A Dynamic Rationalization of Distance Rationalizability

Craig Boutilier
Department of Computer Science
University of Toronto
cebly@cs.toronto.edu

Ariel D. Procaccia
Computer Science Department
Carnegie Mellon University
arielpro@cs.cmu.edu

Abstract

Distance rationalizability is an intuitive paradigm for developing and studying voting rules: given a notion of *consensus* and a *distance function* on preference profiles, a *rationalizable* voting rule selects an alternative that is closest to being a consensus winner. Despite its appeal, distance rationalizability faces the challenge of connecting the chosen distance measure and consensus notion to an *operational* measure of social desirability. We tackle this issue via the decision-theoretic framework of *dynamic social choice*, in which a *social choice Markov decision process (MDP)* models the dynamics of voter preferences in response to winner selection. We show that, for a prominent class of distance functions, one can construct a social choice MDP, with natural preference dynamics and rewards, such that a voting rule is (vote-wise) rationalizable with respect to the unanimity consensus for a given distance function iff it is a (deterministic) optimal policy in the MDP. This provides an alternative rationale for distance rationalizability, demonstrating the equivalence of rationalizable voting rules in a static sense and winner selection to maximize societal utility in a dynamic process.

1 Introduction

Social choice theory—the study of mechanisms for aggregating the preferences of individual agents to determine a suitable consensus outcome—has attracted considerable attention in computer science and AI. It offers various models and mechanisms for coordinating autonomous agent activities; for modeling user interactions; and for group-based decision support. The standard social choice setting includes a set of agents and a set of alternatives. Each agent's vote is a ranking of the alternatives; the collection of the agents' votes is called a *preference profile*. A *voting rule* takes a preference profile as input, and outputs a single alternative as the winner of the election. Decades of research has led to the design of a variety of interesting voting rules, but judging their relative desirability is still the subject of much debate.

In a recent paper, Meskanin and Nurmi (2008) introduced the *distance rationalizability (DR)* framework, building on ideas of 19th century mathematician and author Charles Dodgson (better known as Lewis Carroll). Their key insight is that a voting rule should select an alternative that is the "ideal" candidate in a profile satisfying some notion of *consensus* that is "closest" to the current preference profile, where proximity is measured by a distance function over profiles. For specific notions of consensus, there is general agreement on what constitutes an ideal candidate, so one simply needs to select the appropriate notion of consensus and distance measure to construct a voting rule. Therein lies the attraction of DR: it recasts the debate surrounding the suitability of various voting rules—which traditionally relies on appeal to normative axioms—to one regarding, arguably, more fundamental ingredients. The DR framework has been studied in both social choice (Nurmi 2004; Meskanen and Nurmi 2008) and AI (Elkind, Faliszewski, and Slinko 2009; 2010a; 2010b).

While DR has much to recommend it, it provides no explicit measure of the "societal benefit" associated with a winner, nor does it directly define what it means for one candidate to be "closer" to being a winner than another, except insofar as current societal *preferences* are closer to a profile that supports the first candidate. If we take the notion of *closeness over profiles* literally, one interpretation of the DR framework is to treat winner selection as influencing the preferences of individual voters in a way that *moves* society toward consensus (hence rationalizable rules move society toward the closest such consensus profile). Indeed, most expositions of distance rationalizability informally explain distance measures using phrases like "minimum number of preference changes of voters" (Nurmi 2004). However, distances do not themselves provide an account of the *preference dynamics* that give rise to such changes. In this work, we propose an explicit model of preference dynamics and societal utility for alternatives that supports certain forms of DR. Exploiting the framework of *dynamic social choice* introduced by Parkes and Procaccia (2010)—specifically, their *social choice Markov decision processes (MDPs)*—we develop a model in which individual voter preferences can change in response to winner selection. Within our social choice MDP, an optimal policy is one that maximizes accrued reward (or societal utility) over time.

By aligning this reward function with standard notions of consensus, and making certain assumptions about preference dynamics that directly reflect a distance metric of interest, one might hope to justify specific forms of DR in terms of dynamic social choice. Indeed, we take a step

in this direction with respect to an important class of distance rationalizable voting rules. Specifically, we define a reward function—aligned with the consensus notion *unanimity*—and a class of preference dynamics—tailored to *any* (votewise) distance function (Elkind, Faliszewski, and Slinko 2010a)—that determine a social choice MDP whose optimal policy (i.e., selection of winner for any configuration of voter preferences) is exactly the voting rule given by DR (specifically, using ℓ_1-votewise DR under unanimity).

Our contributions are several. First, our main result offers a new perspective on a particular class of voting rules. More broadly, our use of the dynamic social choice framework provides an alternative interpretation of the DR paradigm, one that views distances over profiles directly in terms of the dynamic evolution of societal preferences, and consensus notions as a surrogate for societal utility. Optimal policies (or voting rules) in our model make appropriate tradeoffs between "nudging" society toward consensus (long-term utility) and immediate utility. Finally, apart from the novel perspective that the AI-inspired dynamic social choice framework brings to social choice, we suspect that the construction of policies governing the dynamic consensus choices of groups of individuals, whose preferences change over time in response to these choices, will be an important component of decision support and recommender systems in a variety of domains (social networks, groupware, etc.). Our framework may justify the use of specific *static* voting rules in such *dynamic* contexts.

The paper is organized as follows: in Sec. 2 we briefly give the necessary formal background on distance rationalizability and dynamic social choice. In Sec. 3 we formulate and prove our main result, and discuss some of its mathematical implications. We discuss more conceptual consequences of our approach and our result in Sec. 4.

2 Preliminaries

The standard social choice setting includes a set of *agents*, which we denote by $N = \{1, \ldots, n\}$, and a set of *alternatives* A, $|A| = m$; we denote specific alternatives by lower case letters x, y, etc. Each agent $i \in N$ holds preferences over A represented by a linear order $P_i \in \mathcal{L}$, where $\mathcal{L} = \mathcal{L}(A)$ is the set of all linear orders over A. A vector $\vec{P} = (P_1, \ldots, P_n) \in \mathcal{L}^n$, which specifies the preferences of all agents, is called a *preference profile*. A *voting rule* $f : \mathcal{L}^n \to A$ receives a preference profile as input, and returns an alternative that is thereby designated as most desirable. Given $P \in \mathcal{L}$, let $P[k]$ denote the alternative that is ranked k'th in P. For $\vec{P} \in \mathcal{L}^n$, $x \in A$, we denote

$$\text{top}(\vec{P}, x) = \{i \in N : P_i[1] = x\}.$$

We also denote by $P[x]$ the position in which alternative x is ranked in P.

Positional scoring rules are voting rules that are induced by a *score vector* of integers $\vec{\alpha} = (\alpha_1, \ldots, \alpha_m)$, where $\alpha_k \geq \alpha_{k+1}$ for all k. For each $k = 1, \ldots, m$, agent $i \in N$ awards α_k points to the alternative $P_i[k]$; an alternative with the most points is selected by the positional scoring rule. Note that the same $\vec{\alpha}$ can induce multiple positional scoring rules due to tie breaking. Well-known positional scoring rules include *plurality*, which is induced by the vector $(1, 0, \ldots, 0)$, and *Borda count*, which is induced by the vector $(m-1, m-2, \ldots, 0)$.

2.1 Distance rationalizability

Let Q be a set. A function $d : Q \times Q \to \mathbb{R}_+$ is a *distance* if the following conditions hold for all $p, q, r \in Q$:

1. *Identity of Indiscernibles*: $d(p, q) = 0 \Leftrightarrow p = q$.
2. *Symmetry*: $d(p, q) = d(q, p)$.
3. *Triangle Inequality*: $d(p, r) \leq d(p, q) + d(q, r)$.

A *pseudodistance* replaces Identity of Indiscernibles with the weaker requirement $d(q, q) = 0$ (i.e., we may have $d(q, r) = 0$ for $q \neq r$). A *quasidistance* satisfies the first and third axioms, (i.e., $d(p, q) \neq d(q, p)$ is permitted). For $R \subseteq Q$ and $q \in Q$ we define $d(q, R) = \min_{r \in R} d(q, r)$. For convenience, we assume any distance function has range $\mathbb{N} \cup \{0\}$ (this is without loss of generality because the domain of these functions will be finite in what follows).

The *distance rationalizability (DR)* framework requires two components: a distance $\tilde{d}$ over the space of preference profiles $\mathcal{L}^n$; and a notion of consensus (Meskanen and Nurmi 2008). Although the framework can support any notion of consensus, we focus on the well-studied notion of *unanimity*. We say that an alternative $x \in A$ is a *unanimous winner* in $\vec{P} \in \mathcal{L}^n$ if $|\text{top}(\vec{P}, x)| = n$ (i.e., all agents rank x first). Let

$$\mathcal{U}_x = \{\vec{P} \in \mathcal{L}^n : |\text{top}(\vec{P}, x)| = n\}$$

be the set of preference profiles where x is a unanimous winner. A voting rule *satisfies unanimity* if it selects a unanimous winner whenever given a profile where one exists.

A voting rule f is *distance rationalizable w.r.t. unanimity via a distance* $\tilde{d}$ if, for every $\vec{P} \in \mathcal{L}^n$, x minimizes $\tilde{d}(\vec{P}, \mathcal{U}_x)$ whenever $f(\vec{P}) = x$. In other words, if f is rationalized by $\tilde{d}$, it must select, in any profile $\vec{P}$, the unanimous winner in the $\tilde{d}$-closest profile $\vec{P}_{\mathcal{U}}$ that has a unanimous winner.

Previous work has considered voting *correspondences* that return *all* alternatives that are closest to being consensus winners. For ease of exposition, we only consider voting rules that select one optimal alternative for each given profile. However, our main result holds for any voting rule that selects winners from the above correspondence, and hence is completely independent of the tie-breaking method adopted.

We illustrate the concept of distance rationalizability using the common *plurality rule*, mentioned above, which selects an alternative $x \in A$ that maximizes $|\text{top}(\vec{P}, x)|$. Define the *Hamming distance* $\tilde{d}_H$ between profiles to be:

$$\tilde{d}_H(\vec{P}, \vec{P}') = |\{i \in N : P_i \neq P'_i\}|$$

(i.e., the number of agents that have different preferences under $\vec{P}$ and $\vec{P}'$). It is not hard to see that plurality is distance rationalizable with respect to unanimity via $\tilde{d}_H$: suppose alternative x is ranked first by k agents in a profile $\vec{P}$

(the plurality score of x is k). By switching the votes of all agents that do not rank x first in $\vec{P}$ to arbitrary rankings with x on top, we obtain a new profile $\vec{P}_x$ where x is a unanimous winner. The Hamming distance between the $\vec{P}$ and $\vec{P}_x$ is $n-k$, and clearly $\vec{P}_x$ is a closest profile to $\vec{P}$ in which x is a unanimous winner. Therefore, the larger the plurality score k of an alternative, the smaller the Hamming distance $n-k$ to some profile satisfying unanimity.

Elkind et al. (2010a) show that any voting rule that satisfies unanimity can be distance rationalized w.r.t. unanimity via *some* distance function. This severely limits the scope of DR as a normative concept. They therefore suggest focusing on *votewise distance rationalizability*. A function $\tilde{d} : \mathcal{L}^n \times \mathcal{L}^n \to \mathbb{N} \cup \{0\}$ is an ℓ_1-*votewise distance* if there exists a distance function d on $\mathcal{L}$ (i.e., that takes individual votes as input) such that

$$\tilde{d}(\vec{P}, \vec{P}') = \sum_{i=1}^{n} d(P_i, P_i')$$

(we say $\tilde{d}$ is the ℓ_1-votewise distance induced by d). Note that Hamming distance is an ℓ_1-votewise distance. A voting rule is ℓ_1-*votewise distance rationalizable* w.r.t. unanimity via d iff it is distance rationalizable w.r.t. unanimity via the ℓ_1-votewise distance induced by d. Elkind et al. (2010a) also consider norms apart from ℓ_1, but the above definitions suffice for our purposes.

2.2 Dynamic social choice

The *dynamic social choice* framework (Parkes and Procaccia 2010) can be used to model settings in which a set of agents have dynamically changing preferences over a set of alternatives. At the heart of this framework is the concept of a *social choice MDP*; this will prove to be a valuable tool in our analysis of DR. We first discuss MDPs more generally, then describe social choice MDPs.

A (finite) *Markov decision process (MDP)* $\mathcal{M} = (\mathcal{S}, A, R, T)$ comprises: a finite set of states $\mathcal{S}$; a finite action set A; a reward function $R : \mathcal{S} \times A \to \mathbb{R}$, where $R(s,a)$ is the reward obtained when action a is taken in state s; and transition function T, where $T(s'|s,a)$ is the probability of moving to state s' when action a is taken in state s. A *deterministic (stationary) policy* is a function $\pi : \mathcal{S} \to A$, specifying which action to take in each state. We assume a discounted, infinite-horizon model, where our objective is to select a policy π that maximizes the expected discounted sum of rewards $E(\sum_{t=0}^{\infty} \gamma^t R^t | \pi)$ w.r.t. discount factor $\gamma \in [0,1)$ (here R^t denotes reward received at time t, and expectation is taken w.r.t. the distribution over state-action trajectories induced by π). The *value* of policy π is given by a value function $V^\pi : \mathcal{S} \to \mathbb{R}$, where $V^\pi(s)$ is this expected sum of rewards starting at $S^0 = s$. Deterministic optimal policies exist and any optimal policy π^* satisfies the Bellman equations:

$$V^{\pi^*}(s) = \max_{a \in A} \left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} T(s'|s,a) \cdot V^{\pi^*}(s') \right], \ \forall s \in \mathcal{S}.$$

See Puterman (1994) for further background on MDPs.

Now assume some social choice setting with agents $N = \{1, \ldots, n\}$ and *alternatives* A. In a *social choice MDP* (Parkes and Procaccia 2010), the state set $\mathcal{S}$ is the set $\mathcal{L}^n$ of preference profiles over A, and the action set is just the set A of alternatives. Intuitively, state transitions reflect the changing preferences of the agents, and an action a taken at state (or profile) s represents the selection of a winning alternative. Crucially, this means that a deterministic policy in a social choice MDP coincides with a voting rule: it selects an action (alternative) given a state (preference profile).[1] Notice that actions, i.e., selection of winners, can influence how agent preferences evolve. Let s_i be the preferences held by agent i in state s. In the sequel we use the terms "state" and "preference profile", the terms "policy" and "voting rule", and their corresponding notation, interchangeably. We also focus on social choice MDPs with *independent agent transition functions*. In particular, for each $P, P' \in \mathcal{L}$ and $x \in A$, let $p(P'|P,x)$ be the probability of an agent with preferences P moving to P' when action x is taken. Then

$$T(s'|s,x) = \prod_{i \in N} p(s_i'|s_i, x).$$

In other words, each agent transitions independently according to p.[2] Conceptually, when alternative x is selected as a winner, each agent's preferences evolve stochastically (and independently) to reflect its revised beliefs about the relative quality of the alternatives.

While we draw on some of the basic technical framework of Parkes and Procaccia (2010), we require none of their technical results. More importantly, conceptually our motivation is quite different. Their approach is *constructive/algorithmic*: they automatically construct decision policies that perform well in dynamic settings (e.g., public policy advocacy), designing algorithms that compute optimal policies for given social choice MDPs. Implementation of their approach in practice faces several obstacles: most crucially it may be difficult, or even impossible, to obtain the agent transition models. In contrast, our approach is *descriptive*: we use a known social choice MDP, with a simple optimal policy, to describe and interpret a distance rationalizable rule (rather than taking a social choice MDP as input to some algorithm).

3 Interpreting Distance Rationalizability Via Dynamic Social Choice

We now develop a social choice MDP that will allow us to interpret specific forms of distance rationalizable voting rules in a dynamic fashion. In particular, we show below that rules that are ℓ_1-votewise distance rationalizable with respect to unanimity, via *any* underlying distance function d over votes, are precisely the optimal policies in a social choice MDP defined relative to d.

The two components that need to be specified to define our MDP are the reward function and the transition function.

[1] We restrict our attention to stationary policies (which are independent of the history); here this restriction plays a crucial role.

[2] Since all agents have the same transition model, in the terminology of Parkes and Procaccia (2010) they have the same *type*.

The properties of the transition function will be detailed in the proof of Thm. 1 below, but intuitively, when an alternative x is selected, an agent's ranking P will transition with a particular probability to the d-closest ranking at which x is top-ranked (otherwise it is unchanged).

Our reward function is the *plurality reward function* R^*, where $R^*(s,x) = |\text{top}(s,x)|$. This is one rather natural measure of (immediate) societal utility: an agent contributes to overall societal utility at any stage of the process if the (current) winner is its (currently) top-ranked choice. Furthermore, in our dynamic context, suppose we attempt to "steer" society towards a consensus profile where x is a unanimous winner. Since each agent contributes a (discounted) utility of one per stage if they rank x first (and zero otherwise), the faster all agents agree on x, the higher the total reward; in this sense, R^* sends a stronger positive signal the closer society is to unanimity (consensus). A technical advantage of the plurality reward function, compared to a coarser reward function that provides a positive reward when x is a unanimous winner and zero reward otherwise, is that it separates the agents and hence facilitates the mathematical analysis.

Our main result is the following theorem. Its proof details the construction of the required social choice MDP, and demonstrates more than the statement of the theorem (as we discuss below).

Theorem 1. *Let $d : \mathcal{L} \times \mathcal{L} \to \mathbb{N} \cup \{0\}$ be a distance function on votes and $D = \max_{P,P' \in \mathcal{L}} d(P,P')$ be an upper bound on d. Assume a discount factor satisfying $\gamma > 1 - 1/D$.[3] Then there is a social choice MDP $\mathcal{M}(d) = (\mathcal{S}, A, R^*, T(d))$ such that voting rule π is ℓ_1-votewise distance rationalizable w.r.t. unanimity via d if and only if π is an optimal deterministic policy for $\mathcal{M}(d)$.*

Proof. We only need to specify the transition function $T(d)$. We construct transitions such that an agent i holding a ranking P, when action $x \in A$ is selected, will either transition to a ranking denoted $S(P,x)$ which is the d-closest ranking with x ranked first, or will maintain its current ranking P. More formally, for each $P \in \mathcal{L}$, let $S(P,x) \in \text{argmin}_{P':\ P'[1]=x} d(P,P')$. Note, if x is ranked first in P, then $S(P,x) = P$.

Define $\hat{d} : \mathcal{L} \times A$ by $\hat{d}(P,x) = d(P, S(P,x))$. We first show that $\hat{d}$ satisfies a variant of the triangle inequality. Specifically, for every $P \in \mathcal{L}$, $x, y \in A$,

$$\hat{d}(P,x) \leq \hat{d}(P,y) + \hat{d}(S(P,y),x). \tag{1}$$

Indeed, it holds that

$$\begin{aligned}&\hat{d}(P,y) + \hat{d}(S(P,y),x)\\ &\quad = d(P,S(P,y)) + d(S(P,y),S(S(P,y),x)).\end{aligned}$$

Hence, by the triangle inequality for d,

$$d(P,S(S(P,y),x)) \leq \hat{d}(P,y) + \hat{d}(S(P,y),x).$$

But in $S(S(P,y),x)$ alternative x is ranked first, and by definition $\hat{d}(P,x) \leq d(P,S(S(P,y),x))$. This establishes (1).

[3]This assumption is mild: the discount factor can be bounded away from 1 by a constant that depends only on D, rather than approaching one as the number of agents or alternatives grows.

We next show, given $s \in \mathcal{S}$ and $x \in A$, that

$$\min_{s' \in \mathcal{U}_x} \sum_{i \in N} d(s_i, s'_i) = \sum_{i \in N} \hat{d}(s_i, x). \tag{2}$$

Indeed,

$$\sum_{i \in N} \hat{d}(s_i,x) = \sum_{i \in N} d(s_i, S(s_i,x)),$$

and $(S(s_1,x), \ldots, S(s_n,x)) \in \mathcal{U}_x$, therefore

$$\min_{s' \in \mathcal{U}_x} \sum_{i \in N} d(s_i, s'_i) \leq \sum_{i \in N} \hat{d}(s_i,x)$$

The inequality in the other direction follows by noting that for each $s' \in \mathcal{U}_x$ and $i \in N$, $s'_i[1] = x$, and hence

$$d(s_i, s'_i) \geq d(s_i, S(s_i,x)) = \hat{d}(s_i,x).$$

We can therefore conclude that a voting rule that is ℓ_1-votewise distance rationalized by d must select alternatives that minimize the expression on the right hand side of (2).

We now turn to constructing the transition function $T(d)$ of $\mathcal{M}(d)$. We define $\hat{p}(P,x)$ and set $p(S(P,x)|P,x) = \hat{p}(P,x)$ and $p(P|P,x) = 1 - \hat{p}(P,x)$ if $P[1] \neq x$, and $p(P|P,x) = 1$ otherwise. Informally, we wish to define $\hat{p}(P,x)$ such that, if the action x is selected repeatedly, the expected "cost" — in the sense of lost reward — that we pay for the agent's preferences to transition is exactly $\hat{d}(P,x)$. More precisely, For $P \in \mathcal{L}$ and $x \in A$ such that $P[1] \neq x$, we want $\hat{p}(P,x)$ to satisfy:

$$\begin{aligned}\hat{d}(P,x) &= \sum_{k=0}^{\infty} \left[\hat{p}(P,x)(1-\hat{p}(P,x))^k \left(\sum_{l=0}^{k} \gamma^l \right) \right]\\ &= \sum_{k=0}^{\infty} \hat{p}(P,x)(1-\hat{p}(P,x))^k \frac{1-\gamma^{k+1}}{1-\gamma}\\ &= \frac{\hat{p}(P,x)}{1-\gamma} \left(\sum_{k=0}^{\infty} (1-\hat{p}(P,x))^k - \gamma \sum_{k=0}^{\infty} ((1-\hat{p}(P,x))\gamma)^k \right)\\ &= \frac{\hat{p}(P,x)}{1-\gamma} \left(\frac{1}{\hat{p}(P,x)} - \gamma \cdot \frac{1}{1-(1-\hat{p}(P,x))\gamma} \right)\\ &= \frac{1}{1-\gamma} \left(1 - \frac{\hat{p}(P,x)\gamma}{1-\gamma+\gamma\hat{p}(P,x)} \right).\end{aligned}$$

Solving for $\hat{p}(P,x)$, we obtain

$$\hat{p}(P,x) = \frac{1}{\gamma} \left(\frac{1}{\hat{d}(P,x)} - (1-\gamma) \right). \tag{3}$$

Our assumption that $\gamma > 1 - 1/D$ ensures $\hat{p}(P,x) \in (0,1)$.

The rest of the proof establishes that $\mathcal{M}(d)$ is as stated in the theorem, i.e., that the set of deterministic optimal policies contains exactly the policies that in state s select an alternative $x \in A$ that (using (2)) minimizes $\sum_{i \in N} \hat{d}(s_i,x)$. We first claim that, if π^* is such a policy, then

$$V^{\pi^*}(s) = \frac{n}{1-\gamma} - \min_{x \in A} \sum_{i \in N} \hat{d}(s_i,x). \tag{4}$$

The intuition behind (4) is simple. If x is the unanimous winner in state s, the sum of discounted rewards is:

$$n \cdot \sum_{k=0}^{\infty} \gamma^k = \frac{n}{1-\gamma}.$$

However, given the way we constructed the transition probabilities in $\mathcal{M}(d)$ we pay an expected cost of $\hat{d}(s_i, x)$ for each agent $i \in N$ that does not rank x first; subtracting these costs from $n/(1-\gamma)$ gives (4).

To formally establish the correctness of (4), let

$$a \in \text{argmin}_{x \in A} \sum_{i \in N} \hat{d}(s_i, x).$$

We must verify that

$$\begin{aligned} V^{\pi^*}(s) &= R(s,a) + \gamma \sum_{s' \in \mathcal{S}} \left(p(s'|s,a) \cdot V^{\pi^*}(s') \right) \\ &= |\text{top}(s,a)| + \gamma \sum_{s' \in \mathcal{S}} \left(p(s'|s,a) \cdot V^{\pi^*}(s') \right). \end{aligned} \tag{5}$$

We argue that if a is selected in s then a also minimizes the sum of distances in every possible next state s'. On the one hand, if agent i transitions when a is selected, we have:

$$0 = \hat{d}(s_i', a) = \hat{d}(s_i, a) - \hat{d}(s_i, a),$$

that is, the sum of distances decreases by $\hat{d}(s_i, a)$. On the other hand, by (1), for any $x \in A$, we have:

$$\hat{d}(s_i', x) \geq \hat{d}(s_i, x) - \hat{d}(s_i, a),$$

because $s_i' = S(s_i, a)$. In other words, the sum of distances to x decreases by at most $\hat{d}(s_i, a)$. Therefore, it is sufficient to consider the expected sum of distances $\sum_{i \in N} \hat{d}(s_i', a)$ for the next state s':

$$\begin{aligned} &\sum_{s' \in \mathcal{S}} \left(p(s'|s,a) \cdot V^{\pi^*}(s') \right) \\ &= \frac{n}{1-\gamma} - \sum_{i \notin \text{top}(s,a)} (1 - \hat{p}(s_i, a)) \hat{d}(s_i, a) \\ &= \frac{n}{1-\gamma} - \frac{1}{\gamma} \sum_{i \notin \text{top}(s,a)} \left(1 - \frac{1}{\hat{d}(s_i, a)} \right) \cdot \hat{d}(s_i, a) \\ &= \frac{n}{1-\gamma} - \frac{1}{\gamma} \sum_{i \in N} \hat{d}(s_i, a) + \frac{n - \text{top}(s,a)}{\gamma}, \end{aligned}$$

where the second equality follows from (3), and the third holds since $\hat{d}(s_i, a) = 0$ for all $i \in \text{top}(s,a)$. Thus, as required, the right hand side of (5) equals:

$$n + \frac{\gamma}{1-\gamma} \cdot n - \sum_{i \in N} \hat{d}(s_i, a) = \frac{n}{1-\gamma} - \sum_{i \in N} \hat{d}(s_i, a).$$

By the Bellman optimality equations, to show that *only* policies minimizing $\sum_{i \in N} \hat{d}(s_i, x)$ in every state are optimal, we must show that for every $b \notin \text{argmin}_{x \in A} \sum_{i \in N} \hat{d}(s_i, x)$,

$$V^{\pi^*}(s) > |\text{top}(s,b)| + \gamma \sum_{s' \in \mathcal{S}} \left(p(s'|s,b) \cdot V^{\pi^*}(s') \right). \tag{6}$$

Indeed, above we have established that $V^{\pi^*}(s)$ is the optimal value in state s. Equation (6) states that for any b that does not minimize the sum of distances, the value of choosing b as a winner and then acting optimally is *strictly* smaller than the value of acting optimally. Hence any policy that chooses such an alternative b violates the fixed point requirement of Bellman optimality.

To upper-bound $\sum_{s' \in \mathcal{S}} \left(p(s'|s,b) \cdot V^{\pi^*}(s') \right)$, it suffices to lower-bound $\mathbb{E}[\min_{x \in A} \sum_{i \in N} \hat{d}(s_i', x)]$, given that b is selected in state s. Note that agents $i \in \text{top}(s,b)$ do not transition. For agents $i \notin \text{top}(s,b)$ it holds by (1) that $\hat{d}(s_i', x) \geq \hat{d}(s_i, x) - \hat{d}(s_i, b)$, where $s_i' = S(s_i, b)$. That is, for each agent i that transitions, the sum of distances to each $x \in A$ can decrease by at most $\hat{d}(s_i, b)$. To derive a lower bound, we start from the minimum sum of distances $\sum_{i \in N} \hat{d}(s_i, a)$ and decrease it by $\hat{d}(s_i, b)$ for every agent $i \in N$ that transitions. Therefore, we can upper-bound the right hand side of (6) by

$$\begin{aligned} &|\text{top}(s,b)| + \gamma \left(\frac{n}{1-\gamma} - \left(\sum_{i \in N} \hat{d}(s_i, a) - \sum_{i \notin \text{top}(s,b)} \hat{p}(s_i, b) \hat{d}(s_i, b) \right) \right) \\ &= |\text{top}(s,b)| + n \frac{\gamma}{1-\gamma} - \gamma \sum_{i \in N} \hat{d}(s_i, a) \\ &\qquad + \sum_{i \notin \text{top}(s,b)} \left(1 - (1-\gamma) \hat{d}(s_i, b) \right) \\ &= |\text{top}(s,b)| + (n - |\text{top}(s,b)|) + n \frac{\gamma}{1-\gamma} - \gamma \sum_{i \in N} \hat{d}(s_i, a) \\ &\qquad - (1-\gamma) \sum_{i \notin \text{top}(s,b)} \hat{d}(s_i, b) \\ &= \frac{n}{1-\gamma} - \gamma \sum_{i \in N} \hat{d}(s_i, a) - (1-\gamma) \sum_{i \in N} \hat{d}(s_i, b) \\ &< \frac{n}{1-\gamma} - \sum_{i \in N} \hat{d}(s_i, a) \\ &= V^{\pi^*}(s). \end{aligned}$$

The third equality follows from the fact $\hat{d}(s_i, b) = 0$ for every $i \in \text{top}(s,b)$, while the fourth inequality is implied by the assumption that $\sum_{i \in N} \hat{d}(s_i, a) < \sum_{i \in N} \hat{d}(s_i, b)$. □

Thus, given any ℓ_1-votewise distance rationalizable rule r with respect to unanimity via some distance d over votes, a social choice MDP can be constructed whose optimal policies correspond to r.

Eq. (4) and its proof can easily be adapted to yield a stronger claim: in $\mathcal{M}(d)$, when starting from state $s \in \mathcal{S}$, the value of the policy that repeatedly selects action $x \in A$ is

$$\frac{n}{1-\gamma} - \sum_{i \in N} \hat{d}(s_i, x) = \frac{n}{1-\gamma} - \min_{s' \in \mathcal{U}_x} \sum_{i \in N} d(s_i, s_i').$$

Because the term $n/(1-\gamma)$ is independent of the policy, this demonstrates that the distance of any alternative x from consensus (i.e., the distance to the closest profile in which x is the unanimous winner) is exactly the social cost incurred by a policy that tries to make x a unanimous winner. In fact, the social choice MDP need not *explicitly select* an alternative at each stage. A "one shot policy" can be used that selects an alternative just once; from that point, the MDP simply reflects the evolution of agent preferences over time while

the selected "incumbent" remains in place. On this view, the distance-minimizing alternative in the DR sense is the alternative selected by the optimal "one-shot" policy given our model of preference dynamics.

Note that our proof does not actually require d to be a distance—it is sufficient that d be a quasidistance (symmetry is not needed). More precisely, we require only the following properties: (i) the triangle inequality; (ii) for every $P \in \mathcal{L}$, $d(P, P) = 0$; and (iii) if $P[1] \neq P'[1]$ then $d(P, P') > 0$. Properties (ii) and (iii) are implied by Identity of Indiscernibles, and are implicitly used in the definition of $\hat{p}$ to ensure that these probabilities are well-defined.

Elkind, Faliszewski, and Slinko (2009) (see also (Elkind, Faliszewski, and Slinko 2010b, Theorem 4.9)) show that positional scoring rules are ℓ_1-votewise rationalizable with respect to unanimity via the following pseudodistance:

$$d_{\vec{\alpha}}(P, P') = \sum_{x \in A} |\alpha_{P[x]} - \alpha_{P'[x]}|.$$

For $P \neq P'$ it may be that $d_{\vec{\alpha}}(P, P') = 0$ (i.e., Identity of Indiscernibles is violated). However, if $\alpha_1 > \alpha_2$, d_α satisfies the weaker property (iii) above (this is true for, e.g., plurality and Borda count). Hence, we have the following corollary:

Corollary 2. *Let $\vec{\alpha}$ be a score vector with $\alpha_1 > \alpha_2$, let $d_{\vec{\alpha}}$ be defined as above, and let $D = \max_{P,P' \in \mathcal{L}} d_{\vec{\alpha}}(P, P')$ be an upper bound on $d_{\vec{\alpha}}$. Assume that the discount factor satisfies $\gamma > 1 - 1/D$. Then there exists a social choice MDP $\mathcal{M}(d_{\vec{\alpha}}) = (\mathcal{S}, A, R^*, T(d_{\vec{\alpha}}))$ such that π is a positional scoring rule induced by $\vec{\alpha}$ if and only if π is an optimal deterministic policy in $\mathcal{M}(d_{\vec{\alpha}})$.*

For instance, consider $\vec{\alpha} = (m-1, m-2, \ldots, 0)$, the score vector corresponding to the Borda count. For each $P \in \mathcal{L}$ and $x \in A$, if $P[1] \neq x$ we let $S(P, x) = P'$, where P' is identical to P except for switching the positions of x and $P[1]$. Then $\hat{d}(P, x) = 2(\alpha_1 - \alpha_{P[x]})$, and this is used directly to define $\hat{p}(P, x)$ using (3).

4 Discussion

What are the conceptual conclusions one can draw from Theorem 1? The theorem and its proof provide an explicit interpretation of distances in terms of rewards and expected societal costs, albeit one that is rather specific. In particular, the plurality reward function R^*—when viewed from the perspective of the expected cumulative discounted rewards accrued under the proposed model of preference dynamics—provides an operational measure of the quality of an alternative. The value of the optimal policy (or an optimal alternative in the one-shot case), directly measures societal utility—both the rate at which ultimate consensus is reached, and the degree of consensus attained along the way—in an arguably natural way. Interestingly, some real-life settings require specific forms of consensus to reach a decision; in particular, in criminal law jury trials in many jurisdictions need a unanimous verdict. In multiagent systems achieving unanimity may be especially desirable as it is extremely robust to failures.

A recent line of work deals with approximating (in the standard multiplicative sense) the score of alternatives according to prominent distance rationalizable voting rules (see, e.g., (Kenyon-Mathieu and Schudy 2007; Caragiannis et al. 2009)), including positional scoring rules (Procaccia 2010). Under these functions, the score of an alternative is exactly its distance from consensus, and therefore the distance rationalizability framework provides a way to quantify the trade-off between approximation quality and desirability of the outcome. For example, 2-approximating the score singles out an alternative that is at most twice as far from consensus. Our approach seems useful in the context of approximation; the specific interpretation suggested by our results is that if x is twice as close as y to being a unanimous winner then the societal cost of disagreement in the dynamic process when making x a unanimous winner is half as large as that of y.

Our approach hinges crucially on several details. First, we consider only one specific reward function R^*, though one we believe can be justified. Second, a possible criticism of our approach is that Theorem 1 holds only for the very specific social choice MDPs, and especially the specific transition functions that are constructed in its proof. Of course, in centrally-designed multiagent systems, agent preference revision processes can be constructed to support such transition models. Nevertheless, in characterizing the behavior of truly autonomous agents (human or artificial), a social choice MDP might reflect the preference dynamics exhibited in the domain of interest. The results in this paper should be seen as a proof of concept. An interesting direction for future research is to characterize the social choice MDPs where distance rationalizable voting rules and optimal policies coincide. But we need not be bound by current approaches in the DR framework. For example, if we allow that different agents have different preference revision mechanisms, votewise distances should be generalized to admit different distance functions over rankings for different agent types. In this way, the dynamic social choice perspective can drive the development of new distance models, and as a consequence, new voting rules.

From a more technical perspective, it may be possible to obtain similar results for votewise distance rationalizability under other prominent norms, such as ℓ_∞ (where the maximum distance between votes rather than the sum of distances is considered). The mathematical advantage of the ℓ_1 norm when used in conjunction with R^* is that agents are fully decoupled via the linearity of expectation, whereas the ℓ_∞ norm inevitably couples the agents. Computing the value of a state then involves estimating the maximum of geometric distributions, which may lead to difficulties in providing exact values as we do in the proof of Theorem 1.

Similarly, an extension of Theorem 1 to other consensus classes, such as *Condorcet* or *majority*, is highly nontrivial. One would have to redefine the reward function to give higher reward the closer one gets to the consensus notion under consideration. Unfortunately, when determining whether $x \in A$ is a Condorcet or majority winner in a given profile, one cannot consider each agent separately (whereas this is possible for unanimity, as one simply needs to verify that

each agent ranks x first). Therefore, once again, some of the independence assumptions that we rely on in our proof break down. This is a technically challenging direction for future research.

References

Caragiannis, I.; Covey, J. A.; Feldman, M.; Homan, C. M.; Kaklamanis, C.; Karanikolas, N.; Procaccia, A. D.; and Rosenschein, J. S. 2009. On the approximability of Dodgson and Young elections. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1058–1067.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2009. On distance rationalizability of some voting rules. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 108–117.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2010a. Good rationalizations of voting rules. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 774–779.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2010b. On the role of distances in defining voting rules. In *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 375–382.

Kenyon-Mathieu, C., and Schudy, W. 2007. How to rank with few errors. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, 95–103.

Meskanen, T., and Nurmi, H. 2008. Closeness counts in social choice. In Braham, M., and Steffen, F., eds., *Power, Freedom, and Voting*. Springer-Verlag.

Nurmi, H. 2004. A comparison of some distance-based choice rules in ranking environments. *Theory and Decision* 57(1):5–24.

Parkes, D. C., and Procaccia, A. D. 2010. Dynamic social choice: Foundations and algorithms. Manuscript. Available from: http://people.seas.harvard.edu/~arielpro/papers/dynamic.pdf.

Procaccia, A. D. 2010. Can approximation circumvent Gibbard-Satterthwaite? In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 836–841.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley.