\mathcal{T}$? Actually, we will require the subfunctions f_i to satisfy a weaker condition called *local truthfulness* which says that the germ of f_i at every point is truthful (see footnote 1) or, equivalently, that f_i satisfies local WMON and vortex freeness at every point.

Theorem 5.1. *Suppose that a finite-dimensional convex type space $\mathcal{T}$ is covered by closed sets $\{\mathcal{T}_i : i \in \mathcal{I}\}$ such that:*

1. *the covering is locally finite;*
2. *each set $\mathcal{T}_i$ is the closure of its interior;*
3. *the pairwise intersections $\mathcal{T}_i \cap \mathcal{T}_j$ are piecewise differentiable and have positive codimension in $\mathcal{T}$.*

Suppose that f is a function on $\mathcal{T}$, and that for each $i \in \mathcal{I}$, we have a locally truthful function f_i on $\mathcal{T}_i$, continuous at each point of the boundary $\partial\mathcal{T}_i$, such that $f = f_i$ on the interior of $\mathcal{T}_i$. If f satisfies local WMON, then f is truthful.

Note that the Saks-Yu Theorem, Theorem 4.1, constitutes the special case in which each f_i is a constant function. The proof of Theorem 5.1 requires the following simple lemma.

Lemma 5.2. *Under the hypotheses of Theorem 5.1, suppose P is a differentiable path in $\partial\mathcal{T}_i$. If P has derivative $\mathbf{h}$ as it passes through a point $\mathbf{x}$, then $\mathbf{h} \cdot f(\mathbf{x}) = \mathbf{h} \cdot f_i(\mathbf{x})$. Consequently,*

$$\int_P f(\mathbf{x})d\mathbf{x} = \int_P f_i(\mathbf{x})d\mathbf{x}.$$

Proof: Assume without loss of generality that P is parametrized by a differentiable function $\gamma : [-1,1] \to \partial\mathcal{T}_i$ such that $\gamma(0) = \mathbf{x}$. For $i = 1, 2, \ldots$ let $\mathbf{x}_i = \gamma(1/i)$ and let $\mathbf{y}_i$ be any point in the interior of $\mathcal{T}_i$ such that $\|\mathbf{y}_i - \mathbf{x}_i\| \leq (1/i)\|\mathbf{x}_i - \mathbf{x}\|$. By construction, the vectors $i(\mathbf{x}_i - \mathbf{x})$ converge to $\mathbf{h}$ and the vectors $i(\mathbf{y}_i - \mathbf{x}_i)$ converge to 0, so $i(\mathbf{y}_i - \mathbf{x}) \to \mathbf{h}$. Also, $f(\mathbf{y}_i) - f(\mathbf{x})$ converges to $f_i(\mathbf{x}) - f(\mathbf{x})$, because f_i is continuous at $\mathbf{x}$ and is equal to f at each of the points $\mathbf{y}_i$. Applying the fact that f satisfies WMON, we obtain

$$\mathbf{h} \cdot (f_i(\mathbf{x}) - f(\mathbf{x})) = \lim_{i \to \infty} [i(\mathbf{y}_i - \mathbf{x}) \cdot (f(\mathbf{y}_i) - f(\mathbf{x}))] \geq 0.$$

A similar argument using the points $\tilde{\mathbf{x}}_i = \gamma(-1/i)$ establishes that $-\mathbf{h} \cdot (f_i(\mathbf{x}) - f(\mathbf{x})) \geq 0$. Hence $\mathbf{h} \cdot f_i(\mathbf{x}) = \mathbf{h} \cdot f(\mathbf{x})$.

The final assertion, that $\int_P f(\mathbf{x})\, d\mathbf{x} = \int_P f_i(\mathbf{x})\, d\mathbf{x}$, follows by approximating each integral by a Riemann sum and comparing the sums term-by-term using the first part of the lemma. ■

Proof of Theorem 5.1: Since f satisfies local WMON, it suffices to prove that it is vortex free. As in the proof of Lemma 3.5, any triangular loop $\triangle$ in $\mathcal{T}$ can be decomposed into a finite number of closed loops L composed of finitely many segments, each of which is either a subinterval of a side of $\triangle$ or a differentible path in one of the sets $\mathcal{T}_i \cap \mathcal{T}_j (i \neq j)$, such that the segments of the second type cancel each other (each is matched by an oppositely-oriented version of the same segment in another loop L') and such that each loop L is completely contained in one of the sets $\mathcal{T}_i$ for some $i = i(L)$. Then we have

$$\oint_{\triangle} f(\mathbf{x}) d\mathbf{x} = \sum_L \oint_L f(\mathbf{x}) d\mathbf{x} = \sum_L \oint_L f_i(\mathbf{x}) d\mathbf{x} = 0,$$

where the second equality holds by Lemma 5.2 and the third one holds because f_i is vortex-free. ■

6. NON-CONVEX TYPE SPACES

In this section we prove that an allocation function on a non-convex type space $\mathcal{T}$ is truthful if and only if it extends to a truthful allocation function on the convex hull of the type space, $\mathcal{T}^\sharp$. The basic idea is that if f is truthful, then by analogy to Proposition 2.1 we can construct a corresponding convex consumer surplus function for it on $\mathcal{T}$, where for a non-convex domain, a function is defined to be convex if it has a subgradient at each point. The trick is to extend this to a convex consumer surplus function on $\mathcal{T}^\sharp$ without invalidating any of the original subgradients, then pick a subgradient at each point of $\mathcal{T}^\sharp - \mathcal{T}$ that lies in $\mathcal{O}$. Our theorem requires the types and outcomes to satisfy a compactness criterion that we now explain.

Definition 6.1. Given an outcome set $\mathcal{O}$ and type space $\mathcal{T} \subseteq \mathbb{R}^{\mathcal{O}}$, for every $\mathbf{a} \in \mathcal{O}$ we may define a function $v_{\mathbf{a}} : \mathcal{T} \to \mathbb{R}$ by $v_{\mathbf{a}}(\mathbf{x}) = v(\mathbf{x}, \mathbf{a})$. We say that $(\Theta, \mathcal{O})$ satisfies *outcome compactness* if the set $\{u_{\mathbf{a}} \,|\, \mathbf{a} \in \mathcal{O}\}$ is a compact subset of $\mathbb{R}^{\mathcal{T}}$ in the product topology.

Remark 6.1. If $\mathcal{O}$ is a compact topological space and each $x \in \mathcal{T} \subseteq \mathbb{R}^{\mathcal{O}}$ is a continuous function from $\mathcal{O}$ to $\mathbb{R}$, then the function $\mathbf{a} \mapsto v_{\mathbf{a}}$ is continuous and hence $(\mathcal{T}, \mathcal{O})$ satisfies outcome compactness. In particular, this holds when $\mathcal{T}, \mathcal{O} \subseteq \mathbb{R}^n$ and $\mathcal{O}$ is compact.

Theorem 6.1. *Let $\mathcal{T}$ be any type space and let $\mathcal{T}^\sharp$ denote the convex hull of $\mathcal{T}$. An allocation function $f : \mathcal{T} \to \mathcal{O}$ is truthful if and only if there exists a truthful allocation function $f^\sharp : \mathcal{T}^\sharp \to \mathcal{O}$ such that f is the restriction of $f^\sharp$ to $\mathcal{T}$.*

Proof: If f is the restriction of a truthful allocation function $f^\sharp$ defined on $\mathcal{T}^\sharp$ then clearly f is truthful. Conversely, assume there exists a truthful mechanism (f, p). For every $\mathbf{x} \in \mathcal{T}$ we may define a function $\sigma_{\mathbf{x}} : \mathcal{T}^\sharp \to \mathbb{R}$ by $\sigma_{\mathbf{x}}(\mathbf{y}) = v(\mathbf{y}, f(\mathbf{x})) - p(\mathbf{x})$. The truthfulness of the mechanism (f, p) is equivalent to the assertion that $\sigma_{\mathbf{x}}(\mathbf{x}) \geq \sigma_{\mathbf{x}'}(\mathbf{x})$ for every $\mathbf{x}, \mathbf{x}' \in \mathcal{T}$.

Any $\mathbf{y} \in \mathcal{T}^\sharp$ can be written as a finite convex combination $\mathbf{y} = \sum_{i=1}^m w_i \mathbf{y}_i$, where $\mathbf{y}_i \in \mathcal{T}$. For all $\mathbf{x} \in \mathcal{T}$ we have

$$\sigma_{\mathbf{x}}(\mathbf{y}) = \sum_{i=1}^m w_i \sigma_{\mathbf{x}}(\mathbf{y}_i) \leq \sum_{i=1}^m w_i \sigma_{\mathbf{y}_i}(\mathbf{y}_i),$$

which implies that $\{\sigma_{\mathbf{x}}(\mathbf{y}) \,|\, \mathbf{x} \in \mathcal{T}\}$ is bounded above. Let $\sigma^*(\mathbf{y})$ be the supremum of this set, and let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be an infinite sequence in $\mathcal{T}$ such that $\sigma_{\mathbf{x}_n}(\mathbf{y}) \to \sigma^*(\mathbf{y})$ as $n \to \infty$.[5] Passing to an infinite subsequence if necessary, we can assume that $v_{f(\mathbf{x}_n)}$ converges, in $\mathbb{R}^{\mathcal{T}}$, to the function $v_{\mathbf{a}}$ for some $\mathbf{a} \in \mathcal{O}$.[6] (This step uses the assumption of outcome compactness.) Now define $f^\sharp(\mathbf{y})$ to be this outcome $\mathbf{a}$, and define a payment $p^\sharp(\mathbf{y}) = v(\mathbf{y}, f^\sharp(\mathbf{y})) - \sigma^*(\mathbf{y})$. We claim that $(f^\sharp, p^\sharp)$ is a truthful mechanism.

To prove the claim, consider any $\mathbf{y}, \mathbf{z} \in \mathcal{T}^\sharp$. Let $(\mathbf{x}_n)_{n=1}^\infty$ and $(\mathbf{x}'_n)_{n=1}^\infty$ be the sequences used in defining $f^\sharp(\mathbf{y})$ and $f^\sharp(\mathbf{z})$, respectively. Then we have $\sigma_{\mathbf{x}_n}(\mathbf{y}) \to \sigma^*(\mathbf{y})$ and $v(\mathbf{y}, f(\mathbf{x}_n)) \to v(\mathbf{y}, f^\sharp(\mathbf{y}))$ as $n \to \infty$, so

$$\begin{aligned} \lim_{n\to\infty} p(\mathbf{x}_n) &= \lim_{n\to\infty} v(\mathbf{y}, f(\mathbf{x}_n)) - \sigma_{\mathbf{x}_n}(\mathbf{y}) \\ &= v(\mathbf{y}, f^\sharp(\mathbf{y})) - \sigma^*(\mathbf{y}) \\ &= p^\sharp(\mathbf{y}). \end{aligned}$$

Similarly, $\lim_{n\to\infty} p(\mathbf{x}'_n) = p^\sharp(\mathbf{z})$. Finally,

$$\begin{aligned} v(\mathbf{y}, f^\sharp(\mathbf{z})) - p(\mathbf{z}) &= \lim_n v(\mathbf{y}, f(\mathbf{x}'_n)) - \lim_n p(\mathbf{x}'_n) \\ &= \lim_n \sigma_{\mathbf{x}'_n}(\mathbf{y}) \\ &\leq \sigma^*(\mathbf{y}) \\ &= v(\mathbf{y}, f^\sharp(\mathbf{y})) - p(\mathbf{y}), \end{aligned}$$

hence $(f^\sharp, p^\sharp)$ is truthful as claimed. ■

7. DOES MON(K) IMPLY CMON?

In the case where $\mathcal{O}$ is finite and $\mathcal{T}$ is convex, WMON implies truthfulness, by Theorem 4.1. We say that f satisfies MON(k) if $G_S(f)$ contains no negative cycles of k or fewer hops, so MON(2) is the same as WMON. By Theorem 2.2, we can recast the previous conclusion as saying MON(2) $\implies$ CMON. Saks and Yu [15] give a pair of examples where MON(2) holds but MON(3) fails. In one of their examples, the domain is convex, but the outcome space is infinite. In their other example, there are only three outcomes, but the domain is non-convex. This prompts the following question: if we relax either the finite range assumption, the convex domain assumption, or both, does there exist some k such that MON(k) implies CMON? The answer is no, as we illustrate below with three examples where $MON(k)$ holds but $MON(k+1)$ does not: a simple example with a non-convex domain *and* infinite outcome space, followed by more complicated examples with one of these properties but not the other (i.e., a convex domain and infinite outcome space, or a non-convex domain and finite outcome space).

7.1 Non-convex domain, infinite outcome space

In our simplest example, the domain $\mathcal{T}$ and outcome space $\mathcal{O}$ are both S^1, i.e., the unit circle in $\mathbb{R}^2$ centered at the origin. Thus, the outcome space is infinite and the domain is non-convex. In this section and the following ones, we will use the following notation for points of S^1: $\hat{\mathbf{e}}_\phi$ denotes the vector $\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix}$. We fix some $k \geq 2$, and pick any $\theta \in (\frac{\pi}{k+1}, \frac{\pi}{k}]$. The allocation function is simply rotation by $-\theta$ (i.e., θ in the clockwise direction), which we denote by $R_{-\theta}(\cdot)$. We claim that MON(k) holds, but MON($k+1$) does not.

First we show that $C_{k+1} = \hat{\mathbf{e}}_{\mathbf{0}} \to \hat{\mathbf{e}}_{\frac{2\pi}{\mathbf{k+1}}} \to \hat{\mathbf{e}}_{\mathbf{2}\cdot\frac{2\pi}{\mathbf{k+1}}} \to \cdots \to \hat{\mathbf{e}}_{\mathbf{k}\cdot\frac{2\pi}{\mathbf{k+1}}} \to \hat{\mathbf{e}}_{\mathbf{0}}$ is a negative $(k+1)$-cycle. In general, consider

[5] If $\mathbf{y} \in \mathcal{T}$, we will insist that this sequence is chosen to be $\mathbf{x}_1 = \mathbf{x}_2 = \ldots = \mathbf{y}$.

[6] If $\mathbf{x}_n = \mathbf{y}$ for all n, we insist that $\mathbf{a}$ is chosen to be outcome $f(\mathbf{y})$.

$v_1, v_2 \in S^1$, where $v_2 = R_\alpha(v_1)$. Then the S-weight of the edge $v_1 \to v_2$ is

$$(v_2 - v_1) \cdot R_{-\theta}(v_2) = v_2 \cdot R_{-\theta}(v_2) - v_1 \cdot R_{-\theta}(R_\alpha(v_1)) = \cos\theta - \cos(\alpha - \theta), \quad (11)$$

because the dot product between two unit vectors equals the cosine of the angle between them. If $\alpha - \theta \in (-\theta, \theta)$, i.e., $\alpha \in (0, 2\theta)$, then this weight is negative. For each hop in C, we have $0 < \alpha = \frac{2\pi}{k+1} < 2\theta$, so each hop has the same negative weight. Thus, C is a negative $(k+1)$-cycle.

Now we prove that no cycle of k or fewer hops has negative weight. Notice that the regular k-gon, oriented in the counterclockwise direction, has non-negative weight, by (11), since $\alpha = \frac{2\pi}{k} \geq 2\theta$. We proceed to prove that the cycle of k or fewer hops with the smallest weight is the regular k-gon.

Since $(S^1)^k$ is compact, there exists some cycle $C_k = v_0 \to v_1 \to \cdots \to v_{k-1} \to v_0$ that attains the minimum weight amongst all k-cycles. Since we can simulate a cycle with fewer than k hops by setting two or more adjacent v_i's to be equal, this cycle actually has the minimum weight amongst all cycles of k or fewer hops. Let us perturb C_k by varying v_1 while keeping all the other vertices fixed. The only edges whose weights change are $v_0 \to v_1$ and $v_1 \to v_2$. The sum of the weights on these two edges is

$$\begin{aligned}&(v_2 - v_1) \cdot R_{-\theta}(v_2) + (v_1 - v_0) \cdot R_{-\theta}(v_1)\\ &= v_2 \cdot R_{-\theta}(v_2) - v_1 \cdot R_{-\theta}(v_2)\\ &\quad + v_1 \cdot R_{-\theta}(v_1) - v_0 \cdot R_{-\theta}(v_1) \quad (12)\\ &= \cos\theta - v_1 \cdot R_{-\theta}(v_2) + \cos\theta - R_\theta(v_0) \cdot v_1 \quad (13)\\ &= 2\cos\theta - v_1 \cdot (R_\theta(v_0) + R_{-\theta}(v_2)), \quad (14)\end{aligned}$$

where we transformed the last term in (12) by applying the rotation R_θ to both vectors, which leaves their dot product unchanged. If $R_\theta(v_0) + R_{-\theta}(v_2) \neq 0$, then (14) is minimized when v_1 points in the same direction as $R_\theta(v_0) + R_{-\theta}(v_2)$, which means that v_1 must bisect the angle between v_0 and v_2. If $R_\theta(v_0) + R_{-\theta}(v_2) = 0$, then every choice of v_1 minimizes (14), yielding another k-cycle of minimum weight. But if we choose any value for v_1 such that $R_\theta(v_{k-1}) + R_{-\theta}(v_1) \neq 0$, and v_0 is not the angle bisector of v_{k-1} and v_1, then we can apply the previous argument to show that this cycle does not have the minimum weight. It is possible to choose such a v_1 unless $k = 2$ and $\theta = \frac{\pi}{2}$. In this one exceptional case, (14) shows that every choice of v_0, v_1 yields a cycle with weight 0. In all other cases, we have shown that the minimum k-cycle must be the regular k-gon. Examination of (14) shows that the counterclockwise orientation of the regular k-gon yields a larger magnitude for $R_\theta(v_0) + R_{-\theta}(v_2)$ and hence a smaller weight than the clockwise orientation.

7.2 Convex domain, infinite outcome space

Fix k and let $\theta \in (\frac{\pi}{k+1}, \frac{\pi}{k}]$, $\mathcal{T} = \mathcal{O} = \mathbb{R}^2$, and $f = R_{-\theta}$, i.e., rotation by angle θ in the clockwise direction.

It is easily verified that any regular $(k+1)$-gon centered at the origin and oriented in the counterclockwise direction has negative weight. We now prove the tricky part, that there are no negative k-cycles. We use exclusively S-weights in this section. The following result is easily derived by examining the behavior of dot products under affine transformations.

Proposition 7.1. *The weight of every cycle is invariant under translations and rotations. If we dilate a cycle by a scale factor $r > 0$, then the weight is multiplied by r^2.*

Corollary 7.2. *If there is a negative k-cycle, then there exists a negative k-cycle with one vertex at the origin and its predecessor at $\hat{\mathbf{e}}_1$.*

In the cycle given by Corollary 7.2, the arc $\hat{\mathbf{e}}_1 \to \mathbf{0}$ weighs $R_{-\theta}(\mathbf{0}) \cdot (-1, 0) = 0$. Thus, the weight of the lightest such k-cycle is equal to that of the lightest $(k-1)$-hop path from the origin to $\hat{\mathbf{e}}_1$. Let p_n be the weight of the lightest n-hop path from $\mathbf{0}$ to $\hat{\mathbf{e}}_1$. Since there is only a single 1-hop path to choose from, we have $p_1 = R_{-\theta}(\hat{\mathbf{e}}_1) \cdot (\hat{\mathbf{e}}_1 - \mathbf{0}) = \cos\theta$.

Lemma 7.3. *For $n \geq 1$ such that $p_n \geq 0$, we have $p_{n+1} = \cos\theta - \frac{1}{4p_n}$.*

Proof: We proceed by induction. The lightest $(n+1)$-hop path from $\mathbf{0}$ to $\hat{\mathbf{e}}_1$ consists of the lightest n-hop path from $\mathbf{0}$ to some type $\mathbf{x}$, followed by $\mathbf{x} \to \hat{\mathbf{e}}_1$. Using Proposition 7.1, the lightest n-hop path from $\mathbf{0}$ to $\mathbf{x}$ weighs $|\mathbf{x}|^2 p_n$, and the last hop has weight $R_{-\theta}(\hat{\mathbf{e}}_1) \cdot (\hat{\mathbf{e}}_1 - t) = \cos\theta - t \cdot R_{-\theta}(\hat{\mathbf{e}}_1)$. The total is minimized when $t = \frac{1}{2p_n} R_{-\theta}(\hat{\mathbf{e}}_1)$, yielding $p_{n+1} = \cos\theta - \frac{1}{4p_n}$. ■

Theorem 7.4. *For $n \leq k-1$, $p_n = \frac{\sin[(n+1)\theta]}{2\sin(n\theta)}$.*

Proof: We want to solve the recurrence $p_1 = a$, $p_{n+1} = a - \frac{b}{p_n}$ for $n \geq 1$, where $a = \cos\theta$, $b = \frac{1}{4}$. We will first transform this into a three-term linear recurrence. A closed-form solution will then be obtained using identities from the theory of orthogonal polynomials, specifically Chebyshev polynomials of the second kind [1, p. 776-7, 782]. As a first step, we multiply through by p_n, define $c_0 = 1$, $c_n = p_n c_{n-1}$ for $n \geq 1$, and substitute to yield the recurrence $c_{n+1} = ac_n - bc_{n-1}$, for $n \geq 1$. Now let $d_n = c_n b^{n/2}$ and substitute to get $d_{n+1} = \frac{a}{\sqrt{b}} d_n - d_{n-1}$. The Chebyshev polynomials of the second kind $U_n(x)$ satisfy the recurrence $U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x)$ and the initial conditions $U_0(x) = 1$ and $U_1(x) = 2x$. When $x = \frac{a}{2\sqrt{b}}$, these are the same recurrence and initial conditions that the d_n satisfy. Therefore, we have established that $d_n = U_n(\frac{a}{2\sqrt{b}}) = U_n(\cos\theta)$ for all $n \geq 0$. Using the identity $U_n(\cos\theta) = \frac{\sin[(n+1)\theta]}{\sin(\theta)}$, we get

$$p_n = \frac{c_n}{c_{n-1}} = \frac{d_n}{d_{n-1}}\sqrt{b} = \sqrt{b}\frac{U_n(\frac{a}{2\sqrt{b}})}{U_{n-1}(\frac{a}{2\sqrt{b}})} = \frac{\sin[(n+1)\theta]}{2\sin(n\theta)}. \quad (15)$$

The recurrence for p_{n+1} is valid as long as $p_n \geq 0$, so (15) holds up to and including the first index where $p_n < 0$. Since $\theta \in (\frac{\pi}{k+1}, \frac{\pi}{k}]$, this occurs at $n = k$. ■

Corollary 7.5. *For $\theta \in (\frac{\pi}{k+1}, \frac{\pi}{k}]$, the allocation function $f = R_{-\theta}$ satisfies MON(k) , but not MON($k+1$).*

7.3 Non-convex domain, finite outcome space

In this example, $\mathcal{T}$ is once again equal to S^1, the unit circle centered at the origin in R^2. We divide the circle into $2k$ intervals of angle π/k each; interval $I_j (j = 0, \ldots, 2k-1)$ runs from $\hat{\mathbf{e}}_{\frac{j\pi}{k}}$ to $\hat{\mathbf{e}}_{\frac{(j+1)\pi}{k}}$. We will always interpret the interval subscripts mod $2k$. For each $t \in I_i$, we define $f(t) = i\pi/k$. We will show that f satisfies MON(k) but not MON($k+1$).

This time it will be more convenient to reason using P-weights, rather than S-weights. Recall that the P-weight of an arc from type u to v is $u \cdot (f(u) - f(v))$. We can label each arc $u \to v$ by its number of *clicks*, i.e., the number of segments from the one containing $f(u)$ to the one containing $f(v)$, which we will always take to be in $\{-k+1, \ldots, k\}$. Thus, by rotational symmetry, the P-weight of arc $u \to v$ depends only on the number of clicks and where u lies in its interval. We now compute $m(j)$, the infimum of the weights amongst all j-click arcs:

$$m(j) = \inf_{\alpha \in (0,\pi/k]} \hat{\mathbf{e}}_\alpha \cdot (\hat{\mathbf{e}}_\mathbf{0} - \hat{\mathbf{e}}_{\mathbf{j}\pi/\mathbf{k}}) \tag{16}$$

$$= \inf_{\alpha \in (0,\pi/k]} \cos\alpha - \cos(j\pi/k - \alpha) \tag{17}$$

$$= \begin{cases} \cos(\pi/k) - \cos((j-1)\pi/k) & \text{for } 0 \leq i \leq k \\ 1 - \cos(j\pi/k) & \text{for } i < 0 \end{cases} \tag{18}$$

Thus, we see that $m(1) < 0$, $m(0) = m(2) = 0$, and $m(j) > 0$ for all other j. From now on, every j-click arc we refer to has weight equal to or arbitrarily close to $m(j)$. Hence, a $(k+1)$-cycle consisting of $(k-1)$ 2-click arcs and two 1-click arcs has negative weight. We now show that every k-cycle has non-negative weight.

Consider the most negative k-cycle C, and suppose it has strictly negative weight. Any combination of clicks is realizable by some cycle, so long as the clicks sum to zero mod $2k$. The most negative k-cycle cannot contain arcs of $-i$ clicks and j clicks (with $i, j > 0$), because we could decrease the weight by replacing them with arcs of $-i+1$ clicks and $j-1$ clicks. Since $m(i) > 0$ for $i < 0$, C must have only arcs with positive clicks. Some arc in C must have only 1 click, since C has negative weight. But if there is a 1-click arc and an i-click arc with $i \geq 3$, then we can decrease the weight by replacing these with a 2-click arc and an $(i-1)$-click arc. Hence, C must contain only arcs of 0, 1 or 2 clicks. But then it is impossible for the clicks to sum to zero mod $2k$, given that at least one of them has 1 click. This leads to a contradiction, so there are no negative k-cycles.

Acknowledgments

We thank Neil Sloane and David Applegate for introducing us to the joy of orthogonal polynomials, and we thank Ron Lavi for suggesting the question of whether $MON(k)$ implies $CMON$. Finally, we thank Rakesh Vohra for helpful discussions relating to our characterization theorem.

8. REFERENCES

[1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions*. Dover, New York, December 1972.

[2] Aaron Archer and Éva Tardos. Truthful mechanisms for one-parameter agents. In *Proceedings of the* 42^{nd} *IEEE Symposium on Foundations of Computer Science*, pages 482–491, 2001.

[3] Sushil Bikhchandani, Shurojit Chatterji, Ron Lavi, Ahuva Mu'alem, Noam Nisan, and Arunava Sen. Weak monotonicity characterizes incentive deterministic dominant strategy implementation. *Econometrica*, 74(4):1109–1132, July 2006.

[4] E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 8:17–33, 1971.

[5] Theodore Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[6] Hongwei Gui, Rudolf Müller, and Rakesh Vohra. Dominant strategy mechanisms with multidimensional types. Research Memorandum 47, METEOR, Maastricht Research School of Economics of Technology and Organization, Maastricht, 2004.

[7] R. Preston McAfee and John McMillan. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, 46:335–354, 1988.

[8] James Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971.

[9] Dov Monderer. Monotonicity and implementability. In *Proceedings of the* 10^{th} *ACM Conference on Electronic Commerce*, 2008.

[10] Rudolf Müller, Andrés Perea, and Sascha Wolf. A network approach to bayes-nash incentive compatible mechanisms. *Games and Economic Behavior*, 61(2):344–358, November 2007.

[11] James R. Munkres. *Topology*. Prentice Hall, 1975.

[12] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, February 1981.

[13] Kevin Roberts. The characterization of implementable choice rules. In Jean-Jacques Laffont, editor, *Aggregation and Revelation of Preferences*, pages 321–348. North-Holland, Amsterdam, 1979.

[14] Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, April 1987.

[15] Michael E. Saks and Lan Yu. Weak monotonicity suffices for truthfulness on convex domains. In *Proceedings of the* 7^{th} *ACM Conference on Electronic Commerce*, pages 286–293, 2005.

[16] Michael Spence. Competitive and optimal responses to signals. *Journal of Economic Theory*, 7:196–232, 1974.

[17] William Vickrey. Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[18] Rakesh Vohra. Paths, cycles and mechanism design. Working paper, 2007.