Dynamic Social Choice with Evolving Preferences

David C. Parkes
School of Engineering and Applied Sciences
Harvard University
parkes@eecs.harvard.edu

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

Abstract

Social choice theory provides insights into a variety of collective decision making settings, but nowadays some of its tenets are challenged by internet environments, which call for dynamic decision making under constantly changing preferences. In this paper we model the problem via Markov decision processes (MDP), where the states of the MDP coincide with preference profiles and a (deterministic, stationary) policy corresponds to a social choice function. We can therefore employ the axioms studied in the social choice literature as guidelines in the design of socially desirable policies. We present tractable algorithms that compute optimal policies under different prominent social choice constraints. Our machinery relies on techniques for exploiting symmetries and isomorphisms between MDPs.

1 Introduction

Social choice theory has its roots in the writings of the marquis de Condorcet and the chevalier de Borda, and over the centuries has evolved so that nowadays we have a comprehensive mathematical understanding of social decision making processes. However, social choice theory falls short in the context of today's online communities. The internet and its myriad of applications has created a need for fast-paced, dynamic social decision making, which begs the question, is it possible to augment social choice theory to make it relevant for this new reality?

In our model of dynamic social decision making, a sequence of decisions must be made in the context of a population with constantly changing preferences, where the evolution of future preferences depends on past preferences and past decisions. As a running example, we consider online public policy advocacy groups, which are quickly gaining popularity and influence on the web and in social networks such as Facebook (via the application *Causes*); to be concrete we focus on MoveOn, which boasts more than five million members. Ideally the causes or issues that are advocated by MoveOn directly stem from the collective preferences of the members. A salient feature of MoveOn is that the time frame between deciding on a cause and acting on it is very short. Crucially, when a cause is chosen and advocated the preferences of the members will usually change, and this should have an impact on the next cause to be chosen. So, we are faced with a situation where both the current cause and the preferences of the members are constantly shifting. This calls for a consistent and socially desirable mechanism that sequentially selects the current cause given the current preferences of MoveOn members.

Our model. The common social choice setting concerns a set of agents (members, in the example) and a set of alternatives (causes or issues, in the example); the preferences of each agent are given by a ranking of the alternatives. A *preference profile* is a collection of the agents' preferences. The outcome is determined by a *social choice function*, which maps a given preference profile to the winning alternative.

We introduce dynamic preferences into this static setting by representing the preferences of agents and their stochastic transitions using a *social choice Markov decision process (MDP)*. One of the virtues of this model is that a state of the MDP corresponds to a preference profile. In other words, a static snapshot of the social choice MDP at any given time reduces, in a sense, to the traditional social choice setting. As in the traditional setting, the set of actions available in each state (which trigger transitions) coincides with the set of alternatives, and indeed in our example members' opinions about a cause that is currently advocated – and hence their preferences — are likely to change.

We say that agents that have identical transition models share the same *type*. In our running example, we imagine MoveOn staffers categorizing their membership according to carefully selected features (e.g., "moderate" or "progressive") and eliciting the vector of features from each member. Each vector of features can then be associated with a transition model in an approximate way. Constructing the transition models is a nontrivial problem that is beyond the scope of this paper (more on this in Section 6).

A deterministic (stationary) policy in an MDP maps each state to the action taken in this state. The crucial insight, which will enable us to relate the dynamic setting to traditional social choice theory, is that we interpret *a deterministic policy in a social choice MDP as a social choice function*. Once this connection has been made, then the usual axiomatic properties of social choice are imposed on policies and not just on decisions in specific states. Still, some axioms such as *Pareto optimality* (in social choice, whenever all the agents prefer x to y, y would not be elected) have a

local interpretation, in the sense that they can be interpreted state by state. In the case of Pareto optimality, if at any point the members all prefer one choice to another then the latter choice should not be made by the organization. But other axioms are *nonlocal*. For example, a policy is *onto* if every possible choice is selected by the policy in some state. Similarly, the requirement that a policy be *nondictatorial* rules out selecting a particular choice in a state only if it is most preferred of the same member who is also getting his most-preferred choice in every other state.

The final component of a social choice MDP model is a reward function, which associates a given action taken in a given state with a reward (for the designer); the goal is to optimize the infinite sum of discounted rewards. The existence of such a *customizable* objective is novel from a social choice perspective (despite much work on optimization in social choice, e.g., of latent utility functions (Boutilier et al. 2012)). Unfortunately, a policy that optimizes discounted rewards may not satisfy basic social choice properties, and hence may be undesirable as a social decision making mechanism. The key algorithmic challenge that we introduce is therefore:

Given a social choice MDP, tractably compute an optimal deterministic policy subject to given social choice constraints.

Our results. Observe that the state space of a social choice MDP is huge; if there are n agents and m alternatives then its cardinality is $(m!)^n$. To make things manageable we assume in our algorithmic results that there is only a constant number of alternatives, i.e., $m = \mathcal{O}(1)$, and moreover that the number of types is also bounded by a constant. We wish to design algorithms that are polynomial time in n. We argue that these assumptions are consistent with our motivation: while the number of MoveOn members is in the millions, the number of causes is usually rather small, and the number of types is limited by the number of features, which must be individually elicited from each member.

As mentioned above, local social choice axioms restrict individual states to certain actions in a way that is independent of the actions selected for other states. A local axiom is *anonymous* if it is indifferent to the identities of the agents. Some of the most prominent social choice axioms are local and anonymous. Our main result is an algorithm that, given a social choice MDP, an anonymous reward function, and an anonymous local axiom, computes an optimal policy that satisfies the axiom in polynomial time in n. Turning to nonlocal axioms, we prove that, given a social choice MDP and an anonymous reward function, we can find an optimal policy that is onto or nondictatorial in polynomial time in n.

Related work. Our work is conceptually related to the literature on dynamic incentive mechanisms (see, e.g., Parkes et al. 2010 for an overview). Specifically, our model is inspired by models where the preferences of agents also evolve via an MDP (Cavallo, Parkes, and Singh 2006; Bergemann and Välimäki 2010). However, the focus of the work on dynamic incentive mechanisms is the design of monetary transfers to the agents in a way that incentivizes truthful reporting of preferences.

Dynamic preferences can also be found in papers on iterative voting (see, e.g., Meir et al. 2010), but this work is fundamentally different in that it fixes a social choice function and studies a strategic best response process; it is not the true preferences that are changing, but rather the reported preferences.

The problem of constructing optimal policies for *constrained MDPs* has received some attention (Altman 1999; Dolgov and Durfee 2005; 2006). Unfortunately the constraints in question usually have a specific structure; to the best of our knowledge the constraints considered in previous work are not sufficiently general to capture our social choice constraints, and moreover we insist on deterministic policies.

A recent paper by Boutilier and Procaccia (2012) builds on our model of dynamic social choice (which was publicly available earlier) to relate a notion of distances in social choice (Elkind, Faliszewski, and Slinko 2009) to an operational measure of social desirability. Very informally, they show that a social choice function selects alternatives that are closest to being consensus winners if and only if that social choice function is an optimal policy of a specific, arguably natural, social choice MDP. The conceptual implication is that alternatives that are close to consensus can become winners faster in a dynamic process.

2 Preliminaries

In this section we formally introduce basic notions from the MDP literature and social choice theory.

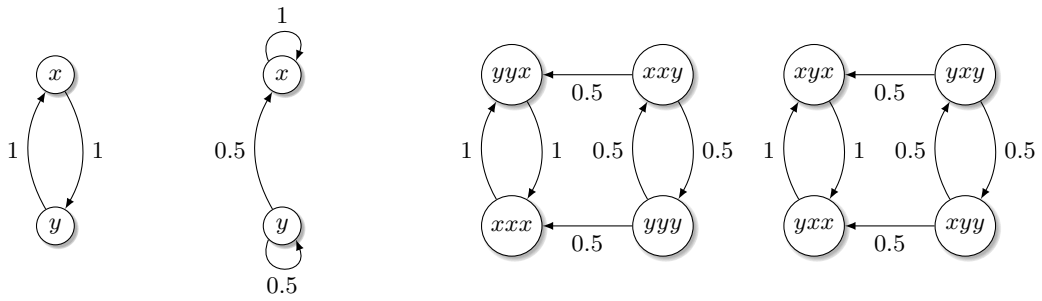
2.1 Markov Decision Processes

Below we briefly review the basics of Markov decision processes; the reader is referred to the book by Puterman (1994) for more details. A *Markov decision process (MDP)* is a 4-tuple $\mathcal{M} = (\mathcal{S}, A, R, P)$ where $\mathcal{S}$ is a finite set of states; A is a finite set of actions; $R : \mathcal{S} \times A \to \mathbb{R}$ is a reward function, where for $s \in \mathcal{S}$ and $a \in A$, $R(s, a)$ is the reward obtained when taking action a in state s; and P is the transition function, where $P(s'|s, a)$ is the probability of moving to state s' when action a is taken in state s. Note that the transition function is Markovian in that it depends only on the current state and not on the history.

A *deterministic policy* is a function $\pi : \mathcal{S} \to A$, which prescribes which action $\pi(s)$ is taken in state $s \in \mathcal{S}$. This definition implicitly assumes that the policy is also stationary, that is, it does not depend on the history.

There are several variations to how the cumulative reward is calculated. We consider the most common approach where there is an infinite horizon and a discount factor $\gamma \in [0, 1)$. Given an MDP $\mathcal{M}$, a policy π is associated with a value function $V_\pi : \mathcal{S} \to \mathbb{R}$, where $V_\pi(s)$ is the cumulative discounted reward that is obtained if the initial state is $s \in \mathcal{S}$ and the action prescribed by π is taken at each step.

It is known that for any (unconstrained) MDP there is an optimal policy π^* that is deterministic. Such an optimal policy can be found in polynomial time, e.g., by computing the optimal values via linear programming and then greedily assigning actions that achieve the maximum at every state.



(a) The two type transition models: type 1 (left) type 2 (right).

(b) The transition model of the MDP.

Figure 1: In this example there are three agents and two alternatives, $A = \{x, y\}$. There are two types and $\widehat{\Theta} = \{\{1, 2\}, \{3\}\}$, that is, agents 1 and 2 are of type 1 and agent 3 is of type 2. The transition models of the two types are illustrated in (a). We only show the transitions that are associated with action x. A node is labeled with the top preference of an agent, e.g., a node labeled by x corresponds to the preference $x \succ y$. The transition model of the MDP is shown in (b), where again only the transitions that are associated with the action x are illustrated. A node is labeled by the top preference of the three agents, e.g., xyx corresponds to the preference profile where $x \succ_1 y, y \succ_2 x, x \succ_3 y$.

2.2 Social choice

Let $N = \{1, \ldots, n\}$ be a set of *agents*, and let A be a set of alternatives where $|A| = m$; we overload the notation A, as below the actions in our MDP coincide with the set of alternatives. Each agent is associated with strict linear preferences $\succ_i$ over the alternatives, that is, a strict ranking of the alternatives; $x \succ_i y$ means that agent i prefers x to y. Let $\mathcal{L} = \mathcal{L}(A)$ denote the set of strict linear preferences over A. A collection $\vec{\succ} = (\succ_1, \ldots, \succ_n) \in \mathcal{L}^n$ of the agents' preferences is called a *preference profile*. A *social choice function* is a function $f : \mathcal{L}^n \to A$ that designates a winning alternative given a preference profile. Given $\succ \in \mathcal{L}$ we denote by top($\succ$) the alternative that is most preferred in $\succ$.

A prominent approach in social choice theory compares social choice functions based on their axiomatic properties. Two properties are considered absolutely essential and are satisfied by all commonly studied social choice functions. A social choice function f is *onto* if for every $a \in A$ there exists $\vec{\succ} \in \mathcal{L}^n$ such that $f(\vec{\succ}) = a$, that is, every alternative can be elected. A social choice function f is *dictatorial* if there exists an agent $i \in N$ such that for every $\vec{\succ} \in \mathcal{L}^n$, $f(\vec{\succ}) = \text{top}(\succ_i)$; f is *nondictatorial* if there is no such agent.

Below we define some other prominent axioms; the first two axioms provide a notion of consensus, and require that a social choice function elect an alternative when this notion is satisfied. We say that $a^* \in A$ is a *Condorcet winner* in $\vec{\succ}$ if for all $a \in A \setminus \{a^*\}$, $|\{i \in N : a^* \succ_i a\}| > n/2$, that is, a majority of agents prefer a^* to any other alternative. A social choice function f is *Condorcet-consistent* if $f(\vec{\succ}) = a^*$ whenever a^* is a Condorcet winner in $\vec{\succ}$; it is *unanimous* if for every $\vec{\succ} \in \mathcal{L}^n$ such that top($\succ_i$) $= a^*$ for every $i \in N$, $f(\vec{\succ}) = a^*$, i.e., it always elects an alternative that is ranked first by all the agents. A related axiom is *Pareto optimality* that requires that for every $\vec{\succ} \in \mathcal{L}$ where $x \succ_i y$ for all $i \in N$, $f(\vec{\succ}) \neq y$.

3 The Model

Let $N = \{1, \ldots, n\}$ be the set of agents and A be the set of alternatives; denote $|A| = m$. We presently describe the MDP $\mathcal{M} = (\mathcal{S}, A, R, P)$ that we deal with. Specifically, we will describe $\mathcal{S}$, A, and P, and leave the restrictions on R for later. The state transitions in our MDP are factored across agents and thus the MDP is a very special case of a *factored MDP* (Boutilier, Dearden, and Goldszmidt 1995).

The set of actions A of our MDP coincides with the set of alternatives (which is also denoted by A). For each agent i we have a random variable X_i, which takes values in $\mathcal{L}$. The current value of X_i indicates the current preferences of agent i. A state $s \in \mathcal{S}$ defines a value $\succ_i \in \mathcal{L}$ for every variable X_i, $i \in N$. Therefore, each state is a preference profile, and the size of the state space is huge: $(m!)^n$. Given a state $s \in \mathcal{S}$, we denote by $s(i)$ the value of X_i in this state, that is, the preferences of agent i. Furthermore, given a permutation $\mu : N \to N$, we also denote by $\mu(s)$ the state such that $s(i) = \mu(s)(\mu(i))$ for all $i \in N$, that is, μ can also be seen as a permutation over states.

For each $i \in N$ we have a transition model P_i, where $P_i(\succ'_i \mid \succ_i, a)$ is the probability of X_i taking the value $\succ'_i$ when the current value is $\succ_i \in \mathcal{L}$ and the action $a \in A$ is taken. Below we assume that there are only t possible transition models, where $t \leq n$. We say that agents with the same transition model have the same *type*. We let $\widehat{\Theta}$ be the partition of agents into types, where each $\theta \in \widehat{\Theta}$ is a set of agents with the same type. In the following we will find it useful to construct partitions that are more refined, and sometimes coarser, than this basic type-based partition.

We define the transition model of the MDP $\mathcal{M}$ by letting

$$P(s'|s, a) = \prod_{i \in N} P_i(s'(i)|s(i), a). \tag{1}$$

See Figure 1 for an illustration. Intuitively, at every step we elect an alternative, and this choice affects the preferences of

all the agents. Note that the preference ranking of each agent transitions independently given the selection of a particular alternative.

Definition 3.1. An MDP $\mathcal{M} = (\mathcal{S}, A, R, P)$ where $\mathcal{S}$, A, and P are as above, is called a *social choice MDP*.

4 Symmetries and Local Axioms

Having defined the social choice MDP, and in particular with states corresponding to preference profiles, we interpret a deterministic policy π as a social choice function. A policy defines a mapping from preference profiles to alternatives. We can therefore seek policies that satisfy the traditional social choice axioms; this is indeed the focus of our technical results.

We ask whether it is possible to efficiently compute an optimal policy despite the large number of states. We will show that we can provide positive answers by exploiting symmetries between states. In particular, we show that the concepts of symmetry are sufficiently general that they facilitate the efficient computation of optimal policies that satisfy what we call *anonymous local axioms*.

4.1 Exploiting symmetries

The symmetries in a social choice MDP stem from the identical transition models associated with agents of the same type. Intuitively, rather than concerning ourselves with which ranking is currently held by each agent, it should be enough to keep track of *how many agents of each type possess each ranking*.

To make this precise we use a formalism introduced by Zinkevich and Balch (2001). Given an *equivalence relation* E over a set B, we denote by $E(x) = \{x' \in B : (x, x') \in E\}$ the *equivalence class* of $x \in B$, and by $\mathcal{E}$ the set of equivalence classes of E; in particular $E(x) \in \mathcal{E}$. $\mathcal{E}$ is a partition of set B.

Definition 4.1. Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be an MDP and let E be an equivalence relation over $\mathcal{S}$.

1. R is *symmetric with respect to* E if for all $(s, s') \in E$ and all $a \in A$, $R(s, a) = R(s', a)$.
2. P is *symmetric with respect to* E if for all $(s, s') \in E$, every $a \in A$, and every equivalence class $S \in \mathcal{E}$,
$$\sum_{s'' \in S} P(s''|s, a) = \sum_{s'' \in S} P(s''|s', a).$$
3. $\mathcal{M}$ is *symmetric with respect to* E if R and P are symmetric with respect to E.
4. A policy π is symmetric with respect to E if for all $(s, s') \in E$, $\pi(s) = \pi(s')$.

Intuitively, symmetry of an MDP with respect to an equivalence relation requires, for every action and every equivalence class on states, that both the reward and the probability of transitioning to any particular equivalence class, is independent of the exact state in the current equivalence class.

Zinkevich and Balch (2001) show that if $\mathcal{M}$ is symmetric with respect to E then there is an optimal deterministic policy that is identical on the equivalence classes of $\mathcal{E}$, and thus symmetric with respect to E. This is useful because an optimal policy can be computed by contracting the state space of $\mathcal{M}$, replacing each equivalence class in $\mathcal{E}$ with one state. More formally, the following lemma is an adaptation of (Zinkevich and Balch 2001, Theorem 2).

Lemma 4.2 (Zinkevich and Balch 2001)**.** *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be an MDP and let E be an equivalence relation over $\mathcal{S}$. Assume that $\mathcal{M}$ is symmetric with respect to E. Then an optimal deterministic policy for $\mathcal{M}$ that is symmetric with respect to E can be computed in time that is polynomial in $|\mathcal{E}|$ and $|A|$.*

In order to employ Lemma 4.2 we must construct an equivalence relation for which the social choice MDP is symmetric. For this, we define a partition Θ on agents that induces an equivalence relation E_Θ on states. As a special case, we will be interested in the partition $\widehat{\Theta}$ of agents into types but a formulation in terms of arbitrary partitions over the agents is useful for the result of Section 5.

Given a partition Θ of agents, we define E_Θ as follows. For all $s, s' \in \mathcal{S}$, $(s, s') \in E_\Theta$ if and only if for all $\theta \in \Theta$ and $\succ \in \mathcal{L}$,
$$|\{i \in \theta : s(i) = \succ\}| = |\{i \in \theta : s'(i) = \succ\}|.$$
Informally, two states s and s' are equivalent given partition Θ on agents if one state can be obtained from the other by a permutation of the preferences of agents in the same subset $\theta \in \Theta$. For example, if $\Theta = \{\{1\}, \{2\}, \{3\}\}$ then all states are distinct. If $\Theta = \{\{1, 2\}, \{3\}\}$, then any pair of states where agents 1 and 2 swap rankings are equivalent.

Note that for each $\theta \in \Theta$, each $\succ \in \mathcal{L}$, and any $s \in \mathcal{S}$, $|\{i \in \theta : s(i) = \succ\}| \in \{0, \ldots, n\}$. This immediately implies the following lemma.[1]

Lemma 4.3. *Let Θ be a partition of the agents such that $|\Theta| = k$. Then $|\mathcal{E}_\Theta| \leq (n + 1)^{k(m!)}$.*

Thus, we see that when m is constant, the number of equivalence classes on states induced by a partition Θ of constant size k is polynomial in n.

Definition 4.4. Given a partition Θ of some (arbitrary) set, a partition Θ' is called a *refinement of* Θ if for every $\theta' \in \Theta'$ there exists $\theta \in \Theta$ such that $\theta' \subseteq \theta$.

We observe that refining the partition over the set of agents also refines the associated partition of states into equivalence classes. Formally:

Observation 4.5. *Let Θ be a partition of agents, and let Θ' be a refinement of Θ. Then $\mathcal{E}_{\Theta'}$ is a refinement of $\mathcal{E}_\Theta$, where these are the associated partitions on states.*

We wish to prove that a social choice MDP is symmetric with respect to equivalence classes induced by refinements of the partition of agents into types. The next lemma establishes this symmetry for the transition model.

Lemma 4.6. *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP, and let $\widehat{\Theta}$ be the partition of the agents into types. Let Θ be a refinement of $\widehat{\Theta}$. Then P is symmetric with respect to E_Θ.*

[1] The bound given in Lemma 4.3 is far from being tight, but it is sufficient for our purposes.

Proof. Let $s, s' \in \mathcal{S}$ such that $(s, s') \in E_\Theta$. Consider a permutation $\mu : N \to N$ such that for every $\theta \in \Theta$ and every $i \in \theta$, $\mu(i) \in \theta$, and $s' = \mu(s)$; μ is guaranteed to exist by the definition of E_Θ.

Since Θ is a refinement of $\widehat{\Theta}$, for every $\theta \in \Theta$ and every $i, j \in \theta$, $P_i \equiv P_j$, and in particular for every $i \in N$, $P_i \equiv P_{\mu(i)}$. It follows from (1) that

$$P(s''|s, a) = P(\mu(s'')|\mu(s), a) = P(\mu(s'')|s', a)$$

for all $s'' \in \mathcal{S}$. Therefore, for every $a \in A$ and every equivalence class $S \in \mathcal{E}_\Theta$,

$$\sum_{s'' \in S} P(s''|s, a) = \sum_{s'' \in S} P(\mu(s'')|s', a). \tag{2}$$

Next, recall that $\mu(i) \in \theta$ for each $\theta \in \Theta$ and $i \in \theta$, and hence for every $S \in \mathcal{E}_\Theta$ and $s'' \in S$ it holds that $(s'', \mu(s'')) \in E_\Theta$. We wish to claim that μ restricted to S is a permutation on S. It is sufficient to prove that μ is one-to-one; this is true since if $s_1'' \neq s_2''$ then there exists $i \in N$ such that $s_1''(i) \neq s_2''(i)$, and therefore $\mu(s_1'')(\mu(i)) \neq \mu(s_2'')(\mu(i))$. We conclude that

$$\sum_{s'' \in S} P(\mu(s'')|s', a) = \sum_{s'' \in S} P(s''|s', a). \tag{3}$$

The combination of (2) and (3) yields

$$\sum_{s'' \in S} P(s''|s, a) = \sum_{s'' \in S} P(s''|s', a),$$

as desired. □

So far we have imposed no restrictions on the reward function R. Below we will require it to be anonymous, that is, it should be symmetric with respect to the equivalence relation $E_{\{N\}}$ induced by the coarsest partition of the agents, $\{N\}$. Equivalently, a reward function R is anonymous if for every permutation $\mu : N \to N$ on the agents, all $s \in \mathcal{S}$ and all $a \in A$, $R(s, a) = R(\mu(s), a)$. This is a stronger restriction than what we need technically, but it seems quite natural.

Definition 4.7. Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP. A reward function $R : \mathcal{S} \times A \to \mathbb{R}$ is *anonymous* if it is symmetric with respect to $E_{\{N\}}$.

As an example of an anonymous reward function, consider a designer who wishes to reach a social consensus. As described in Section 2.2, three common notions of consensus are a *Condorcet winner*, a *majority winner* which is ranked first by a majority of agents, and (the stronger notion of) a *unanimous winner*, that is, an alternative ranked first by all agents. The reward function can then be, e.g., of the form

$$R(s, a) = \begin{cases} 1 & \text{if } a \text{ is a Condorcet winner in } s \\ 0 & \text{otherwise} \end{cases}$$

The abovementioned notions of social consensus are indifferent to permuting the agents and hence lead to anonymous reward functions.

The following observation is a direct corollary of Observation 4.5 and the fact that any partition on agents is a refinement of $\{N\}$.

Observation 4.8. *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP. Let Θ be any agent partition, and let $R : \mathcal{S} \times A \to \mathbb{R}$ be an anonymous reward function. Then R is symmetric with respect to E_Θ.*

To summarize, a social choice MDP with an anonymous reward function is symmetric with respect to refinements of $E_{\widehat{\Theta}}$ (by Lemma 4.6 and Observation 4.8), and this suggests a method for computing an optimal deterministic policy that is polynomial in n by Lemma 4.2.

We want to prove a more general result though, which will also enable the restriction of the actions available in some states. The purpose of handling restrictions is twofold. First, it will directly facilitate the computation of optimal policies that are consistent with certain local axioms. Second, it is crucial for the result of Section 5 that addresses some non-local axioms.

Definition 4.9. Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP.

1. A *restriction* is a subset $\Psi \subseteq \mathcal{S} \times A$. A deterministic policy π is Ψ-consistent if for every $s \in \mathcal{S}$, $\pi(s) \in \{a \in A : (s, a) \in \Psi\}$.
2. Let Θ be a partition of the agents. A restriction Ψ is symmetric with respect to induced equivalence relation E_Θ on states if for all $(s, s') \in E_\Theta$, and all $a \in A$, $(s, a) \in \Psi \Leftrightarrow (s', a) \in \Psi$.

That is, a restriction to particular actions in particular states is symmetric with respect to an equivalence relation if the same restriction holds across all equivalent states. In some of the MDP literature a restriction is considered to be a component of the MDP, but we consider the MDPs and restrictions separately, as we see the restriction as an external constraint that is imposed on the MDP.

We are finally ready to present our main result. A short discussion is in order, though, regarding the parameters of the problem. As mentioned above, the size of the state space of a social choice MDP is $(m!)^n$. Even if m is constant, this is still exponential in n. In order to obtain running time that is tractable with respect to the number of agents n we also assume that the number of types is constant.

Theorem 4.10. *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP, let R be an anonymous reward function, let $\widehat{\Theta}$ be the partition of the agents into types, let Θ be a refinement of $\widehat{\Theta}$, and let $\Psi \subseteq \mathcal{S} \times A$ be symmetric with respect to E_Θ. Furthermore, assume that $|A| = \mathcal{O}(1)$ and $|\Theta| = \mathcal{O}(1)$. Then an optimal deterministic Ψ-consistent policy for $\mathcal{M}$ (which will be symmetric with respect to E_Θ) can be computed in polynomial time in the number of agents n.*

Proof. We define $R' : \mathcal{S} \times A \to \mathbb{R}$ such that[2]

$$R'(s, a) = \begin{cases} R(s, a) & \text{if } (s, a) \in \Psi \\ -\infty & \text{otherwise} \end{cases}$$

From the facts that R is symmetric with respect to E_Θ (by Observation 4.8) and Ψ is symmetric with respect to E_Θ (by

[2]It is possible to replace $-\infty$ with a sufficiently negative constant, e.g., $-(1/(1-\gamma)) \max_{s,a} R(s, a)$.

assumption), it follows that R' is also symmetric with respect to E_Θ. By Lemma 4.6, P is symmetric with respect to E_Θ. Let $\mathcal{M}' = (\mathcal{S}, A, R', P)$; we have that $\mathcal{M}'$ is symmetric with respect to E_Θ. By Lemma 4.2 we can find an optimal policy for $\mathcal{M}'$ in time polynomial in $|A| = \mathcal{O}(1)$ and $|\mathcal{E}_\Theta|$. By Lemma 4.3, using $|A| = \mathcal{O}(1)$ and $|\Theta| = \mathcal{O}(1)$, we have that $|\mathcal{E}_\Theta| \leq (n+1)^{\mathcal{O}(1)}$, therefore an optimal deterministic policy for $\mathcal{M}'$ can be computed in polynomial time. The definition of R' simply rules out the use of state-action pairs that are not in Ψ, hence this policy is an optimal deterministic Ψ-consistent policy for $\mathcal{M}$. □

4.2 Anonymous local axioms

Theorem 4.10 provides us with the means to compute optimal policies that are consistent with restrictions that respect symmetries. Fortunately, many prominent social choice axioms can be expressed with just these kinds of restrictions, hence we can settle the question of computing optimal deterministic policies under such constraints.

Definition 4.11. An axiom is *local*[3] if it can be represented as a restriction $\Psi \subseteq \mathcal{S} \times A$. A local axiom is *anonymous* if it is symmetric with respect to $E_{\{N\}}$.

Informally, a local axiom prescribes which actions are allowed in each state; it is local in the sense that this does not depend on the policy's choices in other states. Anonymity requires that the acceptability, or not, of an action does not depend on the names of agents. For example, Condorcet-consistency is a local axiom since it simply restricts the set of actions (to a singleton) in every state where a Condorcet winner exists. Condorcet-consistency is also anonymous: if an alternative is a Condorcet winner, it would remain one for any permutation of preferences amongst agents. More generally, we have the following observation.

Observation 4.12. *The following axioms are local and anonymous: Condorcet consistency, Pareto-optimality, and unanimity.*[4]

Now the following result is an immediate corollary of Theorem 4.10.

Corollary 4.13. *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP, let R be an anonymous reward function, let $\widehat{\Theta}$ be the partition of the agents into types, and let Ψ be the restriction that represents an anonymous local axiom. Furthermore, assume that $|A| = \mathcal{O}(1)$ and $|\widehat{\Theta}| = t = \mathcal{O}(1)$. Then an optimal deterministic Ψ-consistent policy for $\mathcal{M}$ can be computed in polynomial time in the number of agents n.*

5 Nonlocal Axioms

The axioms that we were able to handle in Section 4 are local, in the sense that the restrictions on the available actions in a state are not conditional on the actions taken in other states. However, many important axioms do not possess this property, e.g., neither Ontoness nor nondictatorship are local. Analogously to Corollary 4.13, we can show:

Theorem 5.1. *Let $\mathcal{M} = (\mathcal{S}, A, R, P)$ be a social choice MDP, let R be an anonymous reward function, let $\widehat{\Theta}$ be the partition of the agents into types, and let $\hat{s} \in \mathcal{S}$. Furthermore, assume that $m = |A| = \mathcal{O}(1)$ and $|\widehat{\Theta}| = \mathcal{O}(1)$. Then an optimal deterministic policy for $\mathcal{M}$ that is either onto or nondictatorial can be computed in polynomial time in the number of agents n.*

The theorem's intricate proof is relegated to the full version of the paper.[5] On a high level, ontoness and nondictatorship stand out in that they admit a tractable guided search approach. For example, there are n dictatorial polices that must be ruled out in obtaining a nondictatorial policy. The technical challenge is to exclude policies through search without breaking the symmetries. To do this we must carefully refine our equivalence classes, without causing an exponential growth in their number. Our algorithms therefore amount to a balancing act, where we refine the equivalence relations on states while avoiding very detailed symmetries that lead to a bad running time.

Conceptually though, we view Corollary 4.13 as a far more important result because it captures many of the most prominent social choice axioms. Theorem 5.1 should be seen as a proof of concept: computing optimal policies subject to some natural nonlocal constraints is tractable (albeit quite difficult). It remains open whether a similar result can be obtained for other prominent nonlocal axioms, e.g., monotonicity (pushing an alternative upwards can only help it), anonymity (symmetry with respect to agents), and neutrality (symmetry with respect to alternatives).

6 Discussion

We have not addressed the challenge of constructing appropriate transition models for the agents; rather these transition models are assumed as input to our algorithms. There is a variety of techniques that may be suitable, ranging from automated methods such as machine learning (see, e.g., Crawford and Veloso 2005) and hidden Markov models, to marketing and psychological approaches. We aim to provide the first principled approach to decision making in environments where currently one is not available. Even a coarse partition of the agents into a few types, and subsequent application of our algorithmic results, would be an improvement over the status quo in organizations like MoveOn. Over time this initial partition can be gradually refined to yield better and better approximations of reality.

Many other challenges remain open, e.g., dealing with dynamically arriving and departing agents, dropping the assumptions on the number of alternatives and types, and tackling prominent nonlocal axioms. In addition to these technical challenges, we hope that the conceptual connection that we have made between social choice and MDPs — and between seemingly unrelated strands of AI research — will spark a vigorous discussion around these topics.

[3] Local axioms are also known as *intraprofile* conditions; see, e.g., (Roberts 1980).

[4] There are additional axioms that are local and anonymous, e.g., Smith consistency, mutual majority consistency, and invariant loss consistency; see (Tideman 2006) for a description of these axioms.

[5] Available from: http://www.cs.cmu.edu/~arielpro/papers.html.

References

Altman, E. 1999. *Constrained Markov Decision Processes*. Chapman and Hall.

Bergemann, D., and Välimäki, J. 2010. The dynamic pivot mechanism. *Econometrica* 78:771–789.

Boutilier, C., and Procaccia, A. D. 2012. A dynamic rationalization of distance rationalizability. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, 1278–1284.

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2012. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, 197–214.

Boutilier, C.; Dearden, R.; and Goldszmidt, M. 1995. Exploiting structure in policy construction. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1104–1111.

Cavallo, R.; Parkes, D. C.; and Singh, S. 2006. Optimal coordinated planning amongst self-interested agents with private state. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 55–62.

Crawford, E., and Veloso, M. 2005. Learning dynamic preferences in multi-agent meeting scheduling. In *Proceedings of the 5th IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT)*, 487–490.

Dolgov, D., and Durfee, E. 2005. Stationary deterministic policies for constrained MDPs with multiple rewards, costs, and discount factors. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, 1326–1331.

Dolgov, D., and Durfee, E. 2006. Resource allocation among agents with MDP-induced preferences. *Journal of Artificial Intelligence Research* 27:505–549.

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.

Meir, R.; Polukarov, M.; Rosenschein, J. S.; and Jennings, N. R. 2010. Convergence to equilibria in plurality voting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 823–828.

Parkes, D. C.; Cavallo, R.; Constantin, F.; and Singh, S. 2010. Dynamic incentive mechanisms. *AI Magazine* 31(4):79–94.

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

Roberts, K. W. S. 1980. Social choice theory: The single-profile and multi-profile approaches. *Review of Economic Studies* 47(2):441–450.

Tideman, N. 2006. *Collective Decisions and Voting*. Ashgate.

Zinkevich, M., and Balch, T. 2001. Symmetry in Markov decision processes and its implications for single agent and multiagent learning. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, 632–640.