Evaluationwise strategy-proofness *

Bora Erdamar[1], M. Remzi Sanver[2], and Shin Sato[3]

[1]University of Rochester, U.S.A.

[2]Université Paris-Dauphine, PSL Research University, CNRS, UMR [7243], LAMSADE, 75016 Paris, France

[2]Murat Sertel Center for Advanced Economic Studies, Istanbul Bilgi University, Turkey

[3]Fukuoka University, Japan

August 27, 2016

Abstract

We consider manipulation of collective decision making rules in a framework where voters not only rank candidates but also evaluate them as "acceptable" or "unacceptable". In this richer informational setting, we adopt a new notion of strategy-proofness, called evaluationwise strategy-proofness, where incentives of manipulation exist if and only if a voter can replace an outcome which he finds unacceptable with an acceptable one. Evaluationwise strategy-proofness is weaker than strategy-proofness. However, we establish the prevalence of a logical incompatibility between evaluationwise strategy-proofness, anonymity and efficiency. On the other hand, we show possibility results when either anonymity or efficiency is weakened.

Keywords: approval voting, efficiency, evaluationwise strategy-proofness, preference-approval, strategy-proofness,

*The work of Erdamar has been supported by the grant NET from ANR and BIDEB 2219 from Tubitak. The work of Sanver has been partly supported by COST Action IC1205 on Computational Social Choice, and the project ANR-14-CE24-0007-01 "CoCoRICo-CoDec". The work of Sato has been supported by JSPS KAKENHI 25780142 and 16K03571. We are grateful to William Thomson, Nick Baigent, Alan Kirman, Maurice Salles, Jean-François Laslier, Jean Lainé, Gabrielle Demange, Fuad Aleskerov, Miguel A. Ballester, Antonin Macé, Rafael Treibich, Patrick Harless, Andrew MacKenzie for their helpful comments.

1 Introduction

Since Gibbard (1973) and Satterthwaite (1975) who establish that every non-dictatorial and surjective social choice function defined over the full domain of preference profiles is manipulable, we observe the growth of a literature which investigates the effect of weakening these conditions which are logically incompatible. In fact, the Gibbard-Satterthwaite impossibility turns out to be quite robust when restricted domains are assumed, or multi-valued/probabilistic social choice rules are allowed.[1]

As another line of research, we see attempts to weaken strategy-proofness in some direction of interest. Campbell and Kelly (2009) define "gains from manipulation" and assume that voters are reluctant to manipulate unless the reward of lying is "sufficiently big".[2] Reffgen (2011) adopts a similar approach and considers manipulations where the manipulating voter can obtain his best or second best alternative. Sato (2013) restricts options of misrepresentation by considering only preferences "adjacent" to the true one, hence ruling out "big lies". None of these weakenings of strategy-proofness allow an escape from the Gibbard-Satterthwaite impossibility. As a recent and strong result in the same spirit, Muto and Sato (2016) define a condition that is much weaker than strategy-proofness, and they show that the condition is sufficient for the impossibility. In brief, the current state-of-the-art gives little hope to find a plausible weakening of strategy-proofness that would overcome an impossibility of the Gibbard-Satterthwaite type, as long as one is confined to the Arrovian framework of preference aggregation.

We carry the analysis into an informationally richer setting, called by Brams and Sanver (2009), the "preference-approval" framework, where voters not only

[1]For example, see Barberà, Dutta, and Sen (2001), Benoît (2002), and Pramanik (2015) among many others.

[2]Their analysis assumes a finite number of alternatives and strict individual preferences where gains from manipulation are measured by the number of ranks the manipulating voter can rise in his (true) preference.

rank alternatives but also evaluate them as acceptable (approved) or unacceptable (disapproved). By making use of this richer setting, we postulate that manipulation occurs if and only if the manipulation leads to a switch from an unacceptable outcome to an acceptable one, i.e., a gain from manipulation is "big". This definition renders manipulation harder than its usual understanding where obtaining any higher ranked outcome is a sufficient incentive for manipulation. As a result, non-manipulability in our sense leads to a weaker version of strategy-proofness, which we call evaluationwise strategy-proofness.[3]

For example, consider an election. Voters often classify the candidates as eligible or ineligible depending on what they attach importance. If a voter thinks that racial or religious diversity is critical, she will consider those candidates who support such diversity as eligible, and the rest ineligible. Moreover, it might be arguable that a voter cares manipulating the outcome if and only if he can change the winner from an ineligible candidate to an eligible one. Of course, the extent to which such an argument reflects behavioral reality is a matter of experimental research. Nevertheless, we see it as a probable description of voter behavior and see evaluationwise strategy-proofness as a concept that is well-worth being investigated.

We first show, by Corollary 1 to Theorem 1, that the Gibbard–Satterthwaite impossibility for strategy-proofness covers the preference-approval framework as well. Moreover, Theorem 1 establishes that the conjunction of evaluationwise strategy-proofness with "approval invariance", i.e., a rule must ignore evaluations, is equivalent to strategy-proofness in the Gibbard–Satterthwaite sense. Thus, on a domain of preferences on which the Gibbard-Satterthwaite impossibility holds,

[3]This idea is in the spirit of Campbell and Kelly (2009) who argue (in p.350) that "if a rule allows only small gains from manipulation, we may be persuaded that the cost to the individual of gathering enough information about the preferences of others to ensure that manipulation is advantageous will not yield a gain once those costs are factored in."

evaluationwise strategy-proof rules which are non-dictatorial and surjective, if any, are among those which are not approval invariant.

Nevertheless, even when we dispense with approval-invariance and benefit from the richness of extended inputs to conceive social choice rules undefined in the ranking aggregation model, a Gibbard–Satterthwaite type of impossibility prevails: Theorem 2 announces the logical incompatibility between efficiency, evaluationwise strategy-proofness, and anonymity of a social choice function when there is an even number of voters. However, this result does not preclude the existence of efficient and evaluationwise strategy-proof social choice functions: Theorem 3 establishes the existence of a social choice function which is evaluationwise strategy-proof, efficient and almost anonymous; in a similar vein, Theorem 4 shows that when efficiency is replaced by a weaker unanimity condition, the tension between anonymity and evaluationwise strategy-proofness vanishes. Hence, evaluationwise strategy-proofness is a weakening of strategy proofness that can serve as a criterion to define a degree of manipulability capable to differentiate among non-dictatorial social choice functions within the preference-approval framework.

We wish to note that our framework is part of a growing literature in social choice theory where individuals are assumed to evaluate alternatives through a common language and the collective decision is seen as an aggregation of evaluations, which are possibly combined with rankings that form the basis of traditional social choice theory. Among the social choice rules that use evaluations as inputs, Approval Voting (AV) is perhaps the most studied. The seminal analysis of AV by Brams and Fishburn (1978, 2007) predates the formal modeling of evaluations as inputs of the collective choice model. As a result, there has been a period where discussions on AV reflected difficulties of expressing evaluations within the framework of ranking aggregations.[4] This incompatibility is handled by Brams and

[4] Among these exchanges, one can non-exhaustively cite Niemi (1984, 1985), Brams and Fishburn (1985), Saari and Newnhizen (1988a,b), and Brams, Fishburn, and Merill (1988a,b). See Laslier and

Sanver (2006) who revisit Approval Voting in the preference-approval framework which is one of the first models that explicitly combines rankings with evaluations and which forms the framework of our analysis.

AV is restricted to binary evaluations, i.e., admits that a voter conceives each alternative as either "approved" or "disapproved". However, there are social choice rules that enrich the inputs by allowing voters to express their evaluations in more than two categories. Among these, one can cite Evaluative Voting analyzed by Hillinger (2005); the threshold aggregation of three graded rankings proposed by Aleskerov, Yakuba, and Yuzbashev (2007) and Range Voting by Smith (2000).

Balinski and Laraki (2011) propose a general theory of evaluation aggregation and advocate that it is evaluations which must be aggregated and not rankings. Nevertheless, as Brams and Sanver (2009) indicate, there is a literature that takes a compromising direction by using a combination of rankings and evaluations as the input of the collective choice problem.[5] As a matter of fact, evaluationwise strategy-proofness cannot be appropriately formulated unless a model that combines rankings and evaluations is considered. It is worth noting that regarding the analysis of strategy-proofness, the preference-approval framework subsumes the standard Arrovian setting (which omits evaluations) as well as the Balinski and Laraki (2011) setting (which omits rankings).[6]

Section 2 of our paper introduces the preference-approval framework and the other notions including evaluationwise strategy-proofness. Our results are presented in Section 3, and finally Section 4 makes some concluding remarks.

Sanver (2010) for an account of the developments on AV.

[5]As pointed out by Sanver (2010), this combination extends the ordinal ranking aggregation framework by introducing cardinal elements into individual preferences.

[6]To be sure, Balinski and Laraki (2011) allow more than two evaluations but the notion of evaluationwise strategy-proofness is naturally extendable to any number of evaluations.

2 Basic notions and the setting

2.1 Preference-approval framework

We begin with the basic terminology for a collective decision making problem. Consider a set of **voters** $N = \{1, \ldots, n\}$ with $n \geq 2$ and a set of **candidates** X with $|X| = m$ where $m \geq 3$.[7] Let $\mathcal{L}$ be the set of all **linear orders** on X.[8] Let 2^X be the set of all subsets of X. Voter i's preference over candidates is a linear order $R_i \in \mathcal{L}$. For each $k \in \{1, \ldots, m\}$, let $r_k(R_i)$ be the kth ranked alternative according to $R_i \in \mathcal{L}$. The strict part of R_i is P_i. We presume that each voter partitions the set of candidates into two subsets: a set of **acceptable** candidates, $A_i \subseteq X$, and a set of **unacceptable** candidates, $U_i = X \setminus A_i$. We also assume a consistency condition: if $y \in X$ is acceptable, all candidates preferred to y should be acceptable as well. Formally, a **preference-approval** of voter $i \in N$ is a pair $p_i = (R_i, A_i) \in \mathcal{L} \times 2^X$ such that

$$\text{for each } x, y \in X,\ \big[(x\ R_i\ y \text{ and } y \in A_i) \ \Rightarrow\ x \in A_i\big].$$

Let Π be the set of all preference-approvals.

We allow domain restrictions by considering a set of **admissible preferences**, $\mathcal{D} \subseteq \mathcal{L}$, and we write $\Pi_{\mathcal{D}} = \{(R_i, A_i) \in \Pi \mid R_i \in \mathcal{D}\}$ for the set of **admissible preference-approvals**. Note that $\Pi = \Pi_{\mathcal{L}}$. Note also that we do not impose any restrictions over the acceptable set of candidates except the consistency condition in the definition of a preference-approval. Thus, if $R_i \in \mathcal{L}$ is admissible, then for each $A_i \in 2^X$ satisfying the consistency condition, (R_i, A_i) is admissible. As a result, for each $R_i \in \mathcal{L}$, there exist $m + 1$ possible preference-approvals.

[7]For each set S, $|S|$ denotes its cardinality.

[8]A binary relation B is a subset of $X \times X$ and we write $x\ B\ y$ for $(x, y) \in B$. A binary relation B is **complete** if for each $x, y \in X$, $x\ B\ y$ or $y\ B\ x$, **transitive** if for each $x, y, z \in X$, $x\ B\ y$ and $y\ B\ z$ imply $x\ B\ z$, **antisymmetric** if for each $x, y \in X$, $x\ B\ y$ and $y\ B\ x$ imply $x = y$. Finally, a binary relation is a **linear order** if it is complete, transitive, and antisymmetric.

$\Pi_{\mathcal{D}}^n$ is the set of all admissible profiles of preference-approvals, and whose typical member is denoted by $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$, where p_i stands for the preference-approval of voter $i \in N$. We write $(p_i', \boldsymbol{p}_{-i})$ to denote that voter i's preference-approval is p_i' and the rest of the other voters' is $\boldsymbol{p}_{-i}$.

2.2 Collective decision rules and axioms

For each $\mathcal{D} \subseteq \mathcal{L}$, a **rule** f **on** $\Pi_{\mathcal{D}}^n$ is a single-valued function from the set of admissible profiles of preference-approvals, $\Pi_{\mathcal{D}}^n$, into the set of candidates, X. So, given a profile, a rule aggregates two pieces of information, orders and binary evaluations, into a social outcome. This outcome may be acceptable for some voters and unacceptable for others.

When a rule depends only on ranking information $(R_1, R_2, \ldots, R_n)$ in the sense of the following approval independence axiom, that rule can be defined in the standard Arrovian model. On the other hand, rules of our framework that are not approval invariant cannot be expressed in the standard model. Thus, our model is an extension of the Arrovian model.

- **Approval Invariance:** For each $\boldsymbol{p}, \boldsymbol{p}' \in \Pi_{\mathcal{D}}^n$ such that $R_i = R_i'$ for each $i \in N$, $f(\boldsymbol{p}) = f(\boldsymbol{p}')$.

We define efficiency as in the Arrovian model. We say that $x \in X$ is **efficient** if there is no $y \in X$ such that $y \; P_i \; x$ for each $i \in N$.

- **Efficiency:** For each $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$, $f(\boldsymbol{p})$ is an efficient candidate.

The following axiom, which is weaker than efficiency, states that when there is a unanimous agreement on the best candidate, a rule should respect this agreement.

- **Unanimity:** For each $x \in X$ and each $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$ such that $r_1(R_i) = x$ for each $i \in N$, $f(\boldsymbol{p}) = x$.

Symmetric treatment of the voters is ensured by defining the standard anonymity axiom in our framework.

- **Anonymity:** For each $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$ and each permutation π of N, $f(\boldsymbol{p}) = f(\boldsymbol{p}')$, where $p'_i = p_{\pi^{-1}(i)}$ for each $i \in N$.

A group of voters $N' \subseteq N$ is **decisive for** $x \in X$ under a rule f on $\Pi_{\mathcal{D}}^n$ if for each profile $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$ such that $A_i = \{x\}$ for each $i \in N'$, $f(\boldsymbol{p}) = x$. We say that a group N' is **decisive** if it is decisive for each candidate, and a voter $i \in N$ is decisive if $\{i\}$ is decisive. Under a unanimous rule, N is decisive.

A rule f on $\Pi_{\mathcal{D}}^n$ is **dictatorial** if there is $i \in N$ such that $f(\boldsymbol{p}) = r_1(R_i)$ for each $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$. Such i is called a **dictator**. By definition, if $i \in N$ is a dictator, then he is decisive. Therefore, if there is no decisive voter, the rule is non-dictatorial.

2.3 Evaluationwise strategy-proofness

We begin with a restatement of the strategy-proofness axiom for the preference-approval domain.

A rule on $\Pi_{\mathcal{D}}^n$ is **manipulable** if there are $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$, $i \in N$ and $p'_i \in \Pi_{\mathcal{D}}$ such that

$$f(p'_i, \boldsymbol{p}_{-i}) \; P_i \; f(\boldsymbol{p}).$$

A rule on $\Pi_{\mathcal{D}}^n$ is **strategy-proof** if it is not manipulable.

In the formulation of strategy-proofness, each voter has an incentive to misrepresent his preference-approval if and only if a more preferred candidate is chosen by this misrepresentation. On the other hand, we also postulate that a voter has an incentive to misrepresent his preference-approval if and only if he can change an unacceptable outcome to an acceptable one. Thus, just achieving a higher ranked candidate is not a sufficient incentive for misrepresentation.

A rule on $\Pi^n_{\mathcal{D}}$ is **evaluationwise manipulable** if there are $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$, $i \in N$ and $p'_i \in \Pi_{\mathcal{D}}$ such that

$$f(\boldsymbol{p}) \in U_i \text{ and } f(p'_i, \boldsymbol{p}_{-i}) \in A_i.$$

A rule on $\Pi^n_{\mathcal{D}}$ is **evaluationwise strategy-proof** if it is not evaluationwise manipulable.

Although strategy-proofness clearly implies evaluationwise strategy-proofness, it is not that clear whether strategy-proofness exhibits a Gibbard–Satterthwaite type of impossibility in our framework where voters have more strategic tools of manipulation (i.e., rankings and approvals) which renders manipulation easier but also due to richer information, there are more conceivable rules than the standard Arrovian framework. The following theorem shows that the impact of additional manipulation tools is dominant and the Gibbard–Satterthwaite impossibility prevails in our framework.

Theorem 1

Let $\mathcal{D} \subseteq \mathcal{L}$, and f be a rule on $\Pi^n_{\mathcal{D}}$. Then, f is strategy-proof if and only if it is evaluationwise strategy-proof and approval invariant.

Proof. Strategy-proofness $\Rightarrow$ evaluationwise strategy-proofness:

Assume that f satisfies strategy-proofness. Let $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$, $i \in N$, and $p'_i \in \Pi_{\mathcal{D}}$. Assume $f(\boldsymbol{p}) \in U_i$. We want to show that $f(p'_i, \boldsymbol{p}_{-i}) \notin A_i$. Suppose $f(p'_i, \boldsymbol{p}_{-i}) \in A_i$. By the definition of a preference-approval, $f(p'_i, \boldsymbol{p}_{-i})\ P_i\ f(\boldsymbol{p})$. This is a contradiction to strategy-proofness.

Strategy-proofness $\Rightarrow$ approval invariance:

Assume that f satisfies strategy-proofness. Let $\boldsymbol{p}, \boldsymbol{p}' \in \Pi^n_{\mathcal{D}}$ be such that $R_i = R'_i$ for each $i \in N$. At $\boldsymbol{p}$, let voter 1 change his preference-approval from p_1 to p'_1. Then, the profile of preference-approvals changes from $\boldsymbol{p}$ to $(p'_1, \boldsymbol{p}_{-1})$. We want to show $f(\boldsymbol{p}) = f(p'_1, \boldsymbol{p}_{-1})$. Suppose $f(\boldsymbol{p}) \neq f(p'_1, \boldsymbol{p}_{-1})$. Then, either

$f(\boldsymbol{p})\ P_1\ f(p'_1, \boldsymbol{p}_{-1})$ or $f(p'_1, \boldsymbol{p}_{-1})\ P_1\ f(\boldsymbol{p})$. In the former case, since $R_1 = R'_1$, we have $f(\boldsymbol{p})\ P'_1\ f(p'_1, \boldsymbol{p}_{-1})$. This is a contradiction to strategy-proofness. The latter case is also a contradiction to strategy-proofness. Thus, $f(\boldsymbol{p}) = f(p'_1, \boldsymbol{p}_{-1})$. At $(p'_1, \boldsymbol{p}_{-1})$, let voter 2 change his preference-approval from p_2 to p'_2. By the same arguments, the social choice remains the same. In this way, after all voters change their preference-approvals from those in $\boldsymbol{p}$ to those in $\boldsymbol{p}'$, the social choice does not change. Thus, $f(\boldsymbol{p}) = f(\boldsymbol{p}')$.

Evaluationwise strategy-proofness and *approval invariance* $\Rightarrow$ *strategy-proofness*:

Assume that f satisfies evaluationwise strategy-proofness and approval invariance. Let $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$, $i \in N$, and $p'_i \in \Pi_{\mathcal{D}}$. Let $p^*_i = (R^*_i, A^*_i) \in \Pi_{\mathcal{D}}$ be such that $R^*_i = R_i$ and $f(p^*_i, \boldsymbol{p}_{-i})$ is the best alternative in U^*_i according to R^*_i. By evaluationwise strategy-proofness, $f(p'_i, \boldsymbol{p}_{-i}) \in U^*_i$. Since $f(p^*_i, \boldsymbol{p}_{-i})$ is the best alternative in U^*_i according to R^*_i, we have $f(p^*_i, \boldsymbol{p}_{-i})\ R^*_i\ f(p'_i, \boldsymbol{p}_{-i})$. Since $R^*_i = R_i$ and f is approval-invariant, $f(p^*_i, \boldsymbol{p}_{-i}) = f(\boldsymbol{p})$. Thus, we have $f(\boldsymbol{p})\ R_i\ f(p'_i, \boldsymbol{p}_{-i})$. This implies that f is strategy-proof. ■

This result has two important implications;

First, it shows that strategy-proofness in the preference-approval model and that in the standard Arrovian model are equivalent. To clarify the meaning of this equivalence, we introduce additional notations. Given $\mathcal{D} \subseteq \mathcal{L}$, let $\mathcal{P}$ be the set of all rule on $\Pi^n_{\mathcal{D}}$ in the preference-approval model, and let $\mathcal{A}$ be the set of all rules on $\mathcal{D}^n$ in the standard Arrovian model, i.e., each $F \in \mathcal{A}$ is such that $F(\boldsymbol{R}) \in X$ for each $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{D}^n$. Moreover, let $\mathcal{P}^* \subset \mathcal{P}$ be the set of all strategy-proof rules on $\Pi^n_{\mathcal{D}}$, and $\mathcal{A}^* \subset \mathcal{A}$ be the set of all strategy-proof rules on $\mathcal{D}^n$. Let φ be a function from $\mathcal{A}$ to $\mathcal{P}$ such that for each $F \in \mathcal{A}$, $f \equiv \varphi(F) \in \mathcal{P}$ is the (approval invariant) **equivalent** of F in $\mathcal{P}$. Formally, for each $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Pi^n_{\mathcal{D}}$, $f(\boldsymbol{p}) = F(\boldsymbol{R})$, where $p_i = (R_i, A_i)$ for each $i \in N$ and $\boldsymbol{R} = (R_1, \ldots, R_n)$.

Similarly, let φ^* be a function from $\mathcal{A}^*$ to $\mathcal{P}^*$ such that for each $F \in \mathcal{A}^*$, $\varphi^*(F) \in \mathcal{P}$ is the (approval invariant) equivalent of F. (Since strategy-proofness of $F \in \mathcal{A}^*$ implies strategy-proofness of its (approval invariant) equivalent, the value of φ^* is in $\mathcal{P}^*$.)

Since $\mathcal{P}$ contains rules that are not approval invariant, φ is not a bijection from $\mathcal{A}$ to $\mathcal{P}$. However, Theorem 1 implies that φ^* is a bijection from $\mathcal{A}^*$ to $\mathcal{P}^*$. In this sense, strategy-proofness is equivalent in the two frameworks.

Corollary 1

φ^* is a bijection from $\mathcal{A}^*$ to $\mathcal{P}^*$.

Proof. It is clear that φ^* is an injection. We show that φ^* is onto. Let $f \in \mathcal{P}^*$. As Theorem 1 shows, f is approval invariant. Since f depends only on ranking information, we can define $F \in \mathcal{A}$ as follows: for each $\boldsymbol{R} \in \mathcal{D}^n$, $F(\boldsymbol{R}) = f(\boldsymbol{p})$, where $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$ is such that $p_i = (R_i, A_i)$ for each $i \in N$. It is clear that $\varphi(F) = f$. Thus, it suffices to show $F \in \mathcal{A}^*$. Since strategy-proofness depends only on ranking information and f is strategy-proof, F is also strategy-proof. Thus, $F \in \mathcal{A}^*$. ■

Second, on each $\mathcal{D} \subseteq \mathcal{L}$ such that strategy-proofness and efficiency imply dictatorship in the Arrovian model, rules on $\Pi_{\mathcal{D}}^n$ (i.e., in the preference-approval model) that are evaluationwise strategy-proof, efficient and non-dictatorial exist only in the family of rules that are sensitive to the binary evaluation component.

Corollary 2

Let $\mathcal{D} \subseteq \mathcal{L}$. Assume that each rule on $\mathcal{D}^n$ satisfying strategy-proofness and efficiency is dictatorial. Then, each non-dictatorial rule on $\Pi_{\mathcal{D}}^n$ satisfying evaluationwise strategy-proofness and efficiency is not approval invariant.

Proof. Suppose that we find an approval invariant rule f on $\Pi_{\mathcal{D}}^n$ that is evaluationwise strategy-proof, efficient and non-dictatorial. By Theorem 1, f is strategy-proof. Since f is approval invariant, it can be considered as a rule in the Arrovian model (Corollary 1). As a rule in the Arrovian model, it inherits each property based only on ranking information from a rule in the preference-approval model. Then, f is a rule in the Arrovian model and it is strategy-proof, efficient, and non-dictatorial. However, this is a contradiction to the assumption that there is no such a rule on $\mathcal{D}^n$. ■

2.4 Relationship to strategy-proofness in dichotomous domains

In Section 2.3, we clarified the relationship between evaluationwise strategy-proofness in the preference-approval framework and strategy-proofness in the standard Arrovian framework. In this section, we clarify the relationship between evaluationwise strategy-proofness in the preference-approval framework and strategy-proofness over (Arrovian) dichotomous domains. Specifically, we will show that the analysis on dichotomous domains can be subsumed in the preference-approval framework.

A preference is **dichotomous** if it has at most two indifference classes. Dichotomous preferences are weak orders,[9] so ties are allowed. When there are exactly two indifference classes, we interpret the upper indifference class as the set of "good" candidates and the lower indifference class as the set of "bad" ones. When there is only one indifference class, we choose the convention that all candidates are "good". Let $\tilde{\mathcal{D}}$ be the set of all dichotomous preferences. A rule F on $\tilde{\mathcal{D}}^n$ is a function from $\tilde{\mathcal{D}}^n$ to X. The following is a key property in expressing rules in the preference-approval framework and rules on the dichotomous domain in terms of each other.

- **Reshuffling Invariance:** For each $\boldsymbol{p}, \boldsymbol{p}' \in \Pi_{\mathcal{D}}^n$ such that $A_i = A_i'$ for each $i \in N$, $f(\boldsymbol{p}) = f(\boldsymbol{p}')$.[10]

[9] A binary relation is a weak order if it is complete and transitive.

[10] Barberà, Berga, and Moreno (2012) consider a condition called "reshuffling invariance" in the standard framework where only rankings are inputs to social choice rules. Their "reshuffling" is within the upper and the lower contour sets of the socially chosen alternative, while ours is within

When a rule in the approval-preference framework is reshuffling invariant, it does not depend on how the candidates are ranked within approved ones and disapproved ones.

We say that a dichotomous preference $Q_i \in \tilde{\mathcal{D}}$ with two indifference classes is **derived from** a preference-approval $p_i \in \Pi$ if for each $x, y \in X$, $x \in A_i$ and $y \in U_i$ imply $x \; Q_i \; y$ and not $y \; Q_i \; x$, i.e., x is a "good" candidate and y is a "bad" candidate according to Q_i. When $Q_i \in \tilde{\mathcal{D}}$ consists of a single indifference class, it is derived from a preference-approval $p_i = (R_i, A_i)$ such that $x \in A_i$ for each $x \in X$.[11] Note that for each $Q_i \in \tilde{\mathcal{D}}$, there is always a preference-approval $p_i \in \Pi$ such that Q_i is derived from p_i. Also, given a profile of dichotomous preferences $\boldsymbol{Q} = (Q_1, \ldots, Q_n) \in \tilde{\mathcal{D}}^n$, if a rule f on Π^n is reshuffling invariant, we have $f(\boldsymbol{p}) = f(\boldsymbol{p}')$ for each $\boldsymbol{p}, \boldsymbol{p}' \in \Pi^n$ such that each Q_i is derived from both p_i and p'_i.

For each rule F on the dichotomous domain $\tilde{\mathcal{D}}^n$, we can define the (reshuffling invariant) **equivalent** f of F in the preference-approval framework as follows: for each $\boldsymbol{p} \in \Pi^n$, let $f(\boldsymbol{p}) = F(\boldsymbol{Q})$, where each Q_i in $\boldsymbol{Q} = (Q_1, \ldots, Q_n)$ is derived from p_i.[12] Note that this f is reshuffling invariant. Conversely, each reshuffling invariant rule f on Π^n can be transformed to the **equivalent** rule F on the dichotomous domain in the following way: for each $\boldsymbol{Q} \in \tilde{\mathcal{D}}^n$, let $F(\boldsymbol{Q}) = f(\boldsymbol{p})$, where

---

the approved and the disapproved alternatives.

[11] We could say that Q_i is derived from $p_i = (R_i, A_i)$ such that $x \in U_i$ for each $x \in X$. Our choice reflects the convention that all candidates are "good" when there is only a single indifference class.

[12] Only in this definition, we disregard the convention that all candidates are "good" when there is one single indifference class, and we assume that Q_i with only one indifference class is derived from also a preference-approval p_i such that $x \in U_i$ for each $x \in X$. Otherwise, we cannot define $f(\boldsymbol{p})$ when some p_i in $\boldsymbol{p}$ is such that $x \in U_i$ for each $x \in X$. As a result of this deviation from the convention, given $\boldsymbol{p}_{-i} \in \Pi^{n-1}$, $f(p_i, \boldsymbol{p}_{-i}) = f(p'_i, \boldsymbol{p}_{-i})$ for each $p_i = (R_i, A_i)$ and $p'_i = (R'_i, A'_i)$ such that $A_i = X$ and $A'_i = \emptyset$.

each p_i in $\boldsymbol{p} \in \Pi^n$ is such that Q_i is derived from p_i.[13] In this sense, although we rule out ties in preferences, the preference-approval model is an extension of the model of the dichotomous domain.

We wish to note that strategy-proofness of F on $\tilde{\mathcal{D}}^n$ and evaluationwise strategy-proofness of its (reshuffling invariant) equivalent f on Π^n are equivalent: if we have a strategy-proof rule F on the dichotomous domain $\tilde{\mathcal{D}}^n$, then its (reshuffling invariant) equivalent f is evaluationwise strategy-proof. To see this, let $\boldsymbol{p} \in \Pi^n$, $i \in N$, and $p'_i \in \Pi$. Let $\boldsymbol{Q} \in \tilde{\mathcal{D}}^n$ be such that for each $j \in N$, Q_j is derived from p_j, and let Q'_i be such that it is derived from p'_i. By strategy-proofness of F, we have $F(\boldsymbol{Q})\ Q_i\ F(Q'_i, \boldsymbol{Q}_{-i})$. Since f is the (reshuffling invariant) equivalent of F, $f(\boldsymbol{p}) = F(\boldsymbol{Q})$ and $f(p'_i, \boldsymbol{p}_{-i}) = F(Q'_i, \boldsymbol{Q}_{-i})$. When Q_i has two indifference classes, there are three cases to consider: according to Q_i, (1) both $F(\boldsymbol{Q})$ and $F(Q'_i, \boldsymbol{Q}_{-i})$ are "good", (2) both of them are "bad", and (3) $F(\boldsymbol{Q})$ is "good" and $F(Q'_i, \boldsymbol{Q}_{-i})$ is "bad". In the first case, since Q_i is derived from p_i, both $f(\boldsymbol{p})$ and $f(p'_i, \boldsymbol{p}_{-i})$ belong to A_i. In the second case, both of them belong to U_i. In the third case, we have $f(\boldsymbol{p}) \in A_i$ and $f(p'_i, \boldsymbol{p}_{-i}) \in U_i$. In each case, f is not evaluationwise manipulable, i.e., f is evaluationwise strategy-proof. When Q_i has only one indifference class, either $A_i = X$ or $A_i = \emptyset$. Thus, f is evaluationwise strategy-proof also in this case. Conversely, we can see in a similar manner that if we have a reshuffling invariant and evaluationwise strategy-proof rule f on Π^n, then its equivalent F on the dichotomous domain is strategy-proof. Thus, results on strategy-proof rules on dichotomous domains can be translated to results on evaluationwise strategy-proof and reshuffling invariant rules in the preference-approval framework, and vice versa.[14]

[13] There are multiple candidates for such p_i. As we note in the previous paragraph, due to reshuffling invariance, the social choice of f is not affected by the choice of p_i.

[14] There are analyses of strategy-proof social choice rules over dichotomous domains such as Brams and Fishburn (1978) who analyze the strategy-proofness of AV; Vorsatz (2007) who con-

However, our results in the next section use efficiency or unanimity in the preference-approval model, and these axioms are incompatible with reshuffling invariance. Thus, none of our results can be translated to results on strategy-proof rules on dichotomous domains.

3 Impossibility and possibility results

We investigate the existence of evaluationwise strategy-proof rules that are efficient and anonymous. Before going to a general result, as an example of the incompatibility between evaluationwise strategy-proofness, efficiency and anonymity, consider a refinement of Approval Voting:

For each profile $\boldsymbol{p} \in \Pi^n$ and each candidate $x \in X$, we denote the number of voters who approve of x as $a(x, \boldsymbol{p}) = |\{i \in N \mid x \in A_i\}|$. We define the **AV winners** as the candidates who are approved by the maximum number of voters, and denote this set by $AV(\boldsymbol{p}) = \{x \in X \mid \forall y \in X,\ a(x, \boldsymbol{p}) \geq a(y, \boldsymbol{p})\}$. Note that some AV winners may not be efficient.

Example 3.1

Define a refinement of AV as follows: Let f^{AV} be a rule on Π^n such that for each profile $\boldsymbol{p} \in \Pi^n$, $f^{AV}(\boldsymbol{p})$ is the efficient AV winner if it is unique, and if not, $f^{AV}(\boldsymbol{p})$ is the first efficient AV winner according to the alphabetical order.

One can easily see that f^{AV} is efficient and anonymous. However, it is evaluationwise manipulable, hence not evaluationwise strategy-proof. We illustrate this

siders AV as a non-resolute social choice rule and gives a characterization in terms of strategy-proofness, anonymity, neutrality and monotonicity; Bogomolnaia, Moulin, and Stong (2005) characterize AV when outcomes are lotteries over candidates. However, we don't know a full characterization of (anonymous) strategy-proof rules over dichotomous domains within the standard Gibbard-Satterthwaite setting. On the other hand, there are considerations of Arrovian social welfare functions over dichotomous domains, such as Sakai and Shimoji (2006) and more recently Maniquet and Mongin (2015).

p_1	p_2	p_3
z	y	z
x	$-$	y
$-$	z	x
y	x	$-$

Table 1: A preference-approval profile at which f^{AV} is evaluationwise manipulable.

for the case of three voters.

Consider the profile $\boldsymbol{p}$ in Table 1.[15] Note that $AV(\boldsymbol{p}) = \{x, y, z\}$. By efficiency, $f(\boldsymbol{p}) \neq x$. Among y and z, according to the alphabetical order, $f^{AV}(\boldsymbol{p}) = y$. This social choice, y, is unacceptable to voter 1. By a misreport of $p_1' = (xz|y)$, the outcome changes from y to x, which is acceptable to voter 1.[16] Thus, f^{AV} is evaluationwise manipulable.

We will show that this incompatibility between evaluationwise strategy-proofness, efficiency, and anonymity always occurs when $\mathcal{D}$ is "rich" in the following sense.

$\mathcal{D} \subseteq \mathcal{L}$ is **circular** if candidates can be arranged on a circle as follows: Let x be an arbitrary candidate, y and z be candidates adjacent to x on the circle. Then, there exist two linear orders in $\mathcal{D}$ such that x is top ranked in both of these orders, whereas in one of the orders y is the second ranked while z is the last ranked, and in the other order, z is the second ranked and y is the last ranked.

Formally, $\mathcal{D} \subseteq \mathcal{L}$ is **circular** if the candidates can be indexed $x_1, x_2, \ldots, x_m$ such that $x_{m+1} = x_1$ and $x_0 = x_m$ and for each $k \in \{1, \ldots, m\}$, there exist two linear orders R, R' in $\mathcal{D}$ with

[15] In Table 1, the horizontal lines between candidates represent the boundary between the acceptable and the unacceptable range.

[16] $p_1' = (xz|y)$ means that $x \ P_1' \ z \ P_1' \ y$ and $A_1' = \{x, z\}$, where $p_1' = (R_1', A_1')$.

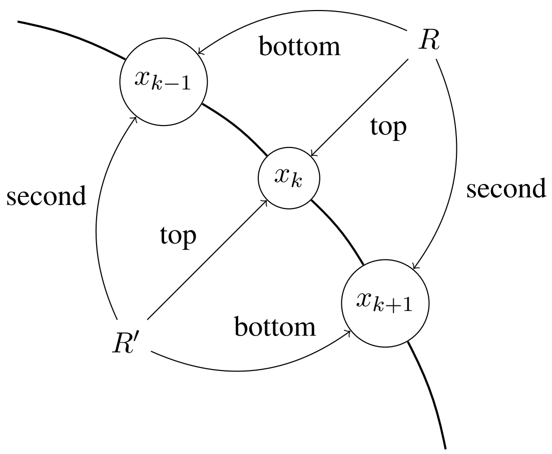

Figure 1: A circular set of preferences

(i) $r_1(R) = x_k$, $r_2(R) = x_{k+1}$, $r_m(R) = x_{k-1}$,

(ii) $r_1(R') = x_k$, $r_2(R') = x_{k-1}$, and $r_m(R') = x_{k+1}$.

This situation is summarized in Figure 1. We say that $\Pi_{\mathcal{D}}$ is **circular** if $\mathcal{D}$ is circular.

The following remarks are worth to note for circular sets of preferences.

- The minimal circular sets of preferences consist of $2m$ preferences since each candidate is top ranked in at least two distinct preferences.

- If $\mathcal{D} \subseteq \mathcal{L}$ is circular, then each $\mathcal{D}' \subseteq \mathcal{L}$ with $\mathcal{D} \subset \mathcal{D}'$ is also circular.

- $\mathcal{L}$ is circular.

- A necessary condition for $\mathcal{D}$ to be circular is that for each $x \in X$, there exist $y \in X$ and a pair $R, R' \in \mathcal{D}$ such that $r_1(R) = x, r_2(R) = y, r_1(R') = x$, and $r_m(R') = y$. Violation of this condition implies that no alternative can be adjacent to x on the circle.

The next result shows the incompatibility of anonymity and efficiency for evaluationwise strategy-proof rules.

Theorem 2

Let n be an even number, and $\mathcal{D}$ be a circular set of preferences. Then, there exists no rule on $\Pi^n_{\mathcal{D}}$ which is anonymous, efficient and evaluationwise strategy-proof.

Proof. Let f be an anonymous, efficient, and evaluationwise strategy-proof rule on $\Pi^n_{\mathcal{D}}$, where $\mathcal{D}$ is circular. Assign a number from 1 to m to each candidate so as to make $\mathcal{D}$ circular. We point out two simple facts.

FACT 1: Let $i \in N$ and $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$. If $A_i = \{f(\boldsymbol{p})\}$, then for each $p'_i = (R'_i, A'_i) \in \Pi_{\mathcal{D}}$ such that $A'_i = \{f(\boldsymbol{p})\}$, evaluationwise strategy-proofness implies $f(p'_i, \boldsymbol{p}_{-i}) = f(\boldsymbol{p})$.

FACT 2: Let $i \in N$ and $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$. If $U_i = \{f(\boldsymbol{p})\}$, then for each $p'_i \in \Pi_{\mathcal{D}}$, evaluationwise strategy-proofness implies $f(p'_i, \boldsymbol{p}_{-i}) = f(\boldsymbol{p})$.

Let $\{N_1, N_2\}$ be a partition of N.

CLAIM 1: For each $k \in \{1, \ldots, m\}$, either N_1 is decisive for x_k or N_2 is decisive for x_{k+1}.

Proof of Claim 1. The following arguments are modification of those by Sato (2010). Let $x_k \in X$. Assume that N_1 is not decisive for x_k.

At $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$ in Table 2, $f(\boldsymbol{p}) \neq x_k$.[17] (Otherwise, by Facts 1 and 2, as long as all voters in N_1 approve only x_k, the social choice is x_k. This is a contradiction to our assumption that N_1 is not decisive for x_k.) By efficiency, $f(\boldsymbol{p}) = x_{k+1}$.

Next, consider $\boldsymbol{p'} \in \Pi^n_{\mathcal{D}}$ in Table 2. (For each $i \in N_1$, $p_i = p'_i$.) By evaluationwise strategy-proofness, $f(\boldsymbol{p'}) = x_{k+1}$. (Suppose not. Then, when the voters in N_2 change their preference-approvals from those in $\boldsymbol{p}'$ to those in $\boldsymbol{p}$ one voter at a

[17]The unspecified parts in Table 2 are arbitrary. The candidate between the brackets is a social outcome at each profile.

	$\boldsymbol{p}$		$\boldsymbol{p'}$		$\boldsymbol{p''}$	
	N_1	N_2	N_1	N_2	N_1	N_2
Best	x_k	$[x_{k+1}]$	x_k	$[x_{k+1}]$	x_k	$[x_{k+1}]$
$\vdots$	–	x_{k+2}	–	–	x_{k-1}	–
	$[x_{k+1}]$	$\vdots$	$[x_{k+1}]$	x_k	$\vdots$	x_k
$\vdots$	$\vdots$	–	$\vdots$	$\vdots$	–	$\vdots$
Worst	x_{k-1}	x_k	x_{k-1}	x_{k+2}	$[x_{k+1}]$	x_{k+2}

Table 2: Profiles of preference-approvals

time, the social choice changes from some candidate in $X \setminus \{x_{k+1}\}$ to x_{k+1} at some step. Since $A'_i = \{x_{k+1}\}$ for each $i \in N_2$, this is a contradiction to evaluationwise strategy-proofness.)

Finally, consider $\boldsymbol{p''} \in \Pi^n_{\mathcal{D}}$ in Table 2. (For each $i \in N_2$, $p'_i = p''_i$.) By efficiency, $f(\boldsymbol{p''}) \in \{x_k, x_{k+1}\}$. Since $f(\boldsymbol{p''}) = x_k$ is a contradiction to evaluationwise strategy-proofness, $f(\boldsymbol{p''}) = x_{k+1}$. By Facts 1 and 2, N_2 is decisive for x_{k+1}.

CLAIM 2: Either N_1 is decisive or N_2 is decisive.

Proof of Claim 2. For each $x_k \in X$, either N_1 is decisive for x_k or N_2 is decisive for x_k. (If not, then by Claim 1, N_1 is decisive for x_{k-1} and N_2 is decisive for x_{k+1}. However, this cannot be the case.)

Let $x \in X$. Then, the argument in the above paragraph implies that either N_1 is decisive for x or N_2 is decisive for x. Without loss of generality, suppose that it is N_1 who is decisive for x. If for some $y \in X \setminus \{x\}$, N_1 is not decisive, then N_2 is decisive for y. Let $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$ be such that for each $i \in N_1$, $A_i = \{x\}$ and for each $i \in N_2$, $A_i = \{y\}$. Then, the rule selects x but also y, which is a contradiction since $x \neq y$.

CLAIM 3: If some $N' \subset N$ is decisive, then each $N'' \subset N$ with $|N'| = |N''|$ is decisive.

Proof of Claim 3. Assume that N' is decisive. We claim that N'' with $|N'| = |N''|$ is also decisive. Let $x \in X$ and $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$ be such that for each $i \in N''$, $A_i = \{x\}$. Let π be a permutation of N such that for each $i \in N'$, $\pi(i) \in N''$. By anonymity, $f(\boldsymbol{p}) = f(\pi(\boldsymbol{p}))$. Since N' is decisive, $f(\pi(\boldsymbol{p})) = x$. Thus, $f(\boldsymbol{p}) = x$. This implies that N'' is decisive for x. Since x was arbitrary, N'' is decisive.

Claim 2 holds for arbitrary partition $\{N_1, N_2\}$ of N. Specifically, it holds with N_1 and N_2 such that $|N_1| = |N_2|$. Then, by Claim 3, both N_1 and N_2 are decisive. However, two disjoint coalitions cannot be both decisive by the definition of the concept. This contradiction completes the proof. ■

Claims 1 and 2 in the above proof use only evaluationwise strategy-proofness and efficiency. On the other hand, Claim 3 uses only anonymity. Thus, for the particular case of two voters, by Claims 1 and 2 of the above proof, the following result can be derived.

Corollary 3

Let $n = 2$. Let $\mathcal{D}$ be a circular domain. If a rule on $\Pi^n_{\mathcal{D}}$ is efficient and evaluationwise strategy-proof, then either voter 1 or voter 2 is decisive.

The following example shows that the decisive voter in Corollary 3 need not to be a dictator.

Example 3.2

Let $n = 2$, and $\mathcal{D}$ be a circular domain. For each $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$, let $f(\boldsymbol{p})$ be the best alternative in A_1 according to R_2. (If $A_1 = \emptyset$, then let $f(\boldsymbol{p}) = r_1(R_2)$.) Then, voter 1 is decisive under this rule. It can be seen that this rule is efficient and evaluationwise strategy-proofness. Also, neither voter 1 nor 2 is a dictator.

Our next two results are positive conclusions for evaluationwise strategy-proof rules either by a weakening of anonymity or by a weakening of efficiency.

Theorem 3

Let $n \geq 3$. For each $\mathcal{D} \subseteq \mathcal{L}$ such that $r_1(R_i) \neq r_1(R_i')$ for some $R_i, R_i' \in \mathcal{D}$, there exists a rule on $\Pi_{\mathcal{D}}^n$ which is efficient and evaluationwise strategy-proof under which no voter is decisive.

Proof. Let f be a rule on $\Pi_{\mathcal{D}}^n$ defined as follows: Fix a linear order L over all sets of voters. Let $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$. For each $x \in X$, let $N(x, \boldsymbol{p})$ be the set of the voters who approve x. Then, among $\{N(x, \boldsymbol{p}) \mid x \text{ is an efficient AV winner at } \boldsymbol{p}\}$, let $K \subseteq N$ be the first coalition according to L. Let $f(\boldsymbol{p})$ be the first candidate in $\{x \in X \mid x \text{ is an efficient AV winner at } \boldsymbol{p}\text{, and } K = N(x, \boldsymbol{p})\}$ in the alphabetical order.

We claim that f is efficient and evaluationwise strategy-proof, and that no voter is decisive under f. Since the social outcome is chosen from the set of efficient AV winners, it is clear that f is efficient. Also, it is clear that no voter is decisive for f. (For each $i \in N$, $N \setminus \{i\}$ is decisive. Since there are $R_i, R_i' \in \mathcal{D}$ such that $r_1(R_i) \neq r_1(R_i')$, this implies that i is not decisive.)

Thus, we will show that f is evaluationwise strategy-proof. Let $\boldsymbol{p} \in \Pi_{\mathcal{D}}^n$, $i \in N$, and $p_i' \in \Pi_{\mathcal{D}}$. Assume $f(\boldsymbol{p}) \in U_i$, i.e., $i \notin N(f(\boldsymbol{p}), \boldsymbol{p})$. Our goal is to show $f(p_i', \boldsymbol{p}_{-i}) \in U_i$.

First, assume that $f(\boldsymbol{p})$ is not an AV winner at $(p_i', \boldsymbol{p}_{-i})$, i.e., $f(\boldsymbol{p}) \notin AV(p_i', \boldsymbol{p}_{-i})$. Since $f(\boldsymbol{p}) \in U_i$, $a(f(\boldsymbol{p}), (p_i', \boldsymbol{p}_{-i})) \geq a(f(\boldsymbol{p}), \boldsymbol{p})$.[18] Let $x \in AV(p_i', \boldsymbol{p}_{-i})$. Since $f(\boldsymbol{p}) \in AV(\boldsymbol{p})$, $a(f(\boldsymbol{p}), \boldsymbol{p}) \geq a(x, \boldsymbol{p})$. Then, we have $a(x, (p_i', \boldsymbol{p}_{-i})) > a(x, \boldsymbol{p})$. (If not, then $a(f(\boldsymbol{p}), (p_i', \boldsymbol{p}_{-i})) \geq a(f(\boldsymbol{p}), \boldsymbol{p}) \geq a(x, \boldsymbol{p}) \geq a(x, (p_i', \boldsymbol{p}_{-i}))$. This is a contradiction to our assumption that $f(\boldsymbol{p}) \notin AV(p_i', \boldsymbol{p}_{-i})$ and $x \in AV(p_i', \boldsymbol{p}_{-i})$.)

[18] Remember that for each $x \in X$ and each $\boldsymbol{p} \in \Pi^n$, $a(x, \boldsymbol{p})$ is the number of voters who approve x at $\boldsymbol{p}$.

This is possible only if $x \in U_i$. Since x was arbitrary in $AV(p'_i, \boldsymbol{p}_{-i})$, we have $AV(p'_i, \boldsymbol{p}_{-i}) \subseteq U_i$. Since $f(p'_i, \boldsymbol{p}_{-i}) \in AV(p'_i, \boldsymbol{p}_{-i})$, it follows that $f(p'_i, \boldsymbol{p}_{-i}) \in U_i$. Thus, when $f(\boldsymbol{p}) \notin AV(p'_i, \boldsymbol{p}_{-i})$, there is no incentive to misreport.

In the following, we assume $f(\boldsymbol{p}) \in AV(p'_i, \boldsymbol{p}_{-i})$.

CASE 1: $AV(\boldsymbol{p}) \cap A_i = \emptyset$.

In this case, $AV(\boldsymbol{p}) \subseteq U_i$. We want to show $AV(p'_i, \boldsymbol{p}_{-i}) \subseteq U_i$. Let $x \in AV(p'_i, \boldsymbol{p}_{-i})$. If $x \in AV(\boldsymbol{p})$, then $AV(\boldsymbol{p}) \subseteq U_i$ implies $x \in U_i$. Thus, assume $x \notin AV(\boldsymbol{p})$. Let $y \in AV(\boldsymbol{p})$. Then, $a(y, \boldsymbol{p}) > a(x, \boldsymbol{p})$. Since $AV(\boldsymbol{p}) \subseteq U_i$, we have $y \in U_i$. Thus, $a(y, (p'_i, \boldsymbol{p}_{-i})) \geq a(y, \boldsymbol{p}) > a(x, \boldsymbol{p})$. Since $x \in AV(p'_i, \boldsymbol{p}_{-i})$, $a(x, (p'_i, \boldsymbol{p}_{-i})) \geq a(y, (p'_i, \boldsymbol{p}_{-i}))$. Thus, we have $a(x, (p'_i, \boldsymbol{p}_{-i})) > a(x, \boldsymbol{p})$. This is possible only when $x \in U_i$. Since x was arbitrary in $AV(p'_i, \boldsymbol{p}_{-i})$, we have $AV(p'_i, \boldsymbol{p}_{-i}) \subseteq U_i$. Since $f(p'_i, \boldsymbol{p}_{-i}) \in AV(p'_i, \boldsymbol{p}_{-i})$, it follows that $f(p'_i, \boldsymbol{p}_{-i}) \in U_i$.

CASE 2: $AV(\boldsymbol{p}) \cap A_i \neq \emptyset$.

For our purpose, we want to show $f(p'_i, \boldsymbol{p}_{-i}) \notin A_i$. Then, it suffices to show $f(p'_i, \boldsymbol{p}_{-i}) \notin (AV(\boldsymbol{p}) \cap A_i)$. (To see this, let $y \in (A_i \setminus AV(\boldsymbol{p}))$. Then, $a(y, (p'_i, \boldsymbol{p}_{-i})) \leq a(y, \boldsymbol{p}) < a(f(\boldsymbol{p}), \boldsymbol{p}) \leq a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))$. Thus, $y \notin AV(p'_i, \boldsymbol{p}_{-i})$, and hence $f(p'_i, \boldsymbol{p}_{-i}) \neq y$.) Let $x \in (AV(\boldsymbol{p}) \cap A_i)$, and examine whether $x = f(p'_i, \boldsymbol{p}_{-i})$ is possible.

SUBCASE 2.1: x is not efficient at $\boldsymbol{p}$.

At $\boldsymbol{p}$, there is $y \in X \setminus \{x\}$ such that $y \ R_j \ x$ for each $j \in N$. Since $x \in A_i$, $y \in A_i$. Without loss of generality, assume that y is efficient at $\boldsymbol{p}$. Then, $a(y, \boldsymbol{p}) \geq a(x, \boldsymbol{p})$. Since $x \in AV(\boldsymbol{p})$, $a(y, \boldsymbol{p}) \leq a(x, \boldsymbol{p})$. Thus, $a(y, \boldsymbol{p}) = a(x, \boldsymbol{p})$. Since $y \ R_j \ x$ for each $j \in N$, $N(y, \boldsymbol{p}) = N(x, \boldsymbol{p})$. Since $y \in A_i$ and $f(\boldsymbol{p}) \in U_i$, $y \neq f(\boldsymbol{p})$. Then y is an efficient AV winner at $\boldsymbol{p}$, but it is not chosen at $\boldsymbol{p}$. Since $i \notin N(f(\boldsymbol{p}), \boldsymbol{p})$ and $i \in N(y, \boldsymbol{p})$, $N(f(\boldsymbol{p}), \boldsymbol{p}) \neq N(y, \boldsymbol{p})$. Thus, $N(f(\boldsymbol{p}), \boldsymbol{p}) \ L \ N(y, \boldsymbol{p}) = N(x, \boldsymbol{p})$. (If $N(y, \boldsymbol{p}) \ L \ N(f(\boldsymbol{p}), \boldsymbol{p})$, then $f(\boldsymbol{p})$ cannot be chosen at $\boldsymbol{p}$, which is a contradiction.)

If x is not an efficient AV winner at $(p'_i, \boldsymbol{p}_{-i})$, then $f(p'_i, \boldsymbol{p}_{-i}) \neq x$. Thus, assume x is an efficient AV winner at $(p'_i, \boldsymbol{p}_{-i})$. For x to be an AV winner at $(p'_i, \boldsymbol{p}_{-i})$, it should be $x \in A'_i$ and $f(\boldsymbol{p}) \in U'_i$. (If $x \notin A'_i$, then $a(x, (p'_i, \boldsymbol{p})) < a(x, \boldsymbol{p}) = a(f(\boldsymbol{p}), \boldsymbol{p}) \leq a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))$. This is a contradiction to $x \in AV(p'_i, \boldsymbol{p}_{-i})$. If $f(\boldsymbol{p}) \notin U'_i$, then $a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i})) > a(f(\boldsymbol{p}), \boldsymbol{p}) = a(x, \boldsymbol{p}) = a(x, (p'_i, \boldsymbol{p}_{-i}))$. This is a contradiction to $x \in AV(p'_i, \boldsymbol{p}_{-i})$.) Therefore, $N(x, (p'_i, \boldsymbol{p}_{-i})) = N(x, \boldsymbol{p})$ and $N(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i})) = N(f(\boldsymbol{p}), \boldsymbol{p})$.

If $f(\boldsymbol{p})$ is an efficient AV winner at $(p'_i, \boldsymbol{p}_{-i})$, then $x \neq f(p'_i, \boldsymbol{p}_{-i})$. (Remember that $N(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))\ L\ N(x, (p'_i, \boldsymbol{p}_{-i}))$.) Since $f(\boldsymbol{p})$ is assumed to be in $AV(p'_i, \boldsymbol{p}_{-i})$, the only chance for x to be chosen at $(p'_i, \boldsymbol{p}_{-i})$ is that $f(\boldsymbol{p})$ is not efficient at $(p'_i, \boldsymbol{p}_{-i})$. Thus, assume that $f(\boldsymbol{p})$ is not efficient at $(p'_i, \boldsymbol{p}_{-i})$. In this case, there is $z \in (X \setminus \{f(\boldsymbol{p})\})$ such that at $(p'_i, \boldsymbol{p}_{-i})$, $z\ R_j\ f(\boldsymbol{p})$ for each $j \in N \setminus \{i\}$ and $z\ R'_i\ f(\boldsymbol{p})$. Without loss of generality, assume that z is efficient at $(p'_i, \boldsymbol{p}_{-i})$.

Since $f(\boldsymbol{p}) \in AV(p'_i, \boldsymbol{p}_{-i})$, $z \in AV(p'_i, \boldsymbol{p}_{-i})$. Thus, z is an efficient AV winner at $(p'_i, \boldsymbol{p}_{-i})$. The fact that $f(\boldsymbol{p})$ is efficient at $\boldsymbol{p}$, but not efficient at $(p'_i, \boldsymbol{p}_{-i})$ implies $f(\boldsymbol{p})\ R_i\ z$ and $z\ R'_i\ f(\boldsymbol{p})$, and for each $j \in N \setminus \{i\}$, $z\ R_j\ f(\boldsymbol{p})$. Since $f(\boldsymbol{p})\ R_i\ z$ and $f(\boldsymbol{p}) \in U_i$ imply $z \in U_i$, $N(z, (p'_i, \boldsymbol{p}_{-i})) \supseteq N(z, \boldsymbol{p})$. Since $f(\boldsymbol{p})$ is an AV winner at $\boldsymbol{p}$, each voter approves $f(\boldsymbol{p})$ if and only if he approves z at $\boldsymbol{p}$. Thus, $N(z, \boldsymbol{p}) = N(f(\boldsymbol{p}), \boldsymbol{p})$. Since $i \notin N(f(\boldsymbol{p}), \boldsymbol{p})$ and $i \in N(x, \boldsymbol{p})$, $x \neq z$. If $N(z, (p'_i, \boldsymbol{p}_{-i})) \supsetneq N(z, \boldsymbol{p})$, then x cannot be an AV winner at $(p'_i, \boldsymbol{p}_{-i})$, which is a contradiction to our assumption. Thus, $N(z, (p'_i, \boldsymbol{p}_{-i})) = N(z, \boldsymbol{p}) = N(f(\boldsymbol{p}), \boldsymbol{p})$. Since $N(f(\boldsymbol{p}), \boldsymbol{p})\ L\ N(x, \boldsymbol{p})$, we have $N(z, (p'_i, \boldsymbol{p}_{-i}))\ L\ N(x, (p'_i, \boldsymbol{p}_{-i}))$. Thus, $x \neq f(p'_i, \boldsymbol{p}_{-i})$.

SUBCASE 2.2: x is efficient at $\boldsymbol{p}$.

Since x is chosen from $AV(\boldsymbol{p})$, x is an efficient AV winner. Since $i \notin N(f(\boldsymbol{p}), \boldsymbol{p})$ and $i \in N(x, \boldsymbol{p})$, $N(f(\boldsymbol{p}), \boldsymbol{p}) \neq N(x, \boldsymbol{p})$. Thus, $x \neq f(\boldsymbol{p})$ implies $N(f(\boldsymbol{p}), \boldsymbol{p})\ L\ N(x, \boldsymbol{p})$.

Then, the arguments from the second paragraph in Subcase 2.1 show $x \neq f(p'_i, \boldsymbol{p}_{-i})$. ■

Note that the rule constructed in the proof of Theorem 3 is nearly anonymous in the sense that for each $\boldsymbol{p}$ such that only one candidate is an efficient AV winner at $\boldsymbol{p}$, renaming voters does not change the social choice.

The following result says that we also have a possibility when we weaken efficiency to unanimity in Theorem 2.

Theorem 4

Let $n \geq 2$. For each $\mathcal{D} \subseteq \mathcal{L}$, there exists a rule on $\Pi^n_{\mathcal{D}}$ which is anonymous, unanimous, and evaluationwise strategy-proof.

Proof. Let f be a rule on $\Pi^n_{\mathcal{D}}$ defined as follows: For each $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$, if there is $x \in X$ such that $r_1(R_i) = x$ for each $i \in N$, then $f(\boldsymbol{p}) = x$. Otherwise, let $f(\boldsymbol{p})$ be the candidate in $AV(\boldsymbol{p})$ who is the first according to the alphabetical order.

It is clear that f is anonymous and unanimous. We will show that f is evaluationwise strategy-proof. Let $\boldsymbol{p} \in \Pi^n_{\mathcal{D}}$ and $i \in N$. Assume $f(\boldsymbol{p}) \in U_i$. Let $x \in A_i$, and examine whether it is possible to have $f(p'_i, \boldsymbol{p}_{-i}) = x$ for some $p'_i \in \Pi_{\mathcal{D}}$.

By definition, $f(\boldsymbol{p})$ is approved by the largest number of voters. (However, since $f(\boldsymbol{p}) \in U_i$, $a(f(\boldsymbol{p}), \boldsymbol{p}) < n$.) Moreover, if there are other such candidates, i.e., $|AV(\boldsymbol{p})| \geq 2$, $f(\boldsymbol{p})$ is the first one among $AV(\boldsymbol{p})$ according to the alphabetical order. Since $x \in A_i$ and $f(\boldsymbol{p}) \in U_i$, $a(x, (p'_i, \boldsymbol{p}_{-i})) \leq a(x, \boldsymbol{p})$ and $a(f(\boldsymbol{p}), \boldsymbol{p}) \leq a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))$. Since $x \neq f(\boldsymbol{p})$, either $a(x, \boldsymbol{p}) < a(f(\boldsymbol{p}), \boldsymbol{p})$, or $a(x, \boldsymbol{p}) = a(f(\boldsymbol{p}), \boldsymbol{p}) < n$ and x comes alphabetically after $f(\boldsymbol{p})$. In the former case, $a(x, (p'_i, \boldsymbol{p}_{-i})) < a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))$, and hence $x \neq f(p'_i, \boldsymbol{p}_{-i})$. In the latter case, $a(x, (p'_i, \boldsymbol{p}_{-i})) < n$, and $a(x, (p'_i, \boldsymbol{p}_{-i})) \leq a(f(\boldsymbol{p}), (p'_i, \boldsymbol{p}_{-i}))$ and x comes alphabetically after $f(\boldsymbol{p})$, and hence $x \neq f(p'_i, \boldsymbol{p}_{-i})$. Thus, f is evaluationwise strategy-proof. ■

4 Concluding remarks

We propose evaluationwise strategy-proofness as a new notion of strategy-proofness in a collective choice framework that combines rankings with evaluations. Our setting, compared to the standard Gibbard–Satterthwaite framework, is richer in inputs. On the other hand, evaluationwise strategy-proofness is more restricted in incentives for manipulation. These opposing effects lead to a mixed picture regarding the existence of non-manipulable rules. On one hand, for an even number of voters, we have an impossibility result in finding evaluationwise strategy-proof rules which are anonymous and efficient.[19] On the other hand, we are able to present results which are much more permissive than those of the Gibbard–Satterthwaite setting where strategy-proofness is incompatible with the conjunction of surjectivity and non-dictatoriality. In fact, the impossibility we establish vanishes when anonymity is slightly compromised or efficiency is weakened to unanimity.

Our possibility results are established through AV refinements, which is not surprising: the analysis of evaluationwise strategy-proofness in the preference-approval framework is closely connected to the analysis of strategy-proofness over the domain of dichotomous preferences where AV is known to have good strategic properties.[20] As discussed in Section 2.4, there are natural channels between the two settings which allow certain results to be translated from one to the other. In particular, results that establish the existence of anonymous and strategy-proof rules over dichotomous domains would imply the existence of anonymous and evaluationwise strategy-proof rules in the preference-approval framework. Surprisingly, a complete picture of strategy-proof social choice rules over dichotomous domains seems to be missing.

We close by restating our conviction on the existence of real-life cases where

[19] Whether this impossibility holds for an odd number of voters remains an open question.

[20] See, for example, the survey papers by Ju (2010) and Xu (2010).

voter behavior would correspond to the behavioral assumptions underlying evaluationwise strategy-proofness. Testing to which extent this conviction is valid presents interesting directions of computational and experimental research.

References

Aleskerov, F., Yakuba, V., and Yuzbashev, D. (2007) A 'threshold aggregation' of three-graded rankings. Mathematical Social Sciences 53, 106–110.

Balinski, M. and Laraki, R. (2011) *Majority Judgment: Measuring, Ranking and Electing*. The MIT press, Cambridge.

Barberà, S., Berga, D., and Moreno, B. (2012) Two necessary conditions for strategy-proofness: On what domains are they also sufficient? Games and Economic Behavior 75, 490–509.

Barberà, S., Dutta, B., and Sen, A. (2001) Strategy-proof social choice correspondences. Journal of Economic Theory 101, 374–394.

Benoît, J.-P. (2002) Strategic manipulation in voting games when lotteries and ties are permitted. Journal of Economic Theory 102, 421–436.

Bogomolnaia, A., Moulin, H., and Stong, R. (2005) Collective choice under dichotomous preferences. Journal of Economic Theory, 122, 165-184.

Brams, S.J. and Fishburn, P.C. (1978) Approval Voting. American Political Science Review, 72, 831-847.

Brams, S.J. and Fishburn, P.C. (1985) Comment on The problem of strategic voting under approval voting (Vol. 78, December 1984, pp. 952-958). American Political Science Review 79, 816–818.

Brams, S.J. and Fishburn, P.C. (2007) *Approval Voting*, 2nd edn. Springer, New York.

Brams, S.J., Fishburn, P.C., and Merrill, S. (1988a) The responsiveness of Approval Voting: Comments on Saari and van Newenhizen. Public Choice 59, 121–131.

Brams, S.J., Fishburn, P.C., and Merrill, S. (1988b) Rejoinder to Saari and Van Newenhizen. Public Choice 59, 149.

Brams, S.J. and Sanver, M.R. (2006) Critical strategies under approval voting: Who gets ruled in and ruled out. Electoral Studies 25, 287–305.

Brams, S.J. and Sanver, M.R. (2009) Voting systems that combine approval and preference, in: *The Mathematics of Preference, Choice and Order: Essays in Honour of Peter C. Fishburn*, Studies in Choice and Welfare (eds. S.J. Brams, W.V. Gherlein, F.S. Roberts), Springer, Berlin, pp. 215–237.

Campbell, D.E. and Kelly, J.S. (2009) Gains from manipulating social choice rules. Economic Theory 40, 349–371.

Gibbard, A. (1973) Manipulation of voting scheme: A general result. Econometrica 41, 587–601.

Hillinger, C. (2005) The case for utilitarian voting. Homo Oeconomicus 23, 295–321.

Ju, B. G. (2010) Collective choice for simple preferences. In Handbook on Approval Voting (pp. 41–90). Springer Berlin Heidelberg.

Laslier, J.-F. and Sanver, M.R. (2010) *Handbook on Approval Voting*. Springer Berlin Heidelberg.

Maniquet, F. and Mongin, P. (2015) Approval voting and Arrow's impossibility theorem. Social Choice and Welfare 44, 519–532.

Muto, N. and Sato, S. (2016) An impossibility under individual bounded response of social choice functions. Unpublished manuscript.

Niemi, R.G. (1984) The problem of strategic behavior under Approval Voting. American Political Science Review 78, 952–958.

Niemi, R.G. (1985) Reply to Brams and Fishburn. American Political Science Review 79, 818–819.

Pramanik, Anup (2015) Further results on dictatorial domains. Social Choice and Welfare 45(2), 379–398.

Reffgen, A. (2011) Generalizing the Gibbard–Satterthwaite theorem: partial preferences, the degree of manipulation, and multi-valuedness. Social Choice and Welfare 37(1), 39–59.

Saari, D.G. and Newenhizen, J.V. (1988a) The problem of indeterminacy in approval, multiple, and truncated voting systems. Public Choice 59, 101–120.

Saari, D.G. and Newenhizen, J.V. (1988b) Is Approval Voting an unmitigated evil: A response to Brams, Fishburn and Merrill. Public Choice 59, 133–147.

Sakai, T. and Shimoji, M. (2006) Dichotomous preferences and the possibility of Arrovian social choice. Social Choice and Welfare 26, 435–445.

Sanver, M.R. (2010) Approval as an intrinsic part of preference. In Handbook on Approval Voting (pp. 469–481). Springer Berlin Heidelberg.

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.

Sato, S. (2010) Circular domains. Review of Economic Design 14, 331–342.

Sato, S. (2013) A sufficient condition for the equivalence of strategy-proofness and nonmanipulability by preferences adjacent to the sincere one. Journal of Economic Theory 148, 259–278.

Smith, W.D. (2000) Range voting. `http://scorevoting.net/WarrenSmithPages/homepage/rangevote.pdf`.

Vorsatz, M. (2007) Approval voting on dichotomous preferences. Social Choice and Welfare, 28, 127-141.

Xu, Y. (2010) Axiomatizations of approval voting. In Handbook on Approval Voting (pp. 91–102). Springer Berlin Heidelberg.