Knowledge-theoretic properties of strategic voting

Samir Chopra[1], Eric Pacuit[2], and Rohit Parikh[3]

[1] Department of Computer Science
Brooklyn College of CUNY
Brooklyn, New York 11210
schopra@sci.brooklyn.cuny.edu
[2] Department of Computer Science
CUNY Graduate Center
New York, NY 10016
epacuit@cs.gc.cuny.edu
[3] Departments of Computer Science, Mathematics and Philosophy
Brooklyn College and CUNY Graduate Center
New York, NY 10016
ripbc@cunyvm.cuny.edu

Abstract. Results in social choice theory such as the Arrow and Gibbard-Satterthwaite theorems constrain the existence of rational collective decision making procedures in groups of agents. The Gibbard-Satterthwaite theorem says that no voting procedure is *strategy-proof*. That is, there will always be situations in which it is in a voter's interest to misrepresent its true preferences i.e., vote strategically. We present some properties of strategic voting and then examine – via a bimodal logic utilizing epistemic and strategizing modalities – the knowledge-theoretic properties of voting situations and note that unless the voter *knows* that it should vote strategically, and how, i.e., knows what the *other* voters' preferences are and *which* alternate preference P' it should use, the voter will not strategize. Our results suggest that opinion polls in election situations effectively serve as the first $n-1$ stages in an n stage election.

1 Introduction

A comprehensive theory of multi-agent interactions must pay attention to results in social choice theory such as the Arrow and Gibbard-Satterthwaite theorems [1, 7, 17]. These impossibility results constrain the existence of rational collective decision making procedures. Work on formalisms for belief merging already reflects the attention paid to social choice theory [9, 6, 12, 11, 13]. In this study we turn our attention to another aspect of social aggregation scenarios: the role played by the states of knowledge of the agents. The study of strategic interactions in game theory reflects the importance of states of knowledge of the players. In this paper, we bring these three issues—states of knowledge, strategic interaction and social aggregation operations—together.

The Gibbard-Satterthwaite theorem is best explained as follows[4]. Let S be a social choice function whose domain is an n-tuple of preferences $P_1 \ldots P_n$, where $\{1, ..., n\}$ are the voters, M is the set of choices or candidates and each P_i is a linear order over M. S takes $P_1 \ldots P_n$ as input and produces some element of M - the winner. Then the theorem says that there must be situations where it 'profits' a voter to vote *strategically*. Specifically, if P denotes the actual preference ordering of voter i, Y denotes the profile consisting of the preference orderings of all the other voters then the theorem says that there must exist P, Y, P' such that $S(P', Y) >_P S(P, Y)$. Here $>_P$ indicates: better according to P. Thus in the situation where the voter's actual ordering is P and all the orderings of the other voters (together) are Y then voter i is better off saying its ordering is P' rather than what it actually is, namely P. In particular, if the vote consists of voting for the highest element of the preference ordering, it should vote for the highest element of P' rather than of P.

Of course, the agent might be *forced* to express a different preference. For example, if an agent, whose preferences are $B > C > A$, is only presented C, A as choices, then the agent will pick C. This 'vote' differs from the agent's true preference, but should not be understood as 'strategizing' in the true sense.

A real-life example of strategizing was noticed in the 2000 US elections when some supporters of Ralph Nader voted for their second preference, Gore,[5] in a vain attempt to prevent the election of George W. Bush. In that case, Nader voters decided that (voting for the maximal element of) a Gore-Nader-Bush expression of their preferences would be closer to their desired ordering of Nader-Gore-Bush than the Bush-Gore-Nader ordering that would result if they voted for their actual top choice. Similar examples of strategizing have occurred in other electoral systems over the years ([4] may be consulted for further details on the application of game-theoretic concepts to voting scenarios). The Gibbard-Satterthwaite theorem points out that situations like the one pointed out above *must* arise.

What interests us in this paper are the *knowledge-theoretic properties* of the situation described above. We note that unless the voter with preference P *knows* that it should vote strategically, and how, i.e., knows that the other voters' preference is Y and that it should vote according to $P' \neq P$, the theorem is not 'effective'. That is, the theorem only applies in those situations where a certain level of knowledge exists amongst voters. Voters completely or partially ignorant about other voters' preferences would have little incentive to change their actual preference at election time. In the 2000 US elections, many Nader voters changed their votes *because* opinion polls had made it clear that Nader stood no chance of winning, and that Gore would lose as a result of their votes going to Nader.

4 Later we use a different formal framework; we have chosen to use this more transparent formalism during the introduction for ease of exposition.

5 Surveys show that had Nader not run, 46% of those who voted for him would have voted for Gore, 23% for Bush and 31% would have abstained. Hereafter, when we refer to Nader voters we shall mean those Nader voters who did or would have voted for Gore.

We develop a logic for reasoning about the knowledge that agents have of their own preferences and other agents' preferences, in a setting where a social aggregation function is defined and kept fixed throughout. We attempt to formalize the intuition that agents, knowing an aggregation function, and hence its outputs for input preferences, will strategize if they know a) enough about other agents' preferences and b) that the output of the aggregation function of a changed preference will provide them with a more favorable result, one that is closer to their true preference. We will augment the standard epistemic modality with a modality for strategizing. This choice of a bimodal logic brings with it a greater transparency in understanding the states that a voter will find itself in when there are two possible variances in an election: the preferences of the voters and the states of knowledge that describe these changing preferences.

Our results will suggest that election-year opinion polls are a way to effectively turn a one-shot game, i.e., an election, into a many-round game that may induce agents to strategize. Opinion polls make voters' preferences public in an election year and help voters decide on their strategies on the day of the election. For the rest of the paper, we will refer to opinion polls also as elections.

The outline of the paper is as follows. In Section 2 we define a formal voting system and prove some preliminary results about strategic voting. In Section 3 we demonstrate the dependency of strategizing on the voters' states of knowledge. In Section 4 we develop a bimodal logic for reasoning about strategizing in voting scenarios.

2 A Formal Voting Model

There is a wealth of literature on formal voting theory. This section draws upon discussions in [4, 5]. The reader is urged to consult these for further details.

Let $\mathcal{O} = \{o_1, \dots, o_m\}$ be a set of candidates, $\mathcal{A} = \{1, \dots, n\}$ be a set of agents or voters. We assume that each voter has a preference over the elements of $\mathcal{O}$, i.e., a reflexive, transitive and connected relation on $\mathcal{O}$. For simplicity we assume that each voter's preference is strict. A voter i's *strict preference relation* on $\mathcal{O}$ will be denoted by P_i. We represent each P_i by a function $P_i : \mathcal{O} \to \{1, \dots, m\}$, where we say that a voter *strictly prefers* o_j to o_k iff $P_i(o_j) > P_i(o_k)$. We will write $P_i = (o_1, \dots, o_n)$ iff $P_i(o_1) > P_i(o_2) > \cdots > P_i(o_n)$. Henceforth, for ease of readability we will use **Pref** to denote preferences over $\mathcal{O}$. A *preference profile* is an element of $(\mathbf{Pref})^n$. Given each agent's preference an *aggregation function* returns the social preference ordering over $\mathcal{O}$.

Definition 1 (Aggregation Function). *An* **aggregation function** *is a function from preference profiles to preferences:*

$$\mathrm{Ag} : \mathbf{Pref}^n \to \mathbf{Pref}$$

In voting scenarios such as elections, agents are not expected to announce their actual preference relation, but rather to select a vote that 'represents' their preference. Each voter chooses a vote v, the aggregation function tallies the votes of each candidate and selects a winner (or winners if electing more than one candidate). There are two components to any voting procedure. First, the type of votes that voters can cast. For example, in *plurality voting* voters can only vote for a single candidate so votes v are simply singleton subsets of $\mathcal{O}$, whereas in *approval voting* voters select a set of candidates so votes v are any subset of $\mathcal{O}$. Following [5], given a set of $\mathcal{O}$ of candidates, let $\mathcal{B}(\mathcal{O})$ be the set of feasible votes, or *ballots*. The second component of any voting procedure is the way in which the votes are tallied to produce a winner (or winners if electing more than one candidate). We assume that the voting aggregation function will select exactly one winner, so ties are always broken[6]. Note that elements of the set $\mathcal{B}(\mathcal{O})^n$ represent votes cast by the agents. An element $\boldsymbol{v} \in \mathcal{B}(\mathcal{O})^n$ is called a *vote profile*. A tallying function $\text{Ag}_v : \mathcal{B}(\mathcal{O})^n \to \mathcal{O}$ maps vote profiles to candidates.

Given agent i's preference P_i, let $S(v, P_i)$ mean that the vote v is a sincere vote corresponding to P_i. For example, in plurality voting, the only sincere vote is a vote for the maximally ranked candidate under P_i. By contrast, in approval voting, there could be many sincere votes, i.e., those votes where, if candidate o is approved, so is any higher ranking o'. Then $\mathcal{B}(\mathcal{O})_i = \{v | S(v, P_i)\}$ is the set of votes which faithfully represent i's preference. The voter i is said to *strategize* if i selects a vote v that is not in the set $\mathcal{B}(\mathcal{O})_i$.

In what follows we assume that when an agent votes, the agent is selecting a preference in the set **Pref** instead of an element of $\mathcal{B}(\mathcal{O})$. A vote is a preference; a vote profile is a *vector* of preferences, denoted by $\boldsymbol{P}$.[7]

Assume that the agents' true preferences are $\boldsymbol{P}^* = (P_1^*, \dots, P_n^*)$ and *fixed* for the remaining discussion. Given a profile $\boldsymbol{P}$ of *actual* votes, we ask whether agent i will change its vote if given another chance to express its preference. Let $\boldsymbol{P}_{-i}$ be the vector of all *other* agents' preferences. Then given $\boldsymbol{P}_{-i}$ and i's true preference P_i^*, there will be a (nonempty) set X_i of those preferences that are i's best response to $\boldsymbol{P}_{-i}$. Suppose that $f_i(\boldsymbol{P}_{-i})$ selects one such best response from X_i.[8] Then $f(\boldsymbol{P}) = (f_1(\boldsymbol{P}_{-1}), \dots, f_n(\boldsymbol{P}_{-n}))$. We call f a *strategizing* function. If $\boldsymbol{P}$ is a fixed point of f (i.e., $f(\boldsymbol{P}) = \boldsymbol{P}$), then $\boldsymbol{P}$ is a *stable* outcome. In other words, such a fixed point $\boldsymbol{P}$ of f is a Nash equilibrium. We define f^n recursively by $f^1(\boldsymbol{P}) = f(\boldsymbol{P})$, $f^n = f(f^{n-1}(\boldsymbol{P}))$, and say that f is **stable at level** n if $f^n(\boldsymbol{P})) = f^{n-1}(\boldsymbol{P})$. It is clear that if f is stable at level n, then f is stable at all levels m where $m \geq n$. Also, if the initial preference of the $\boldsymbol{P}$ is a fixed point of f then all levels are stable.

Putting everything together, we can now define a voting model.

Definition 2 (Voting Model). *Given a set of agents $\mathcal{A}$, candidates $\mathcal{O}$, a* **voting model** *is a 5-tuple $\langle \mathcal{A}, \mathcal{O}, \{P_i^*\}_{i \in \mathcal{A}}, \text{Ag}, f \rangle$, where P_i^* is voter i's true pref-*

[6] [2] shows that the Gibbard-Satterthwaite theorem holds when ties are permitted.

[7] This does not quite work for approval voting where P does not fully determine the sincere vote v, but we will ignore this issue here, as it does not apply in the case of plurality elections, whether of one, or of many 'winners'.

[8] Note that P_i may itself be a member of X_i in which case we shall assume that $f(P_i) = P_i$.

erence; Ag *is an aggregation function with domain and range as defined above; f is a strategizing function.*

Note that in our definition above, we use aggregation functions rather than tallying functions (which pick a winning candidate). This is because we can view tallying functions as selecting a 'winner' from the output of an aggregation function. So in our model, the result of an election is a *ranking* of the candidates. This allows our results to apply not only to conventional plurality voting, but also to those situations where more than one candidate is to be elected. They require some modification to apply to approval voting, as the ballot is not then determined by the preference ordering but also needs a cut-off point between approved and 'dis-approved' candidates.

The following example demonstrates the type of analysis that can be modeled using a strategizing function.

Example 1. Suppose that there are four candidates $\mathcal{O} = \{o_1, o_2, o_3, o_4\}$ and five groups of voters: A, B, C, D and E. Suppose that the sizes of the groups are given as follows: $|A| = 40$, $|B| = 30$, $|C| = 15$, $|D| = 8$ and $|E| = 7$. We assume that all the agents in each group have the same true preference and that they all vote the same way. Suppose that the tallying function is plurality vote. We give the agents' true preferences and the summary of the four elections in the table below. The winner in each round is in boldface.

$$P_A^* = (o_1, o_4, o_2, o_3)$$
$$P_B^* = (o_2, o_1, o_3, o_4)$$
$$P_C^* = (o_3, o_2, o_4, o_1)$$
$$P_D^* = (o_4, o_1, o_2, o_3)$$
$$P_E^* = (o_3, o_1, o_2, o_4)$$

Size	Group	I	II	III	IV
40	A	$\mathbf{o_1}$	o_1	o_4	$\mathbf{o_1}$
30	B	o_2	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
15	C	o_3	$\mathbf{o_2}$	$\mathbf{o_2}$	o_2
8	D	o_4	o_4	o_1	o_4
7	E	o_3	o_3	o_1	$\mathbf{o_1}$

The above table can be justified by assuming that all agents use the following protocol. If the current winner is o, then agent i will switch its vote to some candidate o' provided 1) i prefers o' to o, and 2) the current total for o' plus agent i's votes for o' is greater than the current total for o. By this protocol an agent (thinking only one step ahead) will only switch its vote to a candidate which is currently not the winner.

In round I, everyone reports their top choice and o_1 is the winner. C likes o_2 better than o_1 and its own total plus B's votes for o_2 exceed the current votes for o_1. Hence by the protocol, C will change its vote to o_2. A will not change its vote in round II since its top choice is the winner. D and E also remain fixed since they do not have an alternative like o' required by the protocol. In round III, group A changes its vote to o_4 since it is preferred to the current winner (o_2) and its own votes plus D's current votes for o_4 exceed the current votes for o_2. B and C do not change their votes. For B's top choice o_2 is the current winner and as for C, they have no o' better than o_2 which satisfies condition 2). Ironically, Group D and E change their votes to o_1 since it is prefered to the current winner is o_2 and group A is currently voting for o_1. Finally, in round IV, group A notices that E is voting for o_1 which A prefers to o_4 and so changes its votes back to o_1. The situation stabilizes with o_1 which, as it happens, is also the Condorcet winner.

Much more can be said about the above analysis, but this is a topic for a different paper. We now point out that for every aggregation function Ag and any strategizing f, there *must* be instances in which f never stabilizes:

Theorem 1. *For any given tallying function* Ag_v*, there exists an initial vector of preferences such that* f *never stablizes.*

This follows easily from the Gibbard-Satterthwaite theorem. Suppose not, then we show that there is a strategy-proof tallying function contradicting the Gibbard-Satterthwaite theorem. Suppose that Ag_v is an arbitrary tallying function and $\boldsymbol{P}^*$ the vector of true preferences. Suppose there always is a level k at which f stabilizes given the agents' true preferences $\boldsymbol{P}^*$. But then define Ag' to be the outcome of applying Ag_v to $f^k(\boldsymbol{P}^*)$ where $\boldsymbol{P}^*$ are the agents' true preferences. Then given some obvious conditions on the strategizing function f, Ag' will be a strategy-proof tallying function contradicting the Gibbard-Satterthwaite theorem. Hence there *must be* situations in which f never stabilizes.

Since our candidate and agent sets are finite, if f does not stabilize then f **cycles**. We say that f has a cycle of length n if there are n different votes $\boldsymbol{P}_1, \ldots \boldsymbol{P}_n$ such that $f(\boldsymbol{P}_i) = \boldsymbol{P}_{i+1}$ for all $1 \leq i \leq n-1$ and $f(\boldsymbol{P}_n) = \boldsymbol{P}_1$.

3 Dependency on knowledge

Suppose that agent i knows the preferences of the other agents, and that no other agent knows agent i's preference (and agent i knows *this*). Then i is in a very privileged position, where its preferences are completely secret, but it knows it can strategize using the preferences of the other agents. In this case, i will always know *when* to strategize and when the new outcome is 'better' than the current outcome. But if i only knows the preferences of a certain subset B of the set $\mathcal{A}$ of agents, then there still may be a set of possible outcomes that it could force. Since i only knows the preferences of the agents in the set B, any strategy P will generate a set of possible outcomes. Suppose that there are two strategies P and P' that agent i is choosing between. Then the agent is choosing between two different sets of possible outcomes. Some agents may only choose to strategize if they are *guaranteed* a better outcome. Other agents may strategize if there is even a small chance of getting a better outcome and no chance of getting a worse outcome. We will keep this process of choosing a strategy abstract, and only assume that every agent will in fact choose one of the strategies available to it. Let S_i be agent i's strategy choice function, which accepts the votes of a group of agents and returns a preference P that may result in a better outcome for agent i given the agents report their current preference. We will assume that if $B = \emptyset$ then S_i picks P_i^*. That is, agents will vote according to their true preferences unless there is more information.

As voting takes place or polls reveal potential voting patterns, the facts that each agent knows will change. We assume that certain agents may be in a more privileged position than other agents. As in [14], define a **knowledge graph** to be any graph with $\mathcal{A}$ as its set of vertices. If there is an edge from i to j, then we assume that agent i knows agent j's current vote, i.e., how agent j voted in the current election. Let $\mathcal{K} = (\mathcal{A}, E_{\mathcal{K}})$ be a knowledge graph ($E_{\mathcal{K}}$ is the set of edges of $\mathcal{K}$). We assume that $i \in \mathcal{A}$ knows the current votes of all agents accessible from i. Let $B_i = \{j \mid \text{ there is an edge from } i \text{ to } j \text{ in } \mathcal{K}\}$. Then S_i will select the strategy that agent i would prefer given how agents in B_i voted currently.

We clarify the relationship between a knowledge graph and the existence of a cycle in the knowledge graph $\mathcal{K} = (\mathcal{A}, E_{\mathcal{K}})$ by the following:

Theorem 2. *Fix a voting model $\langle \mathcal{A}, \mathcal{O}, \{P_i^*\}_{i \in \mathcal{A}}, \text{Ag}, f \rangle$ and a knowledge graph $\mathcal{K} = (\mathcal{A}, E_{\mathcal{K}})$. If $\mathcal{K}$ is directed and acyclic then the strategizing function f will stabilize at level k, where k is the height of the graph $\mathcal{K}$; f can cycle* only *if the associated knowledge graph has a cycle.*

Proof. Since $\mathcal{K}$ is a directed acyclic graph, there is at least one agent i such that $B_i = \emptyset$. By assumption such an agent will vote according to P_i^* at every stage. Let

$$A_0 = \{i \mid i \in \mathcal{A} \text{ and } B_i = \emptyset\}$$

and

$$A_k = \{i \mid \text{ if there is } (i, j) \in E_{\mathcal{K}}, \text{then } j \in A_l \text{ for } l < k\}$$

Given (by induction on k) that the agents in A_{k-1} stabilized by level $k - 1$, an agent $i \in A_k$ need only wait $k - 1$ rounds, then choose the strategy according to S_i. □

The following is an example of a situation in which the associated strategizing function never stabilizes:

Example 2. Consider three candidates $\{a, b, c\}$ and 100 agents connected by a complete knowledge graph. Suppose that 40 agents prefer $a > b > c$ (group I), 30 prefer $b > c > a$ (group II) and 30 prefer $c > a > b$ (group III). If we assume that the voting rule is simple majority, then after reporting their initial preferences, candidate a will be the winner with 40 votes. The members of group II dislike a the most, and will strategize in the next election by reporting $c > b > a$ as their preference. So, in the second round, c will win. But now, members of group I will report $b > a > c$ as their preference, in an attempt to draw support away from their lowest ranked candidate. c will still win the third election, but by changing their preferences (and making them public) group I sends a signal to group II that it should report its true preference - this will enable group I to have its second preferred candidate b come out winner. This cycling will continue indefinitely; b will win for two rounds, then a for two rounds, then c for two, etc.

4 An epistemic logic for voting models

In this section we define an epistemic logic for reasoning about voting models. In Example 2, it is clear that voters are reasoning about the states of knowledge of other voters and furthermore, an agent reasons about the change in states of knowledge of other voters on receipt of information on votes cast by them. We now sketch the details of a logic $\mathcal{KV}$ for reasoning about knowledge and the change of knowledge in a fixed voting model $\mathcal{V}$.

4.1 The logic $\mathcal{KV}$ - syntax

In this section we will assume that each vote is an expressed preference, which may or not be the true preference of an agent. So the expression 'preference' without the qualifier 'true' will simply mean an agent's current vote. We assume that for each preference P there is a symbol P that represents it. There are then two types of primitive propositions in $\mathcal{L}(\mathcal{KV})$. First, there are statements with the content "agent i's preference is P". Let P_i represent such statements. Secondly, we include statements with the content "P is the current outcome of the aggregation function". Let P_O represent such statements.

Our language includes the standard boolean connectives, an epistemic modality K_i indexed by each agent i plus an additional modality $\Diamond_i$ (similarly indexed). Formulas in $\mathcal{L}(\mathcal{KV})$ take the following syntactic form:

$$\phi := p \mid \neg\phi \mid \phi \wedge \psi \mid K_i\phi \mid \Diamond_i\phi$$

where p is a primitive proposition, $i \in \mathcal{A}$. We use the standard definitions for $\vee, \rightarrow$ and the duals $L_i, \Box_i$. $K_i\phi$ is read as "agent i knows ϕ; $\Diamond_i\phi$ is read as "after agent i strategizes, ϕ becomes true".

4.2 The logic $\mathcal{KV}$ - semantics

Before specifying a semantics we make some brief remarks on comparing preferences. Strategizing means reporting a preference different from your true preference. An agent will strategize if by reporting a preference other than its true preference, the outcome is 'closer' to its true preference than the outcome it would have obtained had it reported its true preference originally. Given preferences P, Q, R, we use the notation $P \sqsubseteq_R Q$ to indicate that P is at least as compatible with R as Q is. Given the above ternary relation, we can be more precise about when an agent will strategize. Given two preferences P and Q, we will say that an agent whose true preference is R prefers P to Q if $P \sqsubseteq_R Q$ holds. That is, i prefers P to Q if P is at least as 'close' to i's true preference as Q is.

We assume the following conditions on $\sqsubseteq$. For arbitrary preferences P, Q, R, S:

1. (Minimality) $R \sqsubseteq_R P$
2. (Reflexivity) $P \sqsubseteq_R P$

3. (Transitivity) If $P \sqsubseteq_R Q$ and $Q \sqsubseteq_R S$, then $P \sqsubseteq_R S$.
4. (Pareto Invariance) Suppose that $R = (o_1, \ldots, o_m)$ and $P = (o'_1, \ldots, o'_m)$ and Q is obtained from P by swapping o'_i and o'_j for some $i \neq j$. If $R(o'_i, o'_j)$ and $P(o'_i, o'_j)$, i.e., R agrees with P on o'_i, o'_j and disagrees with Q, then P must be at least as close to R as Q ($P \sqsubseteq_R Q$).

(Minimality) ensures that a true preference is always the most desired outcome. (Reflexivity) and (Transitivity) carry their usual meanings. As for Pareto invariance, note that swapping o'_i, o'_j may also change other relationships. Our example below will show that this is not a problem.

The following is an example of an ordering $\sqsubseteq_R$ satisfying the above conditions. Let $R = (o_1, \ldots, o_m)$. For each vector P, suppose that $c_P(o_i)$ is the count of o_i in vector P, i.e., the numeric position of o_i numbering from the right. For any vector P, let $V_R(P) = c_R(o_1)c_P(o_1) + \cdots + c_R(o_m)c_P(o_m)$. This assigns the following value to R, $V_R(R) = m^2 + (m-1)^2 + \cdots + 1^2$. We say that P is closer to R than Q iff $V_R(P)$ is greater than $V_R(Q)$. This creates a strict ordering over preferences, which can be weakened to a partial order by composing V_R with a weakly increasing function.[9]

Let $\mathcal{V} = \langle \mathcal{A}, \mathcal{O}, \{P^*_i\}_{i \in \mathcal{A}}, \mathrm{Ag}, f \rangle$ be a fixed voting model. We define a Kripke structure for our bi-modal language based on $\mathcal{V}$. States in this structure are vectors of preferences[10] together with the outcome of the aggregation function. The set of states W is defined as follows:

$$W = \{(\boldsymbol{P}, O) \mid \boldsymbol{P} \in \mathbf{Pref}^n, \quad Ag(\boldsymbol{P}) = O\}$$

Intuitively, given a state $(\boldsymbol{P}, O)$, $\boldsymbol{P}$ represents the preferences that are reported by the agents and O is the outcome of the aggregation function applied to $\boldsymbol{P}$. So states of the world will be complete descriptions of stages of elections.

Our semantics helps clarify our decision to use two modalities. Let $(\boldsymbol{P}, O)$ be an element of W. To understand the strategizing modality, note that when an agent strategizes it only changes the ith component of $\boldsymbol{P}$, i.e., the accessible worlds for this modality are those in which the remaining components of $\boldsymbol{P}$ are fixed. For the knowledge modality note that all agents know how they voted, which implies that accessible worlds for this modality are those in which the ith component of $\boldsymbol{P}$ remains fixed while others vary.

We now define accessibility relations for each modality. Since the second component of a state can be calculated using Ag we write $\boldsymbol{P}$ for $(\boldsymbol{P}, O)$. For the knowledge modality, we assume that the agents know how they voted and so define for each $i \in \mathcal{A}$ and preferences $\boldsymbol{P}, \boldsymbol{Q}$:

$$(\boldsymbol{P}, O) R_i (\boldsymbol{Q}, O') \quad \text{iff} \quad P_i = Q_i$$

[9] Plurality cannot be produced this way, but other functions satisfying 1–4 can easily be found that do.

[10] Defining Kripke structures over agents' preferences has been studied by other authors. [18] has a similar semantics in a different context.

The above relation does not take into account the fact that some agents may be in a more privileged position than other agents, formally represented by the knowledge graph from the previous section. If we have fixed a knowledge graph, then agent i not only knows how it itself voted, but also the (current) preferences of each of the agents reachable from it in the knowledge graph. Let $\mathcal{K} = (\mathcal{A}, E_{\mathcal{K}})$ be a knowledge graph, and recall that B_i is the set of nodes reachable from i. Given two vectors of preferences $\boldsymbol{P}$ and $\boldsymbol{Q}$ and a group of agents $G \subseteq \mathcal{A}$, we say $P_G = Q_G$ iff $P_i = Q_i$ for each $i \in G$. We can now define an epistemic relation based on $\mathcal{K}$:

$$(\boldsymbol{P}, O) R_i^{\mathcal{K}} (\boldsymbol{Q}, O') \quad \text{iff} \quad P_i = Q_i \text{ and } P_{B_i} = Q_{B_i}$$

Clearly for each agent i and knowledge graph $\mathcal{K}$, $R_i^{\mathcal{K}}$ is an equivalence relation; hence each K_i is an **S5** modal operator. The exact logic for the strategizing modalities depends on the properties of the ternary relation $\sqsubseteq$.
For the strategizing modalities, we define a relation $A_i \subseteq W \times W$ as follows. Given preferences $\boldsymbol{P}, \boldsymbol{Q}$:

$$(\boldsymbol{P}, O) A_i (\boldsymbol{Q}, O') \quad \text{iff} \quad P_{-i} = Q_{-i} \text{ and } O' \sqsubset_{P_i^*} O$$

where P_{-i} is all components of $\boldsymbol{P}$ except for the ith component. So, $(\boldsymbol{P}, O)$ and $(\boldsymbol{Q}, O')$ are A_i related iff they have the same jth component for all $j \neq i$ and agent i prefers outcome O' to outcome O relative to i's true preference P_i^*.

An **election** is a sequence of states. We say that an election $E = (s_1, s_2, \ldots, s_n)$ *respects* the strategizing function f if $f(s_i) = s_{i+1}$ for $i = 1, \ldots, n-1$. We assume always that f is such that $f(s) = s$ unless the agent *knows* that it can strategize and get a better outcome, and *how* it should so strategize. A model for $\mathcal{V}$ is a tuple $\mathbb{M} = \langle W, R_i, A_i, V \rangle$ where $V : W \to 2^{\Phi_0}$ (where Φ_0 is the set of primitive propositions). We assume that all relations R_i are based on a given knowledge graph $\mathcal{K}$. Let $(\boldsymbol{P}, O) \in W$ be any state; we define truth in a model as follows:

1. $(\boldsymbol{P}, O) \models p$ iff $p \in V(\boldsymbol{P}, O)$ and $p \in \Phi_0$
2. $(\boldsymbol{P}, O) \models \neg\phi$ iff $(\boldsymbol{P}, O) \not\models \phi$
3. $(\boldsymbol{P}, O) \models \phi \wedge \psi$ iff $(\boldsymbol{P}, O) \models \phi$ and $\boldsymbol{P} \models \psi$
4. $(\boldsymbol{P}, O) \models K_i\phi$ iff for all $(\boldsymbol{Q}, O')$ such that $(\boldsymbol{P}, O) R_i^{\mathcal{K}} (\boldsymbol{Q}, O')$, $(\boldsymbol{Q}, O') \models \phi$
5. $(\boldsymbol{P}, O) \models \Diamond_i\phi$ iff there is $(\boldsymbol{Q}, O')$ such that $(\boldsymbol{P}, O) A_i (\boldsymbol{Q}, O')$ and $(\boldsymbol{Q}, O') \models \phi$

Nothing in our definition of a model forces primitive propositions to have their intended meaning. We therefore make use of the following definition.

Definition 3. *A valuation function V is an* **appropriate valuation** *for a model $\mathbb{M}$ iff V satisfies the following conditions. Let $\mathcal{V} = \langle \mathcal{A}, \mathcal{O}, \{P_i^*\}_{i \in \mathcal{A}}, \mathrm{Ag}, f \rangle$ be a voting model and $\mathbb{M}$ a model based on $\mathcal{V}$. Let $(\boldsymbol{P}, O) \in W$ be any state. Then:*

1. *For each $i \in \mathcal{A}$, $\mathsf{P}_i \in V(\boldsymbol{P}, O)$ iff P represents the preference P_i.*
2. *For each P_O, $\mathsf{P}_O \in V(\boldsymbol{P}, O)$ iff P represents O.*

We assume that valuation functions are appropriate for the corresponding model.

The following formula implies strategizing for an individual agent. It says that agent i knows that the outcome is P_O and by reporting a different preference a preferred outcome can be achieved.

$$K_i(\mathsf{P}_O \wedge \Diamond_i \top)$$

We are now in a position to present our last main result.

Theorem 3. *Given a voting system $\mathcal{V} = \langle \mathcal{A}, \mathcal{O}, \{P_i^*\}_{i \in \mathcal{A}}, \mathrm{Ag}, f \rangle$, a knowledge graph $\mathbb{K}$ and a model $\mathbb{M}$ for $\mathcal{V}$, let E be an election that respects the strategizing function f. If there is a state $\boldsymbol{P}$ such that $E_l = \boldsymbol{P}$ for some l and $\boldsymbol{P} \models \neg K_i(\mathsf{P}_O \wedge \Diamond_i \top)$ for all i, then $\boldsymbol{P}$ is a fixed point of f. Equivalently, Given an election E that respects f and some k such that $E_{k+1} \neq E_k$, i.e., E_k is not a fixed point of f, then $\exists i \in \mathcal{A}$ such that:*

$$E_k \models K_i(\mathsf{P}_O \wedge \Diamond_i \top)$$

That is, if an agent strategizes at some stage in the election then the agent knows that this strategizing will result in a preferred outcome.

5 Conclusion

We have explored some properties of strategic voting and noted that the Gibbard-Satterthwaite theorem only applies in those situations where agents can obtain the appropriate knowledge. Note that our example in the Introduction showed how strategizing can lead to a rational outcome in elections. In our example the Condorcet winner - the winner in pairwise head-to-head contests - was picked via strategizing. Since our framework makes it possible to view opinion polls as the $n-1$ stages of an n-stage election, it implies that communication of voters' preferences and the results of opinion polls can play an important role in ensuring rational outcomes to elections. A similar line of reasoning in a different context can be found in [15]. Put another way, while the Gibbard-Satterthwaite theorem implies that we are stuck with voting mechanisms susceptible to strategizing, our work indicates ways for voters to avoid irrational outcomes using such mechanisms. Connections such as those explored in this paper are also useful in deontic contexts [10, 16] i.e., an agent can only be obligated to take some action if the agent is in possession of the requisite knowledge.

For future work, we note that in this study, we left the definition of the agents' strategy choice function informal, thus assuming that agents have some way of deciding which preference to report if given a choice. This can be made more formal. We could then study the different strategies available to the agents. For example, some agents may only choose to strategize if they are *guaranteed* to get a better outcome, whereas other agents might strategize even if there is only a small chance of getting a better outcome.

Another question suggested by this framework is: what are the effects of different levels of knowledge of the current preferences on individual strategy choices? Suppose that among agent i and agent j, both i and j's true preferences are common knowledge. Now when agent i is trying to decide whether or not to strategize, i knows that j will be able to simulate i's reasoning. Thus if i chooses a strategy based on j's true preference, i knows that j will choose a strategy based on i's choice of strategy, and so i must choose a strategy based on j's response to i's original strategy. We conjecture that if there is only pairwise common knowledge among the agents of the agents' true preferences, then the announcement of the agents' true preferences is a stable announcement.

On a technical note, the logic of knowledge we developed uses **S5** modalities. We would like to develop a logic that uses **KD45** modalities - i.e., a logic of belief. This is because beliefs raise the interesting issue that a voter - or groups of voters - can have possibly inconsistent beliefs about other voters' preferences, while this variation is not possible in the knowledge case. Another area of exploration will be connections with other distinct approaches to characterize game theoretic concepts in modal logic such as [8, 3, 18]. Lastly, a deeper formal understanding of the relationship between the knowledge and strategizing modalities introduced in this paper will become possible after the provision of an appropriate axiom system for $\mathcal{KV}$. Our work is a first step towards clarifying the knowledge-theoretic properties of voting, but some insight into the importance of states of knowledge and the role of opinion polls is already at hand.

References

1. K. J. Arrow. *Social choice and individual values (2nd ed.)*. Wiley, New York, 1963.
2. Jean-Pierre Benoit. Strategic manipulation in games when lotteries and ties are permitted. *Journal of Economic Theory*, 102:421–436, 2002.
3. Giacomo Bonanno. Modal logic and game theory: two alternative approaches. *Risk Decision and Policy*, 7:309–324, 2002.
4. Steven J. Brams. Voting Procedures. In *Handbook of Game Theory*, volume 2, pages 1055–1089. Elsevier, 1994.
5. Steven J. Brams and Peter C. Fishburn. Voting Procedures. In *Handbook of Social Choice and Welfare*. North-Holland, 1994.
6. Patricia Everaere, Sebastien Konieczny, and Pierre Marquis. On merging strategy-proofness. In *Proceedings of KR 2004*. Morgan-Kaufmann, 2004.
7. Allan Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601, 1973.
8. Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. A modal characterization of nash equilibira. *Fundamenta Informaticae*, 57(2-4):281–321, 2002.
9. Sébastien Konieczny and Ramón Pino-Pérez. On the logic of merging. In A. G. Cohn, L. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR '98)*, pages 488–498, San Francisco, California, 1998. Morgan Kaufmann.
10. Alessio Lomuscio and Marek Sergot. Deontic interpreted systems. *Studia Logica*, 75, 2003.

11. Pedrito Maynard-Zhang and Daniel Lehmann. Representing and aggregating conflicting beliefs. *Journal of Artificial Intelligence*, 19:155–203, 2003.
12. Pedrito Maynard-Zhang and Yoav Shoham. Belief fusion: Aggregating pedigreed belief states. *Journal of Logic, Language and Information*, 10(2):183–209, 2001.
13. Thomas Meyer, Aditya Ghose, and Samir Chopra. Social choice theory, merging and elections. In *Proceedings of Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU-2001*. Springer-Verlag, 2001.
14. Eric Pacuit and Rohit Parikh. A logic of communication graphs. In *Proceedings of DALT-04*. Springer-Verlag, 2004.
15. Rohit Parikh. Social software. *Synthese*, 132, 2002.
16. Rohit Parikh, Eric Pacuit, and Eva Cogan. The logic of knowledge based obligations. In *Proceedings of DALT-2004*. Springer Verlag, 2004.
17. Mark Satterthwaite. *The Existence of a Strategy Proof Voting Procedure: a Topic in Social Choice Theory*. PhD thesis, University of Wisconsin, 1973.
18. J. van Benthem. Rational dynamics and epistemic logic in games. Technical report, ILLC, 2003.