Designing (BREXIT) Referendum: An Economist's Pessimistic Perspective

Siyang Xiong*

November 1, 2016

Abstract

Recently, there is a trend all over the world (e.g., in Colombia, Italy, Switzerland, and U.K.) that referendums are used to determine social decisions. This paper studies designing a referendum. In several setups (e.g., different solution concepts and different types of mechanisms), we prove three impossibility results, i.e., a social goal can be achieved if and only if it is dictatorial.

In a petition to U.K. Parliament singed by more than 2.5 million people, it is proposed that a second referendum will be held if the result of the first referendum does not meet a pre-determined condition. We also provide a possibility result, which fully characterizes when such an option of a second referendum may help us achieve a non-dictatorial social goal.

*Department of Economics, University of Bristol, BS8 1TN, United Kingdom, siyang.xiong@bristol.ac.uk

Democracy: an organization or situation in which everyone is treated equally and has equal rights.

— one of the two definitions[1] *listed in Merriam-Webster's Learner's Dictionary*

And without God, democracy will not and cannot long endure.

—Ronald Reagan (Congress (2009))

Democracy is the worst form of government, except for all those other forms that have been tried from time to time.

—Sir Winston Churchill

1 Introduction

Suppose a society (i.e., a group of agents) needs to jointly choose one option out of a finite set of several social outcomes — let us use $\mathcal{A}$ to denote the set. It is has long been known that if there are three or more alternatives in $\mathcal{A}$, it is difficult to achieve a non-dictatorial social goal, e.g., Arrow's Impossibility Theorem (Arrow (1963)), the Gibbard-Satterthwaite Theorem (Gibbard (1973), Satterthwaite (1975)), the Muller-Satterthwaite Theorem (Muller and Satterthwaite (1977)).

On the contrary, when $\mathcal{A}$ contains only two alternatives, the usual wisdom is that social decisions are much more permissive, and many non-dictatorial social goals can be achieved. In particular, the majority rule is usually considered as being superior in the two-alternative environment. For example, the famous May's Theorem (May (1952)) says that the majority voting rule is the only one that satisfies anonymity, neutrality and monotonicity (see more discussion in Moulin (1988)).[2]

[1]This definition is about the principle idea of demorcracy, and the other definition is about a specific political institution that is built on the idea. We present and discuss the second definition in Section 8.

[2]Besides May's axiomatic characterization, Balinski and Laraki (2016) discuss several other reasons why the majority voting rule is universally accepted.

In light of this positive result, there is a recent trend all over the world that referendums (i.e., voting on two social outcomes among all people in the society) are used to determine social decisions, e.g., Scottish independence referendum in 2014, Greek bailout referendum in 2015, Switzerland referendums in 2016, UK BREXIT referendum in 2016, Colombia peace agreement referendum in 2016, Italian constitutional referendum in 2016. Most of them use the majority rule. For example, the result of Scottish referendum is "55.3% vs 44.7%" with "remain in U.K." winning, and the result of Colombia peace agreement referendum is "50.2% vs 49.8%" with no-agreement winning.

This trend of referendum seems to grow more and more strong. For instance, in year 2016 alone, three referendums have been held in Switzerland in February, June and September, and a fourth one will be held in November. The subjects of the referendums range from "retirement system" to "road tunnel reconstruction," e.g., the following table lists three subjects in Switzerland referendums.

three subjects in the 2016 Switzerland referendums

subject	for	against	result
Gotthard road tunnel reconstruction	57%	43%	accepted
Retirement system	41%	59%	rejected
Green economy	36%	64%	rejected

Also, in his presidential campaign, Mr. Nicolas Sarkozy announces that two referendums[3] will be held if he is elected as the president of France in 2017. As reported by Le Monde (the French newspaper), Mr. Sarkozy argues: *To choose a referendum is to "choose democracy."*[4]

Among these referendums, arguably, the most influential and controversial one is the U.K. BREXIT referendum, which was held on June 23, 2016, with exit winning by "51.9% vs 48.1%." The economic impact of BREXIT is huge. On June 24 (i.e., one day after the voting), the stock indexes of many countries fell sharply (see the table below), and the exchange rate of British pound to US dollar tumbled to 1.3224 USD/1 GBP, the lowest

[3] The subjects of the referendums are: (1) the suspension of automatic family reunion and (2) administrative detention of individuals who are most dangerous for state security.

[4] La solution référendaire, c'est faire le choix de la démocratie , a justifié M. Sarkozy, qui avait déjà proposé des référendums lors de sa campagne de 2012, invoquant le général de Gaulle. – LeMonde (2016).

level since 1985.

stock index changes in several countries on June 24, 2016

London FTSE	France CAC	Germany DAX	Italy FTSE MIB	Spain IBEX	Japan Nikkei
↓ 3.2%	↓ 8%	↓ 6.8%	↓ 12%	↓ 12%	↓ 7.9%

Before the BREXIT voting, a petition was set up by Mr. William Oliver Healey at the website of U.K. parliament:

> *We the undersigned call upon HM Government to implement a rule that if the remain or leave vote is less than 60% based on a turnout less than 75% there should be another referendum.*

At the time when the BREXIT result was revealed, 22 people signed the petition. However, by June 25 (i.e., two days after the BREXIT voting), more than 2.5 million people have signed the petition.

Clearly, quite a few people think that the original voting mechanism for BREXIT was not successful, and demand a second referendum using, hopefully, a better voting mechanism. However, in what sense did the original voting mechanism fail? Furthermore, in what sense would a ″better″ voting mechanism succeed? In particular, the petition proposes a voting rule which involves an option of a second referendum. Does such an option of a second referendum help? Suppose an economist (i.e., a mechanism designer) is called upon to solve this problem, i.e., design an ″optimal″ voting mechanism to ″properly″ aggregate people's preference. Can the designer find such a mechanism?

To answer these questions, we define the problem rigorously, and model it as a classical implementation problem *à la* Maskin (1999). Specifically, we assume $\mathcal{A} = \{L, R\}$, where, for instance, L and R stand for left (or liberal) and right (or conservative), respectively. Every agent in the society has her[5] own strict preference on L and R, i.e., she strictly prefers either L to R, or R to L. A preference profile specifies the strict preference of every agent in the society, and let $\mathcal{P}$ denote the set of all preference profiles that are deemed possible. For each preference profile $P \in \mathcal{P}$, all people in the society, *a priori*, agree on a

[5]Throughout the paper, we use ″she″ to denote an agent.

desired social outcome — we use the function $f : \mathcal{P} \longrightarrow \mathcal{A}$ to represent all of the agreed outcomes for all preference profiles in $\mathcal{P}$. Therefore, the *goal* of the mechanism designer is to select $f(P)$ when P is the true preference profile. The function f is called a social choice function (hereafter, SCF).

However, a fundamental difficulty in mechanism design is that the designer cannot directly observe P (i.e., agents' preference). Furthermore, agents are not obliged to tell the designer their true preference. Hence, the tool that the designer can use is to build an appropriate game (or equivalently, a mechanism), so that incentive compatibility of agents under any true preference profile P would enforce the agents to take the targeted strategies in the game, and to induce the desired outcome $f(P)$.

Roughly, a game is a tuple $\left\langle S \equiv (S_i)_{i \in I}, \, g : S \longrightarrow \mathcal{A} \right\rangle$,[6] where each S_i is a finite set of strategies for agent i. When a strategy profile $s \in S$ is chosen by the agents, the designer selects $g(s)$ as the final choice of the society. To predict outcomes of a game, we have to fix a solution concept. A solution is a strategy profile in the game that is considered as a reasonable outcome. Notable solution concepts include Nash equilibria (Nash (1950, 1951)), rationalizability (Bernheim (1984), Pearce (1984)), etc.

Given a particular solution concept, a preference profile P and a game G, let $\Gamma(G, P)$ denote the set of all solutions in the game. Then, the objective of the designer is to construct a game $G^* = \langle S, \, g : S \longrightarrow \mathcal{A} \rangle$, such that under any preference profile $P \in \mathcal{P}$, we have

$$g\left[\Gamma\left(G^*, P\right)\right] = \left\{f(P)\right\}.$$

That is, for any $P \in \mathcal{P}$ and any solution $s \in \Gamma(G^*, P)$, the outcome $g(s)$ selected by the game matches the targeted social outcome $f(P)$. If such a game exits, we say f can be implemented in this particular solution concept.

According to Merriam-Webster's Learner's dictionary cited at the beginning of the paper, democracy is *an organization or situation in which everyone is treated equally and has equal rights.* The opposite of democracy is *dictatorship*, i.e., social decisions are fully determined by the preference of *one* particular agent (i.e., the dictator) in the society. We say a SCF f is dictatorial, if a dictator exists.

[6] We consider more general mechanisms in Sections 4, 5 and 6.

In this paper, we study what kind of SCFs can be implemented in different solution concepts and different types of mechanisms. In particular, we provide three impossibility results of the following form: a SCF can be implemented if and only if it is dictatorial, which challenges the usual wisdom that social decisions are permissive in the two-alternative environment. Different from the axiomatic approach in May (1952), we take a game-theoretic approach, and the message we bring is that, if we take the entirety of game-theoretic analysis seriously, it is still difficult to achieve a non-dictatorial social goal, even in the two-alternative environment.

First, we study Nash equilibria and focus on voting mechanisms which generalize all of the existing voting rules. A voting mechanism is a game in which agents can vote for the two social outcomes, and for any social outcome, the more votes, the more likely it is selected.

Our first impossible result is that a SCF can be implemented in Nash equilibria by a voting mechanism if and only if it is dictatorial. One implication of the result is that none of the existing voting rules can help us implement a non-dictatorial social choice function.

Two assumptions on games are critical to achieve our impossibility result. First, only deterministic mechanisms are allowed. Second, only voting mechanisms are allowed. Our second and third main results extend the impossibility to the cases in which we relax one or both of the two assumptions.

We then propose a general notion of stochastic voting mechanisms, which allow for both objective and subjective lotteries on social outcomes. The option of a second referendum discussed above can be thought as a subjective lottery in the eyes of the agents.[7] Furthermore, we consider an assumption on preference, called "uncertainty in swing voters," which is likely to hold in many situations, e.g., the US electoral voting. For any $\alpha \in \mathcal{A}$, agent i is called an α-voter if i always prefers α to the other social outcome under any possible preference profile. For example, in BREXIT, Mr. David Cameron and Mr. Nigel Farage are the stay-voter and the exit-voter, respectively. A swing voter is an agent who is neither an L-voter nor a R-voter, i.e., a swing voter prefers L to R under some prefer-

[7] Regarding the second referendum, each agent forms her subjective belief regarding the outcomes L and R, i.e., it is a subjective lottery in the eyes of agents.

ence profile, and prefers R to L under another. "Uncertainty in swing voters" holds if for any two distinct agents i and j with j being a swing voter, knowing i's preference would not reveal j's preference. For example, in US electoral voting, knowing Georgia voting for Republican does not reveal whether Ohio will vote for Democrat or Republican, i.e., uncertainty in swing voters holds.

Our second impossible result is that, given uncertainty in swing voters, a SCF can be implemented in Nash equilibria by a stochastic voting mechanism if and only if it is dictatorial.

Third, we consider the solution concept of rationalizability and any general stochastic mechanisms (which may not be voting mechanisms). Our third impossible result is that, a SCF can be implemented in rationalizability by a general stochastic mechanism if and only if it is dictatorial.

Finally, we fully characterize Nash implementation by a stochastic voting mechanism, when uncertainty in swing voters is violated. This result also fully characterizes when the option of a second referendum may help us achieve a non-dictatorial social goal.

The remainder of the paper proceeds as follows: we describe the model in Section 2; we study Nash implementation in voting mechanisms and stochastic voting mechanisms in Sections 3 and 4, respectively; we study rationalizable implementation in Section 5; we fully characterize Nash implementation by a stochastic voting mechanism in Section 6; we discuss extensions and related literature in Section 7; we provide some final thoughts in Section 8. Some technical proofs are relegated to the appendix.

2 Model

Let I denote a *finite* set of agents, and $\mathcal{A}$ a set of two social outcomes L and R, i.e., $\mathcal{A} = \{L, R\}$ with $L \neq R$. We use α and β to denote unidentified elements in $\mathcal{A}$.

Every agent $i \in I$ is endowed with a strict preference (denoted by $\succ_i$), i.e., she strictly prefers either L to R, or R to L. Let $\mathcal{P}_i^*$ denote the set of all possible strict preference of agent i on $\mathcal{A}$,[8] and $\mathcal{P}^* \equiv \times_{i \in I} \mathcal{P}_i^*$.

An implementation problem, denoted by $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, is fully defined by some $\mathcal{P} \subset \mathcal{P}^*$ and a function f which maps $\mathcal{P}$ to $\mathcal{A}$. The interpretation is that $\mathcal{P}$ is the set of all *possible* profiles of agents' preference, and f is the targeted social choice function. That is, when $P \in \mathcal{P}$ occurs, we aim to select the social outcome $f(P) \in \mathcal{A}$. We say $[\mathcal{P},\ f]$ is a non-trivial implementation problem, if and only if $f(\mathcal{P}) = \mathcal{A}$.[9]

The following notion generalize all of the existing voting rules.

Definition 1 *A voting mechanism is a tuple* $\langle S \equiv (S_i)_{i \in I},\ g : S \longrightarrow \mathcal{A} \rangle$ *such that* $|S| < \infty$, $\mathcal{A} \subset S_i$ *for every* $i \in I$ *and*

$$\forall (i, s, \alpha) \in I \times S \times \mathcal{A},$$
$$g(s) = \alpha \implies g(s_i' = \alpha,\ s_{-i}) = \alpha. \tag{1}$$

In a voting mechanism, $\mathcal{A} \subset S_i$ implies that every agent can vote for L and R, and if $S_i \diagdown \mathcal{A} \neq \varnothing$, agent i is allowed to choose a different strategy, e.g., abstain. Condition (1) is called voting monotonicity[10], and it says that if $g(s) = \alpha$, then any agent switching her vote to α will make the voting result unchanged, i.e., a new supporter can do no harm, as described in Moulin (1988).[11]

For any $P \in \mathcal{P}$ and any voting mechanism $G = \langle S \equiv (S_i)_{i \in I},\ g : S \longrightarrow \mathcal{A} \rangle$, define

$$NE\left[G,\ P = (\succ_i)_{i \in I}\right] := \left\{ s \in S : \forall i \in I, \nexists s_i' \in S_i \text{ such that } g(s_i', s_{-i}) \succ_i g(s_i, s_{-i}) \right\}.$$

I.e., $NE(G, P)$ is the set of all pure-strategy Nash equilibrium[12] (hereafter, NE) in G.

[8] Clearly, $\mathcal{P}_i^* = \{\succ_i', \succ_i''\}$ for every $i \in I$, with $L \succ_i' R$ and $R \succ_i'' L$.

[9] If $f(\mathcal{P}) = \{L\}$ (resp., $f(\mathcal{P}) = \{R\}$), we should choose L (resp., R) under any possible preference profile. As a result, we can implement f trivially.

[10] Voting monotonicity is very similar to Maskin's monotonicity (Maskin (1999)). However, Maskin's monotonicity is defined on a social choice function and the true preference, while voting monotonicity is defined on a game and reported preference.

[11] Moulin (1988) assumes $S_i = \mathcal{A}$ when it defines voting monotonicity.

[12] We discuss mixed-strategy Nash equilibria in Section 7.1.

Definition 2 *Given an implementation problem* $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, *we say* f *can be voting-implemented in NE if there exists a voting mechanism G such that*

$$g\left[NE\left(G, P\right)\right] = \left\{f\left(P\right)\right\}, \forall P \in \mathcal{P}.$$

Definition 3 *Given an implementation problem* $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, *we say* f *is dictatorial if there exists* $i \in I$ *such that for any* $P = \left(\succ_j\right)_{j \in I} \in \mathcal{P}$,

$$L \succ_i R \Leftrightarrow f\left(P\right) = L, \text{ or equivalently, } R \succ_i L \Leftrightarrow f\left(P\right) = R.$$

In particular, we say agent i *is a dictator.*

3 Nash implementation in voting mechanisms

We now present our first impossibility result.

Theorem 1 *In any non-trivial implementation problem, the following two statements are equivalent:*

[1] f *is dictatorial;*

[2] f *can be voting-implemented in NE.*

Suppose f is dictatorial with agent i being a dictator. Define the i-dictatorial mechanism as the voting mechanism in which we invite i *only* to vote on L and R; if i votes for L, we select L, and if i votes for R, we select R. Clearly, the i-dictatorial mechanism implements f in NE, i.e., [1] $\Longrightarrow$ [2].

Proof of [2] $\Longrightarrow$ [1]: Fix any non-trivial implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, and suppose a voting mechanism $G = \langle S,\ g : S \longrightarrow \mathcal{A}\rangle$ implements f. Then, there exist $\left\{P' = \left(\succ'_i\right)_{i \in I}, P'' = \left(\succ''_i\right)_{i \in I}\right\} \subset \mathcal{P}$ and $\{s', s''\} \subset S$ such that

$$f\left(P'\right) = g\left(s'\right) = L, \text{ and } s' \in NE\left(G, P'\right); \tag{2}$$

$$f\left(P''\right) = g\left(s''\right) = R, \text{ and } s'' \in NE\left(G, P''\right). \tag{3}$$

By voting monotonicity, (2) and (3) imply

$$g\left[s^* \equiv (s_i^* = L)_{i\in I}\right] = L;$$
$$g\left[s^{**} \equiv (s_i^{**} = R)_{i\in I}\right] = R.$$

Since G implements f in NE, $f(P'') = R$ and $g[s^*] = L$ imply that s^* cannot be a NE under P''. Hence, there exists an agent $j \in I$ with $R \succ_j'' L$ who wants to deviate from s^*, and her deviation would indeed change the voting result to R, i.e.,

$$\exists \widehat{s}_j \in S_j, \text{ such that } g\left[\widehat{s}_j,\ (s_i^* = L)_{i\in I\diagdown\{j\}}\right] = R,$$

which, together with voting monotonicity, further implies

$$g\left[s_j^{**} = R,\ (s_i^* = L)_{i\in I\diagdown\{j\}}\right] = R. \tag{4}$$

Applying the same argument to s^{**} and P', we get some agent $h \in I$ such that

$$g\left[s_h^{*} = L,\ (s_i^{**} = R)_{i\in I\diagdown\{h\}}\right] = L. \tag{5}$$

We now show $j = h$. Suppose $j \neq h$. Without loss of generality, suppose $j = 1$ and $h = 2$. Then, (4) and (5) become

$$g[R, L, L, ..., L] = R; \tag{6}$$
$$g[R, L, R, ..., R] = L; \tag{7}$$

Then, (6) and voting monotonicity imply

$$g[R, L, R, ..., R] = R,$$

which contradicts (7). Therefore, $j = h$.

Finally, we show agent $j\,(= h)$ is a dictator. By voting monotonicity, (4) and (5) imply[13]

$$g\left[s_j^{**} = R,\ s_{-j}\right] = R, \forall s_{-j} \in S_{-j};$$
$$g\left[s_j^{*} = L,\ s_{-j}\right] = L, \forall s_{-j} \in S_{-j}.$$

[13]For example, if $g\left[s_j^{**} = R,\ s_{-j}\right] = L$ for some s_{-j}, then voting monotonicity implies $g\left[s_j^{**} = R,\ (s_i^* = L)_{i\in I\diagdown\{j\}}\right] = L$, contradicting (4).

As a result, for any $P = (\succ_i)_{i \in I} \in \mathcal{P}$ with $L \succ_j R$, a strategy profile is a NE if and only if agent j plays $s_j^* = L$ under that strategy profile. Since G implements f, we conclude $f[P] = L$. Similarly, for any $P = (\succ_i)_{i \in I} \in \mathcal{P}$ with $R \succ_j L$, we have $f[P] = R$. Therefore, j is a dictator.■

3.1 Interpretation of Theorem 1

To some extent, Theorem 1 is not too surprising, because a similar but simpler argument (i.e., without resorting to voting monotonicity) has been known for the majority voting rule. Theorem 1 can be thought as a rigorous formalization of this argument, and it generalizes the argument in two dimensions. First, Theorem 1 applies to any voting rule, rather than the majority rule only. Second, for the general voting mechanism, we allow for any arbitrary S_i such that $\mathcal{A} \subset S_i$, i.e., we allow for any other strategies, besides L, R and abstain.

How should we interpret Theorem 1? Two elements are critical to get this impossibility, i.e., the solution concept of NE and the voting mechanism. The usual interpretation takes the voting mechanism for granted, and consider NE as being too weak, i.e., not a good solution concept in this setup. From a normative view, the stronger the solution concept (e.g., weakly dominant strategy), the better, because it not only provides a sharper prediction[14], but also imposes fewer requirement on agents (e.g., their beliefs on their opponents' strategies) when they play the stronger solution. For instance, it is straightforward to show that truthfully voting is a weakly dominant strategy in any voting mechanism. As an immediate consequence, the majority SCF[15] can be implemented in dominant strategy by the majority voting rule.

However, from a positive view, the interpretation above is problematic. Even though playing a weakly dominant strategy imposes few requirements on players, not playing other best replies actually imposes much more requirements — partial implementation considers the former only, but full implementation (as adopted here) has to consider

[14] For full implementation as considered in this paper, the stronger the solution, the easier to implement.

[15] The majority SCF is the one in which we select the social outcome that is preferred by the majority.

both.[16] More specifically, for the weak solution concept of undominated strategies, agents can play any best reply, i.e., we assume they are rational; for a stronger solution of rationalizability, we have to assume the stronger assumption of common knowledge of rationality; for a even stronger solution of NE, we have to, in addition, assume that players have the correct conjecture about each other's strategies. Therefore, if we require agents play a dominant strategy, and meanwhile, not play any other best replies, we have to impose much stronger assumptions. From a positive view, it is easy to show that adopting the solution of dominant strategy is not a good description of what happens in reality. For example, the turnout rates of the 2015 Greece bailout referendum and the 2016 Colombia peace agreement referendum are 62.5% and 37.44%, respectively. As discussed above, abstain is a weakly dominated strategy, but a large proportion of voters still play it. Clearly, playing a dominant strategy is not a good description of their behavior. Rather, a better explanation is that those voters do not expect to be pivotal voters, and choose to abstain, i.e., they play a Nash equilibrium, or rationalizable actions.

In light of this positive view, we propose an alternative interpretation of Theorem 1: the voting mechanism, which requires (I) $g(S) = \mathcal{A}$ and (II) voting monotonicity, is the source that brings the impossibility. It is straightforward to see that both (I) and (II) are critical in the proof of Theorem 1. Furthermore, we can construct examples in which the impossibility fails when either (I) or (II) is violated. Therefore, in the sections that follow, we relax one or both of (I) and (II), and provide more leeway for the designer to choose mechanisms. We then study what SCFs can be implemented in such more general mechanisms.

4 Nash implementation in stochastic voting mechanisms

To provide more leeway, we allow the designer to use stochastic mechanisms defined as follows.

[16] This is why, for partial implementation, the stronger solution, the better, while for full implementation, the weaker, the better.

Definition 4 *A stochastic mechanism is a tuple,* $\left\langle S \equiv (S_i)_{i \in I}, \; g : S \longrightarrow \mathcal{O} \right\rangle$*, such that*

$$|S| < \infty, \mathcal{A} \subset \mathcal{O}, \text{ and}$$
$$\alpha \succ_i \beta \Longrightarrow \alpha \succ_i o \succ_i \beta, \forall (i, \alpha, \beta, o) \in I \times \mathcal{A} \times \mathcal{A} \times [\mathcal{O} \diagdown \mathcal{A}]. \tag{8}$$

In the definition above, we assume that every agent i's preference on $\mathcal{A}$ can be extended to a preference on $\mathcal{O}$. With abuse of notation, we continue to use $\succ_i$ to denote the extended preference on $\mathcal{O}$.

If $\mathcal{O} \subset \triangle(\mathcal{A})$, it is a stochastic mechanism with *objective* lotteries on social outcomes. Furthermore, it also accommodates *subjective* lotteries. For example, we may define a mechanism $\langle S, \; g : S \longrightarrow \mathcal{O} \rangle$ for the BREXIT voting, with

$$\mathcal{A} = \{L = \text{"exit"}, R = \text{"stay"}\};$$
$$\mathcal{O} = \{L, R, o = \text{"a second referendum"}\}.$$

Clearly, o is an outcome that involves uncertainty, and every agent has her own subjective belief regarding the chance of "stay" and "exit," if o is chosen. Nevertheless, an agent (e.g., a expected utility maximizer) always strictly prefers her top choice in $\mathcal{A}$ with certainty to o, and strictly prefers o to the worst choice in $\mathcal{A}$ with certainty, i.e., (8) is satisfied.[17]

Like above, define Nash equilibria of a stochastic mechanism G as:

$$NE\left[G, \; P = (\succ_i)_{i \in I}\right] := \left\{s \in S : \forall i \in I, \nexists s_i' \in S_i \text{ such that } g\left(s_i', s_{-i}\right) \succ_i g\left(s_i, s_{-i}\right)\right\}.$$

Similarly, a stochastic voting mechanism is $\left\langle S \equiv (S_i)_{i \in I}, \; g : S \longrightarrow \mathcal{O} \right\rangle$ such that $A \subset S_i$ for every $i \in I$, and voting monotonicity is satisfied, i.e.,

$$\forall (i, s, \alpha) \in I \times S \times \mathcal{A},$$
$$g(s) = \alpha \implies g\left(s_i' = \alpha, \; s_{-i}\right) = \alpha.$$

Definition 5 *Given an implementation problem* $[\mathcal{P}, \; f : \mathcal{P} \longrightarrow \mathcal{A}]$*, we say* f *can be stochastically-voting-implemented in NE if there exists a stochastic voting mechanism* G *such that*

$$g\left[NE(G, P)\right] = \{f(P)\}, \forall P \in \mathcal{P}.$$

[17]We make a minimal assumption (8), and allows for general preference (e.g., risk aversion and uncertainty aversion), as long as (8) is satisfied.

We now define an assumption on preference before we present our second impossibility result. Any $\mathcal{P}$ partitions agents into three groups.

$$I^{(\mathcal{P},\{L\})} := \left\{ i \in I : L \succ_i R, \forall P = \left(\succ_j\right)_{j \in I} \in \mathcal{P} \right\};$$
$$I^{(\mathcal{P},\{R\})} := \left\{ i \in I : R \succ_i L, \forall P = \left(\succ_j\right)_{j \in I} \in \mathcal{P} \right\};$$
$$I^{(\mathcal{P},\mathcal{A})} := I \diagdown \left[I^{(\mathcal{P},\{L\})} \bigcup I^{(\mathcal{P},\{R\})} \right].$$

The set $I^{(\mathcal{P},\{L\})}$ contains all the agents who always prefer L to R under any possible preference profiles—call these agents *L-voters*. Similarly, the set $I^{(\mathcal{P},\{R\})}$ contains all the *R-voters*. The set $I^{(\mathcal{P},\mathcal{A})}$ contains all the *swing voters*, i.e., every agent in $I^{(\mathcal{P},\mathcal{A})}$ prefers L to R under some preference profile, and prefers R to L under some other. For example, in US presidential electoral voting with $L =$ [Democratic party] and $R =$ [Republican party], $I^{(\mathcal{P},\{L\})}$ is the set of Democratic states, e.g., California, and $I^{(\mathcal{P},\{R\})}$ is the set of Republican states, e.g., Texas, and $I^{(\mathcal{P},\mathcal{A})}$ contains all the swing states, e.g., Ohio. Consider the following assumption, which is likely to be satisfied in reality.

Assumption 1 (uncertainty in swing voters) *$\mathcal{P}$ $(\subset \mathcal{P}^*)$ satisfies "uncertainty in swing voters," if for every $(i, j) \in I \times I^{(\mathcal{P},\mathcal{A})}$, and any $\alpha, \beta \in \mathcal{A}$,*

$$[\exists P = (\succ_h)_{h \in I} \in \mathcal{P} \text{ such that } \alpha \succ_i \beta \text{ and } \alpha \succ_j \beta]$$
$$\Longleftrightarrow$$
$$[\exists P' = \left(\succ'_h\right)_{h \in I} \in \mathcal{P} \text{ such that } \alpha \succ'_i \beta \text{ and } \beta \succ'_j \alpha].$$

Uncertainty in swing voters means that knowing any voter's preference does not reveal another swing voter's preference. In US presidential electoral voting, it means that knowing Georgia voting for Democrat does not help us predict the result of Ohio.

We now present our second impossibility result.

Theorem 2 *In any non-trivial implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$ with $\mathcal{P}$ satisfying "uncertainty in swing voters," the following two statements are equivalent:*

[1] f is dictatorial;

[3] f can be stochastically-voting-implemented in NE.

Suppose f is dictatorial with agent i being a dictator. Then, the i-dictatorial mechanism defined in Section 3 is a voting mechanism that implements f, i.e., [1] $\Longrightarrow$ [3].

To prove [3] $\Longrightarrow$ [1], we need the following notion.

Definition 6 *For any implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$ and any $\alpha, \beta \in \mathcal{A}$ with $\alpha \neq \beta$, we say agent $i \in I$ is an α-dictator if there exists some $P = (\succ_j)_{j \in I} \in \mathcal{P}$ such that $\alpha \succ_i \beta$, and for every $P' = \left(\succ'_j\right)_{j \in I} \in \mathcal{P}$, we have*

$$\alpha \succ'_i \beta \Longrightarrow f(P) = \alpha.$$

We say agent i is a semi-dictator if she is either a L-dictator or a R-dictator. We say f is semi-dictatorial if there exist $i, j \in I$ such that i is a L-dictator and j is a R-dictator.

A dictatorial SCF is semi-dictatorial, with the dictator being both a L-dictator and a R-dictator. However, the converse may not be true, because the L-dictator and the R-dictator may be distinct for a semi-dictatorial SCF. We now provide a property of semi-dictators.

Lemma 1 *In any non-trivial implementation problem, if agent i is a semi-dictator, then i is a swing voter.*

Proof of Lemma 1: If agent i is a L-voter, there exists no preference profile in $\mathcal{P}$ under which i prefers R to L, so i is not a R-dictator; furthermore, f maps some preference profile in $\mathcal{P}$ to R in a non-trivial implementation problem, so i is not a L-dictator either. Similarly, if agent i is a R-voter, i is not a semi-dictator. Therefore, if agent i is a semi-dictator, then i is a swing voter.■

The following lemma provides a necessary condition for implementation in NE by a stochastic voting mechanism, and it is implied by Lemma 5, whose proof can be found in Appendix A.3.

Lemma 2 *In any non-trivial implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, f is semi-dictatorial if f can be stochastically-voting-implemented in NE.*

Proof of [3] $\Longrightarrow$ [1]: By Lemma 2, [3] implies that f is semi-dictatorial, i.e., there exist $i, j \in I$ such that i is a L-dictator and j is a R-dictator. By Lemma 1, both i and j are swing voters. We now prove $i = j$ by contradiction, and as a result, f is dictatorial. Suppose $i \neq j$. Then, by "uncertainty in swing voters," there exists $P = (\succ_h)_{h \in I} \in \mathcal{P}$, such that $L \succ_i R$ and $R \succ_j L$. If $f(P) = L$, it contradicts j being a R-dictator, and if $f(P) = R$, it contradicts i being an L-dictator.$\blacksquare$

5 Rationalizable implementation

In this section, we continue to provide more leeway for the designer, and consider any general stochastic mechanism $\langle S \equiv (S_i)_{i \in I}, g : S \longrightarrow \mathcal{O} \rangle$ as defined above, which may not be a voting mechanism, i.e., we allow for arbitrary finite S_i, and voting monotonicity may not be satisfied. Furthermore, we consider two different solution concepts: undominated strategies and rationalizable strategies, which are rigorously defined below.

We allow for mixed strategies, which induces random outcomes in $\triangle(\mathcal{O})$. We further assume that every agent i's preference on $\mathcal{O}$ can be extended to $\triangle(\mathcal{O})$, which is denoted by $\succ_i^{\circ}$. With abuse of notation, we continue to use L (resp. R) to denote the Dirac measure[18] on L (resp. R), and use $\mathcal{A}$ to denote the set of two Dirac measures on L and R. Throughout the paper, we assume

$$\alpha \succ_i \beta \Longrightarrow \alpha \succ_i^{\circ} \epsilon \succ_i^{\circ} \beta, \ \forall (i, \alpha, \beta, \epsilon) \in I \times \mathcal{A} \times \mathcal{A} \times [\triangle(\mathcal{O}) \diagdown \mathcal{A}]. \tag{9}$$

(9) says that every agent prefers her top choice in $\mathcal{A}$ to any non-degenerated mixture of the two social outcomes, and the latter is also strictly preferred to her worst choice in $\mathcal{A}$. Clearly, the expected utility satisfies (9).

With slight abuse of notation, we continue to use g to denote the extended game on mixed strategies, i.e.,

$$\begin{aligned} g &: \triangle(S) \longrightarrow \triangle(\mathcal{O}), \\ g[\rho](E) &= \sum_{s \in \{s' \in S:\, g(s') \in E\}} \rho(s), \forall \rho \in \triangle(S), \forall E \subset \mathcal{O}. \end{aligned}$$

[18] I.e., the Dirac measure on L (resp. R) assigns probability 1 on L (resp. R).

We use s to denote the Dirac measure on s. Furthermore, for any $s_i \in S_i$ and any $\rho_{-i} \in \Delta(S_{-i})$, we use $(s_i, \rho_{-i}) \in \triangle(\times_{j \in I} S_j)$ to denote the distribution whose marginal distributions on S_i and S_{-i} are s_i and ρ_{-i}, respectively.

Given a preference profile P and a stochastic mechanism G, for every $i \in I$, define

$$R_i^0(G,P) \equiv S_i; \text{ and for every positive integer } k,$$

$$R_i^k(G,P) \equiv \left\{ s_i \in R_i^{k-1}(G,P) : \begin{array}{c} \nexists s_i' \in S_i \text{ such that } g(s_i', \rho_{-i}) \succ_i^{\circ} g(s_i, \rho_{-i}), \\ \forall \rho_{-i} \in \Delta\left(\times_{j \neq i} R_j^{k-1}(G,P)\right) \end{array} \right\};$$

$$R_i^{\infty}(G,P) \equiv \cap_{k=1}^{\infty} R_i^k(G,P).$$

That is, $R_i^k(G,P)$ is the set of strategies of agent i in G that survives k rounds of iterative deletion of strictly dominated strategies, and $R_i^{\infty}(G,P)$ is the set of rationalizable strategies. Define

$$R^1(G,P) \equiv \times_{i \in I} R_i^1(G,P) \text{ and } R^{\infty}(G,P) \equiv \times_{i \in I} R_i^{\infty}(G,P).$$

Definition 7 *Given an implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, we say f can be implemented in undominated strategies if there exists a stochastic mechanism G such that*

$$g\left[R^1(G,P)\right] = \{f(P)\}, \forall P \in \mathcal{P}.$$

Definition 8 *Given an implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$, we say f can be implemented in rationalizable strategies if there exists a stochastic mechanism G such that*

$$g\left[R^{\infty}(G,P)\right] = \{f(P)\}, \forall P \in \mathcal{P}.$$

We now present our third impossibility result.

Theorem 3 *In any non-trivial implementation problem, the following three statements are equivalent:*

[1] f is dictatorial;

[4] f can be implemented in undominated strategies;

[5] f can be implemented in rationalizable strategies.

Suppose f is dictatorial with agent i being a dictator. Then, the i-dictatorial mechanism defined in Section 3 implements f, i.e., [1] $\Longrightarrow$ [4] and [1] $\Longrightarrow$ [5]. Furthermore, it is straightforward to show [4] $\Longrightarrow$ [5].

To prove [5] $\Longrightarrow$ [1], we need the following two lemmas, and their proofs can be found in Appendix A.1 and A.2.

Lemma 3 *For any $P = (\succ_i)_{i \in I} \in \mathcal{P}^*$, any $\alpha, \beta \in \mathcal{A}$ with $\alpha \neq \beta$, and any stochastic mechanism $G = \langle S,\ g \rangle$, if $g\left[R^\infty(G, P)\right] = \{\beta\}$, then*

$$\alpha \succ_j \beta \Longrightarrow R_j^\infty(G, P) = S_j, \forall j \in I. \tag{10}$$

Lemma 3 implies that if we can implement f in rationalizable strategies under P and $f(P)$ is the worst outcome in $\mathcal{A}$ for agent j, then every strategy is rationalizable for j. The intuition is straightforward: if the *worst* outcome $f(P)$ is induced by a *best* reply give a rationalizable conjecture, then every strategy must be a best reply given that conjecture, and hence also rationalizable.

Given any $G = \left\langle S \equiv (S_i)_{i \in I},\ g : S \longrightarrow \mathcal{O} \right\rangle$ and any $(i, \alpha) \in I \times \mathcal{A}$, define

$$S^{(G,\alpha)} = \{s \in S :\ g(s) = \alpha\},$$
$$S_i^{(G,\alpha)} = \{s_i \in S_i : \exists s_{-i} \in S_{-i} \text{ such that } g(s_i, s_{-i}) = \alpha\},$$

i.e., $S^{(G,\alpha)}$ is the set of strategy profiles in G that result in α, and $S_i^{(G,\alpha)}$ is the set of strategies of agent i that cannot exclude the possibility of α.

Lemma 4 *For any $P = (\succ_i)_{i \in I} \in \mathcal{P}^*$, any $\alpha \in \mathcal{A}$, and any stochastic mechanism $G = \langle S,\ g \rangle$ such that $g\left[R^\infty(G, P)\right] = \{\alpha\}$, we have*

$$R^\infty(G, P) = S^{(G,\,\alpha)} = \times_{i \in I} S_i^{(G,\,\alpha)}, \tag{11}$$

and in particular,

$$R_i^\infty(G, P) = S_i^{(G,\,\alpha)}, \forall i \in I. \tag{12}$$

Lemma 4 implies that if G implements f under P, then $S^{(G,\, f(P))}$ and $S_i^{(G,\, f(P))}$ fully describe the sets of rationalizable strategy profiles and i's rationalizable strategies, respectively.

Proof of [5] $\Longrightarrow$ [1]: Fix any non-trivial implementation problem $[\mathcal{P},\ f : \mathcal{P} \longrightarrow \mathcal{A}]$. Given [5], there exists a stochastic mechanism G such that

$$g\left[R^{\infty}\left(G, P\right)\right] = \left\{f\left(P\right)\right\}, \forall P \in \mathcal{P}.$$

Since $f\left(\mathcal{P}\right) = \mathcal{A}$, there exist $\widehat{P}, \widetilde{P} \in \mathcal{P}$ such that $g\left[R^{\infty}\left(G, \widehat{P}\right)\right] = \{L\}$ and $g\left[R^{\infty}\left(G, \widetilde{P}\right)\right] = \{R\}$. Then, by Lemma 4 (more precisely, (11)), we have

$$S^{(G,\, L)} = \times_{i \in I} S_i^{(G,\, L)} \text{ and } S^{(G,\, R)} = \times_{i \in I} S_i^{(G,\, R)}.$$

We now prove [1] by contradiction. Suppose f is not dictatorial. As a result, for every $j \in I$, j is not a dictator, i.e., either there exists $P' = \left(\succ_i'\right)_{i \in I} \in \mathcal{P}$ such that $f\left(P'\right) = L$ and $R \succ_j' L$, or there exists $P'' = \left(\succ_i''\right)_{i \in I} \in \mathcal{P}$ such that $f\left(P''\right) = R$ and $L \succ_j'' R$. By (10) and (12), the former case implies $S_j = R_j^{\infty}\left(G, P'\right) = S_j^{(G,\, L)}$ and the latter case implies $S_j = R_j^{\infty}\left(G, P''\right) = S_j^{(G,\, R)}$. Each of the two case implies

$$S_j^{(G,\, L)} \cap S_j^{(G,\, R)} \neq \varnothing, \forall j \in I.$$

As a result,

$$\times_{j \in I}\left[S_j^{(G,\, L)} \cap S_j^{(G,\, R)}\right] \neq \varnothing.$$

Pick any $s \in \times_{j \in I}\left[S_j^{(G,\, L)} \cap S_j^{(G,\, R)}\right]$, and by (11), we have

$$s \in \times_{j \in I} S_j^{(G,\, L)} = S^{(G,\, L)} \text{ and } s \in \times_{j \in I}\left[S_j^{(G,\, R)}\right] = S^{(G,\, R)},$$

which further implies

$$g\left(s\right) = L \neq R = g\left(s\right),$$

i.e., we reach a contradiction.■

6 Full characterization of NE implementation

Finally, we fully characterize when a non-dictatorial SCF can be implemented by a stochastic voting mechanism, if uncertainty in swing voters does not hold. As discussed above, being dictatorial is a sufficient condition for NE implementation, and being semi-dictaorial is a necessary condition. Hence, a necessary and sufficient condition must be weaker than the former and stronger than the latter. The following is such a notion.

Definition 9 *For any implementation problem $[\mathcal{P},\ f:\mathcal{P}\longrightarrow\mathcal{A}]$, we say f is pseudo-dictatorial if there exist two subsets of swing voters, $I^L\subset I$ and $I^R\subset I$, such that for every $P=(\succ_h)_{h\in I}\in\mathcal{P}$,*

$$f(P)=L \text{ if and only if } L\succ_i R \text{ for some } i\in I^L, \tag{13}$$
$$\text{and } f(P)=R \text{ if and only if } R\succ_j L \text{ for some } j\in I^R.$$

Several points are worthy of noting regarding Definition 9. First, a dictatorial SCF is pseudo-dictatorial, and a pseudo-dictatorial SCF is semi-dictatorial. Second, every agent in I^L is an L-dictator and every agent in I^R is a R-dictator. Third, an equivalent statement of (13) is:

$$f(P)=R \text{ if and only if } R\succ_i L \text{ for all } i\in I^L,$$
$$\text{and } f(P)=L \text{ if and only if } L\succ_j R \text{ for all } j\in I^R.$$

The following result fully characterizes all SCFs that can be implemented in NE by a stochastic voting mechanism.

Theorem 4 *In any non-trivial implementation problem $[\mathcal{P},\ f:\mathcal{P}\longrightarrow\mathcal{A}]$, the following two statements are equivalent:*

[3] f can be stochastically-voting-implemented in NE;

[6] f is pseudo-dictatorial.

[3] $\Longrightarrow$ [6] is implied by the following lemma, and its proof can be found in Appendix A.3.

Lemma 5 *In any non-trivial implementation problem $[\mathcal{P},\ f:\mathcal{P}\longrightarrow\mathcal{A}]$, f is pseudo-dictatorial if f can be stochastically-voting-implemented in NE.*

To prove [6] $\Longrightarrow$ [3], we define a stochastic voting mechanism which involves the option of a second referendum.

For every $\gamma \in (0,1)$, define the voting mechanism $G^{\gamma} \equiv \left\langle S = (S_i = \mathcal{A})_{i \in I}, \, g^{\gamma} : S \longrightarrow \mathcal{A} \right\rangle$ as

$$g^{\gamma}(s) = \begin{cases} L, & \text{if } \frac{|\{i \in I : s_i = L\}|}{|I|} \geq \gamma, \\ R, & \text{otherwise.} \end{cases}$$

That is, under G^{γ}, we select L if and only if L gets at least a proportion of γ of the total votes. Define

$$o^{\gamma} \equiv \text{"a second referendum using the voting rule } G^{\gamma}\text{,"}$$
$$\text{and } \mathcal{O}^* = \{o^{\gamma} : \gamma \in (0,1)\} \cup \mathcal{A}.$$

We assume that for any $\gamma', \gamma'' \in (0,1)$ with $\gamma' < \gamma''$,

$$L \succ_i R \Longrightarrow L \succ_i o^{\gamma'} \succ_i o^{\gamma''} \succ_i R;$$
$$R \succ_i L \Longrightarrow R \succ_i o^{\gamma''} \succ_i o^{\gamma'} \succ_i L.$$

I.e., if L is agent i's top choice, i prefers a second referendum with a low threshold (for L) to a second referendum with a high threshold. Similarly, if R is agent i's top choice, i prefers a second referendum with a high threshold to a second referendum with a low threshold.

Fix any non-trivial implementation problem $[\mathcal{P}, \, f : \mathcal{P} \longrightarrow \mathcal{A}]$. Suppose f is pseudo-dictatorial with two subsets of swing voters, $I^L \subset I$ and $I^R \subset I$, such that (13) holds for every $P = (\succ_i)_{i \in I} \in \mathcal{P}$. We now define the stochastic voting mechanism $G^* \equiv \left\langle S = (S_i = \mathcal{A})_{i \in I}, \, g^* : S \longrightarrow \mathcal{O}^* \right\rangle$ as follows, which implements f.

$$g^*(s) = \begin{cases} L, & \text{if } s_j = L \text{ for every } j \in I^R \text{ and } s_i = L \text{ for some } i \in I^L, \\ R, & \text{if } s_i = R \text{ for every } i \in I^L \text{ and } s_j = R \text{ for some } j \in I^R, \\ o^{\frac{|\{i \in I : s_i = R\}|}{|I|}}, & \text{otherwise.} \end{cases}$$

The voting outcome is determined as follows: if all voters in I^R and at least one voter in I^L vote for L, then g^* selects L; if all voters in I^L and at least one voter in I^R vote for R, then g^* selects R; otherwise, a second referendum will be held with the threshold for L being $\frac{|\{i \in I: s_i = R\}|}{|I|}$.

We now show f is implemented by G^*. For any $P = (\succ_i)_{i \in I} \in \mathcal{P}$ such that $f(P) = L$, (13) implies

$$L \succ_j R \text{ all } j \in I^R \text{ and } L \succ_k R \text{ for some } k \in I^L.$$

Clearly, any strategy profile with agent k and all agents in I^R voting for L is a NE.

We now show that any s is not a NE if $g^*(s) \neq L$. First, consider $g^*(s) = R$, i.e., all agents in I^L and at least one agent in I^R vote for R. Then, agent $k \in I^L$ with $L \succ_k R$ would want to deviate to vote for L, which would induce an outcome other than R, and agent k gets strictly better off, i.e., s is not a NE. Second, consider $g^*(s) \notin \mathcal{A}$, i.e., s induces a second referendum. Since $g^*(s) \neq L$, not everyone in $I^R \cup \{k\}$ votes for L. Let h be one of the agents in $I^R \cup \{k\}$ who vote for R. Suppose h deviates to vote for L. Then, the voting outcome cannot be converted to R because of voting monotonicity. Hence, there are only two cases left: either the outcome is converted to L, or the outcome remains a second referendum but with a lower threshold for L. In both cases, agent h gets strictly better off, so s is not a NE.

Similar argument shows that for any $P \in \mathcal{P}$ with $f(P) = R$, there is a NE inducing outcome R, and any s is not a NE if $g^*(s) \neq R$.■

7 Discussion

7.1 Other equilibrium solutions

It is not clear to us how to fully characterize NE implementation by a general stochastic mechanism, which is left for future research.

However, both Theorems 1 and 2 can be easily extended to other equilibrium solutions, e.g., mixed-strategy Nash equilibria (hereafter, MSNE) and correlated equilibria. For simplicity, let us focus on MSNE. First, if a SCF is dictatorial, it is straightforward to see that it can be implemented in MSNE.

Second, if a SCF can be implemented in MSNE, then a pure strategy NE always exists. To see this, fix any SCF f and any game $\langle S, g : S \longrightarrow E \rangle$ that implements f in MSNE. For any $P = (\succ_i)_{i \in I} \in \mathcal{P}$ and any MSNE $\sigma = (\sigma_i, \sigma_{-i})$ under P, pick any pure strategy profile (s_i, s_{-i}) in the support of σ. Since f is implemented, we have $g(s_i, s_{-i}) = f(P)$. Let $\alpha\,(\neq f(P))$ denote the other outcome in A. Note that $P = (\succ_i)_{i \in I}$ partitions I into two subsets.

$$I^{\alpha} = \{i \in I : \alpha \succ_i f(P)\} \text{ and } I^{f(P)} = \{i \in I : f(P) \succ_i \alpha\}.$$

For every $i \in I^{f(P)}$, $g(s_i, s_{-i}) = f(P)$ is the best outcome for i, i.e., s_i is a best reply for i, given s_{-i}. For every $i \in I^{\alpha}$, since f is implemented under P, we have $g\left(s_i, s'_{-i}\right) = f(P)$ for every s'_{-i} in the support of σ_{-i}, and $f(P)$ is the worst outcome for i. Nevertheless, s_i is a best reply given σ_{-i} under MSNE σ. Hence, by (9), we conclude,

$$g\left(s'_i, s_{-i}\right) = f(P), \ \forall s'_i \in S_i,$$

i.e., s_i is a best reply for i, given s_{-i}. Therefore, (s_i, s_{-i}) is a pure strategy NE.

Hence, the argument above shows that implementation in MSNE implies implementation in pure-strategy NE, which further implies that f is dictatorial when conditions in Theorems 1 and 2 hold.

7.2 Incomplete information

All of our analysis above are based on the complete-information setup, which implicitly imposes common-prior and common-knowledge assumptions. To see the roles played by these assumptions, we extend our analysis to incomplete-information setups. An incomplete-information setup is a type space

$$\left\langle T = (T_i)_{i \in I}, \, v = (v_i : T_i \rightarrow \mathcal{P}^*_i)_{i \in I}, \, \kappa = [\kappa_i : T_i \rightarrow \triangle(T_{-i})]_{i \in I} \right\rangle.$$

Clearly, this definition takes an interim view, and does not impose the common prior assumption, i.e., the usual common-prior model is a special case. For notational ease, let us focus on finite types, i.e., we require $|T| < \infty$. Given a type space $\langle T, v, \kappa \rangle$, the set of all possible type profiles that may occur is

$$\mathcal{T} = \left\{ (t_i)_{i \in I} : \exists j \in I, \text{such that } \kappa_j \left(t_j \right) \left[t_{-j} \right] > 0 \right\}.$$

$\kappa_j \left(t_j \right) \left[t_{-j} \right] > 0$ implies that $\left(t_j, t_{-j} \right)$ is one possible type profile in the eyes of agent j. Then, the set of all possible preference profiles that may occur is

$$\mathcal{P} = \left\{ \left[v_i \left(t_i \right) \right]_{i \in I} : (t_i)_{i \in I} \in \mathcal{T} \right\}.$$

We thus can define implementation in incomplete-information setups as before.

It is straightforward to extend Theorem 1 to any incomplete-information setups. In fact, the same proof as in Section 3 applies. Therefore, the common-prior and common-knowledge assumptions do not play any role in Theorem 1.

Theorem 2 can be extended to the class of incomplete-information models with 1st-order knowledge of the desired social outcome. More rigorously, it is defined as: for every $i \in I$ and every $t_i \in T_i$

$$\kappa_i \left(t_i \right) \left[t'_{-i} \right] \times \kappa_i \left(t_i \right) \left[t''_{-i} \right] > 0 \Longrightarrow f \left[v_i \left(t_i \right), v_{-i} \left(t'_{-i} \right) \right] = f \left[v_i \left(t_i \right), v_{-i} \left(t''_{-i} \right) \right],$$

i.e., with probability 1, every type t_i believes $f \left[v_i \left(t_i \right), v_{-i} \left(t'_{-i} \right) \right]$ is the desired social outcome. Nevertheless, it allows different agents to know different social outcomes, i.e., we do not impose assumptions on higher-order knowledge. In particular, the proofs of Lemma 2 and Theorem 2 can adapted with little modification.

Finally, it is straightforward to extend Theorem 3 to the class of incomplete-information models with common knowledge of the desired social outcome. I.e., it is common knowledge among the agents that a particular outcome is the desired social outcome, though they may know the preference of each other.

7.3 Related literature

Voting on binary social outcomes has been studied in many papers, e.g., Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996, 1997), Krishna and Morgan (2011, 2012). To the best of our knowledge, we are the first to model it as an implementation problem, and provide impossibility results.

This paper is also closely related to the implementation literature. In various setups, impossibility in implementation has been proved in many papers, e.g., Gibbard (1973), Satterthwaite (1975), Dasgupta, Hammond, and Maskin (1979), Maskin (1979), Jackson (1992), Jackson and Srivastava (1996). Among them, Börgers (1995) is the one which is most close to this paper. Compared to it, our setup is less general in the sense that we focus on a binary social choice, while Börgers (1995) allows for two or more than two social outcomes. However, we study Nash equilibria and stochastic mechanisms besides rationalizability and deterministic mechanisms, while Börgers (1995) restricts attention to the latter two only. In particular, Börgers (1995) raises an open question regarding whether its impossibility result extends to stochastic mechanisms. Theorem 3 provides a partial answer for the question, i.e., the impossibility indeed extends if we focus on two social outcomes.

Assumptions on preference domain (e.g., full domain) are usually needed for previous impossibility results in implementation. For example, Börgers (1995) assumes that "unanimous" preference profiles are always possible. In BREXIT, this means that the two preference profiles, "everyone prefers stay to exit" and "everyone prefers exit to stay," are always possible — clearly, this is not true, because Mr. David Cameron and Mr. Nigel Farage are stay-voter and exit-voter, respectively. Hence, the impossibility result in Börgers (1995) does not apply to BREXIT.

Different from most of previous results, one distinct feature of our impossibility results (Theorems 1 and 3) is that we impose no assumption on preference domain.

8 Some final thoughts

As Sir Winston Churchill once said: *democracy is the worst form of government, except for all those other forms that have been tried from time to time.* One way to interpret it is: democracy may be the best form of government among those that have been tried from time to time, but it is far from being good. This paper provides a rigorous foundation on why democracy cannot help us achieve our social goal in a specific situation, i.e., voting on a binary choice.

Furthermore, our results show that only dictatorial, or more generally, pseudo-dictatorial social choice functions can be implemented in (stochastic) voting mechanisms. Then, why do we vote? Or equivalently, why not let the dictator or the semi-dictators decide on social outcomes? Hence, maybe, we should not vote for social outcomes. Rather, we should elect a ”dictator” or ”semi-dictators” in voting, who will represent all the people to decide on social outcomes. This is not as bad as it looks[19], and we can find real-life analogs, e.g., in US political institutions. A president is elected once every four years, and she or he is the ”dictator” we choose. Congressmen are elected regularly[20], and they are the semi-dictators we choose. This echoes the second definition of democracy in Merrian-Webster Learner's Dictionary:

> *Democracy: a. [noncount]: a form of government in which people choose leaders by voting; b. [cout]: a country ruled by democracy.*

Once a ”dictator” or ”semi-dictators” are elected, other political institutions will enforce them to choose social outcomes for the best interests of most people in the society, e.g., if they perform badly, they will not be re-elected for the next term.

[19] We follow the tradition in the literature, and use ”dictator” and ”semi-dictator,” which are associated with negative meaning, to name those agents who (fully or partially) determine social outcomes. A more appropriate and neutral name should be ”decider” and ”semi-decider.”

[20] United States Senate elections are held once every 6 years, and United States House of Representatives elections are held once every 2 years.

A Proofs

A.1 Proof of Lemma 3

Fix any $(s_j, s_{-j}) \in R^{\infty}(G, P)$. Given s_{-j} played by agents $-j$, let s_j^* denote the best pure strategy for agent j. Note that $s_j^* \in R_j^{\infty}(G, P)$, because $s_{-j} \in R_{-j}^{\infty}(G, P)$. Then, $g[R^{\infty}(G, P)] = \{\beta\}$ implies

$$g\left(s_j^*, s_{-j}\right) = \beta.$$

If β is the worst outcome in $\mathcal{A}$ for agent j under P, (9) implies that β remains the unique worst outcome in $\triangle(\mathcal{O})$ for j. Hence, to make s_j^* a best reply for j given s_{-j}, we must have $g\left(s_j', s_{-j}\right) = \beta$ for every $s_j' \in S_j$. That is, every strategy in S_j is a best reply to s_{-j}, and hence also rationalizable.■

A.2 Proof of Lemma 4

Fix any $P = (\succ_i)_{i \in I} \in \mathcal{P}^*$, any $\alpha \in \mathcal{A}$, and any stochastic mechanism $G = \langle S, g \rangle$ such that $g[R^{\infty}(G, P)] = \{\alpha\}$. We first show $R^{\infty}(G, P) = S^{(G, \alpha)}$. Clearly, $g[R^{\infty}(G, P)] = \{\alpha\}$ implies $R^{\infty}(G, P) \subset S^{(G, \alpha)}$. We now show $S^{(G, \alpha)} \subset R^{\infty}(G, P)$. Fix any $s = (s_j)_{j \in I} \in S^{(G, \alpha)}$, i.e., $g(s) = \alpha$. Let $\beta (\neq \alpha)$ denote the other element in $\mathcal{A}$, i.e., $\mathcal{A} = \{\alpha, \beta\}$. Note that $P = (\succ_i)_{i \in I}$ partitions I into two subsets.

$$I^{\alpha} = \{i \in I : \alpha \succ_i \beta\} \text{ and } I^{\beta} = \{i \in I : \beta \succ_i \alpha\}.$$

Then, Lemma 3 (more precisely, (10)) implies

$$s_i \in R_i^{\infty}(G, P) = S_i, \ \forall i \in I^{\beta}. \tag{14}$$

For every $i \in I^{\alpha}$, (9) implies that α remains a best outcome in $\triangle(\mathcal{O})$ for i. Hence, $g(s_i, s_{-i}) = \alpha$ implies that s_i is a best reply for i, given s_{-i}. As a result, $s_i \in R^k(G, P)$ for every positive integer k, i.e.,

$$s_i \in R_i^{\infty}(G, P), \ \forall i \in I^{\alpha}. \tag{15}$$

(14) and (15) imply $s \in R^{\infty}(G, P)$. Therefore, $S^{(G, \alpha)} \subset R^{\infty}(G, P)$.

Recall $R^{\infty}(G,P) \equiv \times_{i\in I} R_i^{\infty}(G,P)$, which, together with $R^{\infty}(G,P) = S^{(G,\alpha)}$ proved above, implies

$$R_i^{\infty}(G,P) \equiv S_i^{(G,\alpha)}, \forall i \in I.$$

Hence, $R^{\infty}(G,P) = \times_{i\in I} S_i^{(G,\alpha)}$.■

A.3 Proofs of Lemmas 2 and 5

Lemma 5 implies Lemma 2. Hence, it suffices to prove Lemma 5 only.

We first provide two intermediate results, which will be used in the proofs later.

Lemma 6 *For any $P = (\succ_i)_{i\in I} \in \mathcal{P}^*$, any $\alpha, \beta \in \mathcal{A}$, and any stochastic mechanism $G = \langle S,\ g : S \longrightarrow \mathcal{O}\rangle$, if $\alpha \succ_j \beta$ for some $j \in I$, then*

$$\left(s_j, s_{-j}\right) \in NE(G,P) \ \& \ g\left(s_j, s_{-j}\right) = \beta \Longrightarrow g\left(s'_j, s_{-j}\right) = \beta, \forall s, s' \in S.$$

Proof of Lemma 6: Fix any $\left(s_j, s_{-j}\right) \in NE(G,P)$ such that $g\left(s_j, s_{-j}\right) = \beta$. Since β is the uniquely worst outcome in $\mathcal{A}$ for j, by (8), β remains the uniquely worst outcome in $\mathcal{O}$ for j. Thus, given s_{-j}, to make s_j a best reply for j, we must have $g\left(s'_j, s_{-j}\right) = \beta$ for every $s' \in S$.■

Recall

$$I^{(\mathcal{P},\{L\})} \equiv \left\{i \in I : L \succ_i R, \forall P = \left(\succ_j\right)_{j\in I} \in \mathcal{P}\right\};$$
$$I^{(\mathcal{P},\{R\})} \equiv \left\{i \in I : R \succ_i L, \forall P = \left(\succ_j\right)_{j\in I} \in \mathcal{P}\right\};$$
$$I^{(\mathcal{P},\mathcal{A})} \equiv I \diagdown \left[I^{(\mathcal{P},\{L\})} \bigcup I^{(\mathcal{P},\{R\})}\right].$$

i.e., $I^{(\mathcal{P},\{L\})}$, $I^{(\mathcal{P},\{R\})}$ and $I^{(\mathcal{P},\mathcal{A})}$ are the sets of L-voters, R-voters and swing voters, respectively. Furthermore, Lemma 1 says that a semi-dictator must be a swing voter. Hence, we define

$$I^{(\mathcal{P},\mathcal{A},L)} \equiv \left\{i \in I^{(\mathcal{P},\mathcal{A})} : i \text{ is an } L\text{-dictator}\right\};$$
$$I^{(\mathcal{P},\mathcal{A},R)} \equiv \left\{i \in I^{(\mathcal{P},\mathcal{A})} : i \text{ is a } R\text{-dictator}\right\},$$

i.e., $I^{(\mathcal{P},\mathcal{A},L)}$ and $I^{(\mathcal{P},\mathcal{A},R)}$ are the sets of L-dictators and R-dictators, respectively.

Lemma 7 *In any non-trivial implementation problem* $[\mathcal{P},\ f:\mathcal{P}\longrightarrow\mathcal{A}]$*, if* f *is implemented in NE by a stochastic voting mechanism* G*, then for every* $j\in I\diagdown\left(I^{(\mathcal{P},\{L\})}\bigcup I^{(\mathcal{P},\mathcal{A},R)}\right)$*, we have*

$$g\left(s_j,\ (s_i^*=L)_{i\in I\diagdown\{j\}}\right)=L,\forall s_j\in S_j, \tag{16}$$

and for every $j\in I\diagdown\left(I^{(\mathcal{P},\{R\})}\bigcup I^{(\mathcal{P},\mathcal{A},L)}\right)$*, we have*

$$g\left(s_j,\ (s_i^{**}=R)_{i\in I\diagdown\{j\}}\right)=R,\forall s_j\in S_j, \tag{17}$$

Proof of Lemma 7: For any $j\in I\diagdown\left(I^{(\mathcal{P},\{L\})}\bigcup I^{(\mathcal{P},\mathcal{A},R)}\right)$, j is either a R-voter or a swing voter who is not an R-dictator. In the former case, j always prefers R to L, and there exists some preference profile P' such that $f(P')=L$. In the latter case, there exists some preference profile P'' under which j prefers R to L but $f(P'')=L$. To sum, for every $j\in I\diagdown\left(I^{(\mathcal{P},\{L\})}\bigcup I^{(\mathcal{P},\mathcal{A},R)}\right)$, there exists $P=(\succ_i)_{i\in I}\in\mathcal{P}$ such that $f(P)=L$ and $R\succ_j L$. Pick any NE $\widehat{s}$ under P. Since f is implemented, we have $g(\widehat{s})=f(P)=L$. Since $R\succ_j L$, Lemma 6 implies

$$g\left(s_j,\ \widehat{s}_{-j}\right)=L,\forall s_j\in S_j. \tag{18}$$

Then, (18) and voting monotonicity implies (16). A similar argument proves (17)■

We now prove Lemma 5 by contradiction: In any non-trivial implementation problem $[\mathcal{P},\ f:\mathcal{P}\longrightarrow\mathcal{A}]$, suppose f is implemented in NE by a stochastic voting mechanism $G=\langle S,\ g:S\longrightarrow\mathcal{O}\rangle$. Recall that $I^{(\mathcal{P},\mathcal{A},L)}$ and $I^{(\mathcal{P},\mathcal{A},R)}$ are the sets of L-dictators and R-dictators, respectively. We show that for every $P=(\succ_i)_{i\in I}\in\mathcal{P}$,

$$f(P)=L \text{ if and only if } L\succ_j R \text{ for some } j\in I^{(\mathcal{P},\mathcal{A},L)}, \tag{19}$$

$$\text{and } f(P)=R \text{ if and only if } R\succ_j L \text{ for some } j\in I^{(\mathcal{P},\mathcal{A},R)}, \tag{20}$$

i.e., f is pseudo-dictatorial. Fix any $P=(\succ_i)_{i\in I}\in\mathcal{P}$, and we now show (19). The "if" part: if $L\succ_j R$ for some $j\in I^{(\mathcal{P},\mathcal{A},L)}$, we have $f(P)=L$, because j is an L-dictator. The "only if" part: suppose $f(P)=L$. By (17), we have

$$g\left[(s_i^{**}=R)_{i\in I}\right]=R.$$

Then $\left[(s_i^{**}=R)_{i\in I}\right]$ cannot be a NE under P, because $f(P)=L$ and f is implemented in G. Thus, some agent must want to deviate from $\left[(s_i^{**}=R)_{i\in I}\right]$. By Lemma 7 (more precisely, by (17)), the deviator must come from $I^{(\mathcal{P},\,\{R\})} \bigcup I^{(\mathcal{P},\,\mathcal{A},\,L)}$. Since R is the top choice for every agent in $I^{(\mathcal{P},\,\{R\})}$, none of them wants to deviate from $\left[\left(s_i^{**} = R\right)_{i \in I}\right]$. Therefore, the deviator must come from $I^{(\mathcal{P},\,\mathcal{A},\,L)}$, i.e., there exists some $j \in I^{(\mathcal{P},\,\mathcal{A},\,L)}$ such that $L \succ_j R$ (and she wants to deviate from $\left[\left(s_i^{**} = R\right)_{i \in I}\right]$).—This completes the ”only if” part.

A similar argument shows (20).■

References

ARROW, K. (1963): *Social Choice and Individual Values*. Wiley, New York.

AUSTEN-SMITH, D., AND J. S. BANKS (1996): “Information aggregation, rationality, and the Condorcet jury theorem,” *American Political Science Review*, 90, 34–45.

BALINSKI, M., AND R. LARAKI (2016): “Majority Judgement vs. Majority Rule,” *miemo*.

BERNHEIM, B. D. (1984): “Rationalizable Strategic Behavior,” *Econometrica*, 52, 1007–1028.

BÖRGERS, T. (1995): *A note on implementation and strong dominance*, Social Choice, Welfare, and Ethics (edited by William A. Barnett, Hérve Moulin, Maurice Salles and Norman J. Schofield). Cambridge University Press.

CONGRESS, U. S. (2009): *Congressional Record*, vol. 150. United States Congress, Page 11954, ISBN-13: 9780160825644.

DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): “The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility,” *Review of Economic Studies*, 46, 185–216.

FEDDERSEN, T. J., AND W. PESENDORFER (1996): “The swing voters curse,” *American Economic Review*, 86, 408–424.

——— (1997): “Voting behavior and information aggregation in elections with private information,” *Econometrica*, 65, 1029–1058.

GIBBARD, A. (1973): "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41, 587–602.

JACKSON, M. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, 59, 757–775.

JACKSON, M., AND S. SRIVASTAVA (1996): "A Characterization of Game-Theoretic Solutions Which Lead to Impossibility Theorems," *Review of Economic Studies*, 63, 23–38.

KRISHNA, V., AND J. MORGAN (2011): "Overcoming ideological bias in elections," *Journal of Political Economy*, 119, 183–211.

——— (2012): "Voluntary voting: Costs and benefits," *Journal of Economic Theory*, 147, 2083–2123.

LEMONDE (2016): "Nicolas Sarkozy propose des référendums sur le regroupement familial et sur les fichés S," *Le Monde (July 10, 2016)*.

MASKIN, E. (1979): *Implementation and Strong Nash Equilibrium*, Aggregation and Revelation of Preferences (edited by J.-J. Laffont). Amsterdam: North-Holland.

MASKIN, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

MAY, K. O. (1952): "A set of independent necessary and sufficient conditions for simple majority decisions," *Econometrica*, 20, 680–684.

MOULIN, H. (1988): *Axioms of Cooperative Decision Making*. Cambridge University Press.

MULLER, E., AND M. A. SATTERTHWAITE (1977): "The equivalence of strong positive association and strategy-proofness," *Journal of Economic Theory*, 2, 412–418.

NASH, J. (1950): "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Sciences of the United States of America*, 36, 48–49.

——— (1951): "Non-Cooperative Games," *Annals of Mathematics*, 54, 286–295.

PEARCE, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

Satterthwaite, M. A. (1975): "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, 187–217.