WINPEC Working Paper Series No.E2110
June 2021

(The Impossibility of)

Deliberation-Consistent Social Choice

Tsuyoshi Adachi and Hun Chung and Takashi Kurihara

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

(The Impossibility of) Deliberation-Consistent Social Choice

Tsuyoshi Adachi[1] Hun Chung[1] Takashi Kurihara[2]

[1]Faculty of Political Science and Economics, Waseda University
[2]School of Political Science and Economics, Tokai University

June 16, 2021

Abstract There is now a growing consensus among democratic theorists that we should incorporate both 'democratic deliberation' and 'aggregative voting' into our democratic processes, where democratic deliberation precedes aggregating people's votes. But how should the two democratic mechanisms of deliberation and voting interact? The question we wish to ask in this paper is which social choice rules are consistent with successful deliberation once it has occurred. For this purpose, we introduce a new axiom, which we call "Non-Negative Response toward Successful Deliberation (NNRD)." The basic idea is that if some individuals change their preferences toward other individuals' preferences through successful deliberation, then the social choice rule should not make everybody who has successfully persuaded others through reasoned deliberation worse-off than what s/he would have achieved without deliberation. We prove an impossibility theorem that shows that there exists no aggregation rule that can simultaneously satisfy (NNRD) along with other mild axioms that reflect deliberative democracy's core commitment to unanimous consensus and democratic equality. We offer potential escape routes: however, it is shown that each escape route can succeed only by compromising some core value of deliberative democracy.

Keywords Social Choice Theory; Deliberative Democracy; Deliberation; Aggregation; NNRD

1 Introduction

It is now widely endorsed by contemporary democratic theorists that the sole reliance on aggregative voting procedures is insufficient to lend full democratic legitimacy or democratic justification of its resulting outcomes. The numerous negative results of modern social choice theory have repeatedly shown that virtually all democratic voting procedures tend to (at least sometimes) generate arbitrary and unstable voting outcomes that make them susceptible to strategic manipulation and agenda control. (Arrow 1951/1963; Plott 1967; McKelvey 1976, 1979; Schofield 1978; Riker 1982) This, according to William Riker, implies that all voting outcomes are "uninterpretable and meaningless" (Riker 1982: 237) and thus cannot be regarded as "fair and true amalgamations of voters' judgments." (Riker 1982: 238) After what scholars have called the "deliberative turn" (Dryzek 2000: 1), democratic theorists have instead sought to ground democratic legitimacy on a process of democratic deliberation by insisting that "outcomes are democratically legitimate if and only if they could be the object of a free and reasoned agreement among equals." (Cohen 1997a: 73) The underlying thought is an account of political justification that requires "that a person must be given a reason that is acceptable by his or her own lights for a policy in order for that policy not to be oppressive." (Christiano 1997: 272) In this way, deliberative democracy "puts public reasoning at the center of political justification." (Cohen 1997b: 413)

Then, what exactly is the normative role of democratic deliberation and how does it facilitate the achievement of democratic legitimacy? Many people believe that one key function of democratic deliberation lies in its potential to change or transform people's preferences. According to Amy Gutmann and Dennis Thomson, "By engaging in deliberation, citizens acknowledge the possibility that they may change their preferences. ⋯ The very nature of the deliberative process of justification sends a signal that its participants are willing to enter into a dialogue in which the reasons given, and the reasons responded to, have the capacity to change minds." (Gutmann and Thomson 2004: 20) By deliberating with others, participants of democratic deliberation may change and/or transform their preferences by acquiring new factual information, detecting logical mistakes in their previous reasoning, seeing the issue from other participants' perspectives, and also by forming a newly developed "commitment to justice, which now overrides or modifies the self-interested perspective with which they entered the deliberation." (Mansbridge et al. 2010: 79) In these and other ways, the point is that "[p]ublic deliberation transforms, modifies, and clarifies the beliefs and preferences of the citizens of a political society." (Christiano 1997: 244) As Jane Mansbridge and her co-authors

explain, "[d]eliberation would have no point if it did not produce change in the views of at least some participants." (Mansbridge et al. 2010: 78) In such cases, "[w]hether or not deliberation is desirable, it would be futile." (Mackie 2006: 299)

Furthermore, we do not expect democratic deliberation to merely change people's preferences; we expect democratic deliberation to change people's preferences into a particular direction. If democratic deliberation changed people's preferences randomly, there would be no point in incorporating democratic deliberation into a society's collective decision-making process as we would have no reason to think that the social choices reached through people's post-deliberation preferences would be better than those reached through people's pre-deliberation preferences. The reason that we think that incorporating deliberative institutions into our democratic process is both meaningful and valuable is that, after deliberation, we expect people's preferences to be enlightened so that they become more logical, better-structured, consistent with known facts, and better grounded in reasons and arguments. Moreover, as people's preferences change on the basis of mutual persuasion through the exchange of reasoned arguments with other participants in deliberation, we would normally expect people's preferences to become closer (if not totally converge) to one another (rather than farther apart) after deliberation than what they were prior to deliberation.

Although some critics have pointed out that "[d]iscussion only rarely eliminates differences of opinion on matters of politics" (Christiano 1997: 264) and worried that democratic deliberation may actually "[produce] more disagreement and diversity of opinion" (ibid.) and even lead certain groups (that are engaged in "enclave deliberation") to shift toward more extreme positions that exacerbates "group polarizations" that may potentially cause "danger to social stability" (Sunstein 2002: 176-177), we follow other deliberative democratic theorists and regard this as rather a case of unsuccessful deliberation rather than the norm, which can be not too difficultly avoided. For instance, according to Amy Gutmann and Dennis Thomson, if we make sure that deliberation occurs prior to voting, is inclusive and "large enough to represent random samples rather than skewed samples of opinions, have moderators who oversee the deliberations to ensure that all perspectives receive a fair hearing, enlist experts to answer questions and clarify matters of fact, and have extensive information available to all participants ahead of time," the participants of democratic deliberation will "tend not to polarize but rather to find greater common ground than they had before." (Gutmann and Thomson 2004: 54) This is consistent with what Cass Sunstein finds in James Fishkin's "deliberative opinion polls" (Fishkin 1995: 206-207), where "the existence of monitors, an absence of a group decision, the great heterogeneity of the people," together with having access to

"a set of written materials that attempted to be balanced and that contained detailed arguments on both sides" helped prevent deliberation from having a polarizing effect. (Sunstein 2002: 194-195)

Then, how does such a process of changing or transforming people's preferences through democratic deliberation help overcome the alleged pitfalls of aggregative voting mechanisms and help us achieve democracy legitimacy? Early proponents of deliberative democracy, such as Jürgen Habermas (1990), Joshua Cohen (1997a), and Jon Elster (1997) thought that once we introduce democratic deliberation into the democratic process "there would not be any need for an aggregating mechanism, since rational discussion would tend to produce unanimous preferences." (Elster 1997: 11) The basic thought was that once we require people to engage in democratic deliberation, they will tend to "go beyond private self-interests of the "market" and orient themselves to public interests of the "forum"" (Bohman and Rehg 1997: xiv), and "[w]hen the private and idiosyncratic wants have been shaped and purged in public discussion about the public good, uniquely determined rational desires would emerge." (Elster 1997: 11) If this is correct, then democratic deliberation would be able to restore democratic legitimacy by completely replacing aggregating voting mechanisms with rationally motivated unanimous consensus.

Many scholars have pointed out that unanimous agreement is seldom achievable even under ideal circumstances, especially, in modern pluralistic democracies, whose basic characteristics, according to John Rawls, is "the fact of reasonable pluralism – the fact that a plurality of conflicting reasonable comprehensive doctrines, religious, philosophical and moral, is the normal result of its culture of free institutions." (Rawls 1993/2005: 441) Of course, one important regulative ideal of deliberative democracy is that "[d]eliberation is reasoned in that the parties to it are required to state their reasons for advancing proposals, supporting them, or criticizing them ⋯ with the expectation that those reasons (and not, for example, their power) will settle the fate of their proposal." (Cohen 1997a: 74) However, Gerald Gaus explains that in modern pluralistic democracies, "[s]incere reasoners will find themselves in principled disagreements," and "[b]ecause this is so, we will inevitably have competing judgments about what is public justified." (Gaus 1997: 231) Gerry Mackie has also argued that unanimous consensus, even if practically achievable, might not be so desirable because a deliberative environment in which unanimous consensus is easily achieved will be "a system with extreme confirmation bias: unless it is overwhelming, contradictory new evidence will be rejected as false, even if it is true." (Mackie 2006: 283) As a result, "a unanimity rule unjustifiably enshrines [the] status quo." (Mackie 2018: 225) Hence, "[i]n a pluralistic world," claims John Dryzek,

"consensus is [not only] unattainable, [but also] unnecessary and undesirable." (Dryzek 2000: 170)

If it is neither practically feasible (nor desirable) to aim at deliberative democracy's regulative ideal of unanimous consensus, we would have to eventually rely on some form of aggregative voting mechanism to reach a democratic decision. Even Joshua Cohen, who thinks that unanimous consensus is the ultimate aim of ideal deliberation acknowledges that "[e]ven under ideal conditions there is no promise that consensual reasons will be forthcoming" and "[i]f they are not, then deliberation concludes with voting, subject to some form of majority rule." (Cohen 1997a: 75) However, if it is true that we cannot completely dispense with aggregative voting mechanisms, with all its alleged imperfections and defects, even with the introduction of ideal democratic deliberation, what would the point of introducing deliberation into the democratic process be in the first place?

Many people have argued that even if democratic deliberation may seldom lead to full unanimity, it may still help us better achieve democratic legitimacy by making it possible for aggregative voting mechanisms to avoid many of its alleged pitfalls and shortcomings. For instance, it has been pointed out that the problems of instability and cycling of aggregative voting mechanisms demonstrated in social choice theory usually occur when there is more than one dimension of conflict. (Plott 1967; McKelvey 1976, 1979; Schofield 1978) To this, a number of scholars have argued that democratic deliberation may significantly reduce the possibilities of instability and cycling by inducing its participants to arrive at, if not unanimous consensus toward a specific social outcome, a shared understanding regarding the single, underlying, dimension of political conflict. (Miller 1992; Knight and Johnson 1994, 2007) Whenever the participants' disagreements are reduced to a disagreement along a single, shared policy/issue dimension in this way, this will facilitate the participants to restructure their preferences to become single-peaked, (which informally means that each participant has an ideal policy point located somewhere along the single issue/policy dimension and that his/her preferences over various social alternatives decreases as they move farther way from his/her ideal policy point), which makes it possible for us to avoid majority voting cycles when using pairwise majority vote as the aggregation voting rule. (Black 1958; Miller 1992; Dryzek and List 2003; List et al. 2013) This particular way of avoiding the problem of voting cycles through single-peaked preferences can be seen as a case of domain restriction that relaxes one of the conditions (namely, universal preference domain) that Arrow relies on to derive his famous impossibility result. According to John Dryzek and Christian List, "[e]ach condition of Arrow's theorem and of the Gibbard-Satterthwaite theorem

points towards a potential escape-route from the impossibility theorems" because "[i]f any one of these conditions is relaxed, there exist social choice procedures satisfying all the others, and such procedures can, in principle, be employed in democratic decision making." (Dryzek and List 2003: 7) An important value of democratic deliberation, then, even when it fails to achieve unanimous consensus, is that "[d]eliberation facilitates pursuits of several escape-routes from the impossibility results commonly invoked by social choice-theoretic critics of democracy." (Dryzek and List 2003: 27)

So, although we cannot practically expect democratic deliberation to achieve full unanimous consensus, we can at least reasonably expect democratic deliberation to sufficiently improve and restructure people's preferences so that, later, their aggregation will lead to stable, non-arbitrary, and meaningful democratic decisions in the voting stage. In short, ideal democratic deliberation, when successful, enlightens and changes people's preferences and makes them one step closer toward reasoned consensus so that their aggregation leads to better-informed, rational social outcomes than what the aggregation of their pre-enlightened unconsidered preferences would have generated.

There is now a growing consensus among democratic theorists that 'deliberation' and 'aggregation (or voting)' have their own respective virtues and that each plays an important role in the democratic process that cannot be properly reduced to the role performed by the other. According to Robert Goodin, although democratic deliberation is excellent as a 'discovery procedure' that may inform the participants about what may constitute the best alternative, it is not particularly a good 'decision procedure' due to its inherent path dependency. (Goodin 2008: 111; Knight and Johnson 1997: 291; see also Chung and Duggan 2020: 21-23) Conversely, we might say that although aggregation is excellent at generating a final decision even "when interests conflict irreconcilably, negation to agreement is impossible, or an assembly simply runs out of time" (Mansbridge et al. 2010: 85), the decision so arrived will in many cases lack democratic significance and lead us astray without being properly informed by a prior stage of reasoned deliberation. According to Gerry Mackie, "voting and discussion are complements, not substitutes."(Macike 2003: 107) Many people now think that in order to achieve democratic legitimacy/justification of the resulting outcomes, democratic institutions should incorporate both 'democratic deliberation' and 'aggregative voting' into its process, where democratic deliberation precedes aggregating people's votes. Robert Goodin neatly summarizes this into the following slogan: 'first talk, then vote.' (Goodin 2008: 124)

However, "[d]eliberative democrats have often downplayed the virtues and even anathematized the aims and mechanisms of voting." (Mansbridge et al. 2010 84) As Adam

Przeworski points out, early "deliberation theorists ⋯ [have] wish[ed] away the vulgar fact that under democracy ends in voting." (Przeworski 1998: 141) To many deliberative democratic theorists, the second aggregation stage after deliberation is still regarded as an addendum or a necessary evil, which is included into the democratic process to merely arrive at a final decision when deliberation fails to deliver full unanimity. In this way, "deliberative democrats concede the pragmatic point [of having a second aggregation stage of voting], typically in a spirit of resigned acceptance." (Goodin 2008: 109) As a result, most of the previous philosophical literature on deliberative democracy have focused almost exclusively on investigating the deliberation stage and characterizing the set of ideal conditions [1] that define deliberative democracy as a "regulative ideal" (Mansbridge et al. 2010: 65) as well as "the ideal deliberative procedure ⋯ [that is] meant to provide a model for institutions to mirror" (Cohen 1997a: 73) rather than focusing on the proper normative relationship between the prior deliberation stage and the post aggregation stage.

However, we should remember that "[t]he process of voting is integrated with deliberation, and not just complementary to it, when the deliberation structures the voting, for example by ruling out options, creating single-peaked (or other) preference orderings, or, on a more macro level, choosing the form of voting itself. The expectation of voting also structures deliberation, for example by forcing the choices into a simple yes-or-no vote." (Mansbridge et al. 2010: 88-89) It has been shown in the game-theoretical literature on deliberation that how likely the participants would reveal their private information truthfully during the deliberation stage crucially depends on the post-deliberation voting rule. (Austen-Smith and Feddersen 2006; Coughlan 2000; Mathis 2011) It has also been shown through real-world case studies that the post-deliberation voting rule crucially affects the substantive contents discussed in the pre-aggregation deliberation stage as well as whether or not the participants will engage in deliberation with a strategic motive to manipulate the outcome. (Mackie 2018: 226-229) Hence, "[a]ny "systems" approach to deliberation should take into account not only how different kinds of deliberative forums contribute to or detract from the broader patterns of deliberation in the system but also how other non-deliberative mechanisms, particularly voting, can affect public de-

[1] The conditions that define deliberative democracy as a regulative ideal include the requirements that deliberation be "open to all those affected by the decision"; the participants should have "equal opportunity to influence the process, have equal resources, and be protected by basic rights"; the participants should "treat one another with mutual respect and equal concern" and "speak truthfully"; the participants should "listen to one another and give reasons to one another that they think the other can comprehend and accept"; the participants should "aim at finding fair terms of cooperation among free and equal persons"; and "coercive power should be absent." (Mansbridge et al. 2010: 65-66)

liberation in many venues. The two democratic mechanisms of voting and deliberation interact." (Mackie 2018: 229)

But how should the two democratic mechanisms of voting and deliberation interact? Suppose we have achieved highly successful deliberation, and, as a result, people's preferences are changed and transformed for the better in the specific way prescribed by our best normative theories of deliberative democracy. Even so, if we want successful deliberation to be fully translated into better social outcomes, we should not be complacent by the mere fact that our aggregative voting rule can now, by the restructuring of people's preferences, avoid, say, voting cycles. Rather, we should go beyond and require that our aggregative voting rule positively generates better social outcomes that fully respects and accommodates the direction of the preference changes and transformations that have occurred after successful deliberation.

To our best knowledge, no prior work has specifically investigated this task; that is, no prior work has examined what social choice rules can properly accommodate the effects of successful deliberation once it occurs. The question we wish to ask in this paper is which social choice rules are consistent with successful deliberation. The reason that answering this question is fundamental to the success of deliberative democracy is that if it turns out that no social choice function or aggregation rule is compatible with successfully performed ideal deliberation, then the whole purpose of introducing deliberation into the democratic process becomes futile. This would be so even if it can be shown that, with democratic deliberation, we can avoid voting cycles and other problems of aggregative voting rules identified by social choice theory. This is because the normative aims of deliberative democracy is not and should not be confined to merely finding some convenient escape routes that allow us to circumvent the many impossibility results raised by social choice theory; rather, the true success of deliberative democracy requires that the results of successful deliberation in the first deliberative stage be fully incorporated and translated into better social outcomes generated by the second aggregation stage. We require this as one important regulative ideal of deliberation-combined social choice.

2 Motivating Example: A Social Choice Rule that Fails to Respect Successful Deliberation

Consider the following example:

Example 1: Suppose that our democratic society employs the plurality rule with a tie-breaker as its aggregation rule. Specifically, each individual votes for his/her best alternative and the alternative that receives the most number of votes gets socially chosen; however, if there is a tie, then the alternative that comes later in the alphabetical order is chosen. Suppose that there are three social altneratives x, y, and z, and three social groups – viz., Group A, Group B, and Group C – each consisting of six, five, and three members, who have the following preferences:

Group A: six individuals consider that x is the best;

Group B: five individuals consider that y is the best; and

Group C: three individuals prefer z to x to y.

Suppose that our democratic society implements the social choice rule (i.e., the plurality rule with the specified tie-breaker) without deliberation. Then, the initial social choice is: x.

Now, suppose that the members of each group engage in democratic deliberation. Suppose that after deliberation, one member of Group A is persuaded and convinced by the arguments presented by members of Group C, and now thinks that alternative z is better than alternative x, which s/he still believes to be better than alternative y. Then, as a result of successful deliberation, we have the following list of post-deliberation preferences:

Group A': five individuals consider that x is the best; and one individual prefer z to x to y.

Group B': five individuals consider that y is the best; and

Group C': three individuals prefer z to x to y.

By implementing the plurality rule, it turns out that we now have a tie for the two alternatives x and y. So, we follow our protocol and use our pre-decided tie-breaker, which requires us to choose the alternative that comes later in the alphabetical order. As a result, alternative y now becomes the new winner. But note: alternative y is worse than alternative x for members of Group C. What this means is that although the members of Group C were able to, through reasoned deliberation, persuade a member of Group A to change her preferences toward those of the members of Group C, the resulting social choice turned out to be what Group C considers to be their *worst* alternative,

y! For the members of Group C, democratic deliberation was futile; not because it was unsuccessful in persuading other individuals, but rather, because the success of democratic deliberation in persuading others was not properly reflected into the final implemented social choice. In this case, it would have been better for the members of Group C to decide not to engage in democratic deliberation at all; surely, it would be a normative failure of deliberative democracy if it provided incentives for participants to avoid democratic deliberation! Hence, this example illustrates an important normative property we wish our aggregative social choice rule to ideally incorporate:

Non-Negative Response toward (Successful) Deliberation (NNRD) If some individuals, through successful deliberation, change their preferences toward other individuals' preferences, then the aggregative voting rule (i.e., social choice function) should non-negatively respond to the preferences of those who have successfully persuaded others – specifically, the social choice rule should not make those who have successfully persuaded others through reasoned deliberation worse-off than what they would have achieved without deliberation.

We believe that it is important for a social choice rule to satisfy NNRD because if it does not, then this implies that the results of successful deliberation that occur in the first deliberative stage might not get properly translated into producing better social outcomes in the second aggregation stage, which would defeat the very purpose of introducing democratic deliberation into our democratic process. As we have just seen, the social choice rule, "use plurality rule, and when there is a tie, choose the alternative that comes later in the alphabet order" does not satisfy NNRD. The purpose of this paper is to investigate what social choice rules, if there are any, are able to satisfy NNRD.

3 The Model

3.1 Preliminaries

We consider a democratic society that employs a two stage democratic process to arrive at a social decision. In the first *deliberation stage*, all individuals are assumed to engage in democratic deliberation on the basis of which they update and change their preferences over the alternatives under consideration. In the second *aggregation stage*, our democratic society aggregates each individual's *post-deliberation* preferences to arrive at a specific social choice.

Let N be the set of individuals. Let X be the set of alternatives. We assume that both N and X are finite and have at least three members, i.e., $|N| \geq 3$ and $|X| \geq 3$.

For each individual $i \in N$, a preference order on X is denoted by $\succsim_i \in \mathcal{R}_i$, where $\mathcal{R}_i$ is the set of all weak orders on X, which are both complete and transitive. We assume $\mathcal{R}_i = \mathcal{R}_j$ for all $i, j \in N$. For each $\succsim_i \in \mathcal{R}_i$, let $\succ_i$ and $\sim_i$ denote the asymmetric and symmetric components of $\succsim_i \in \mathcal{R}_i$. Let $\succsim = (\succsim_i)_{i \in N}$ be a preference profile that lists each individual's preferences on X. Let $\mathcal{R} = \prod_{i \in N} \mathcal{R}_i$ be the set of all preference profiles. Let $\succsim_{-i} = (\succsim_j)_{j \neq i}$ and $\mathcal{R}_{-i} = \prod_{j \neq i} \mathcal{R}_j$ respectively.

As already explained, we assume that the individuals can change their preferences after engaging in democratic deliberation. For this purpose, we will use $\succsim_i^0$ to represent i's initial pre-deliberation preferences and use $\succsim_i^1$ to represent i's post-deliberation updated preferences, whenever we need to emphasize the distinction.

A social choice function (hereafter, SCF) is any function from $\mathcal{R}$ mapping to X. Here, we take $\mathcal{R}$ to denote the set of all possible preference profiles *after deliberation*. The reason that we take the domain of our SCF (i.e., $\mathcal{R}$) to be the set of all post-deliberation (as opposed to pre-deliberation) preferences is obvious. We want our SCF to reflect the many positive effects of democratic deliberation that have taken place in the prior deliberation stage and such positive effects of democratic deliberation, if there are any, will be reflected in the individuals' *post*-deliberation (as opposed to *pre*-deliberation) preferences. Note further that since we are considering a social choice 'function' (that produces a unique outcome) and not a social choice 'correspondence' (that produces a set of outcomes) as our social aggregation mechanism, analyzing standard voting rules will require us to supplement some kind of tie-breaking method to break ties whenever they occur.

We now explain how the individuals are assumed to transform or change their preferences through democratic deliberation. For this purpose, let $\mathcal{U}_i$ be the set of all utility functions on X. For any $\succsim_i \in \mathcal{R}_i$, let $\mathcal{U}_{\succsim_i} \subsetneq \mathcal{U}_i$ be the set of all utility functions representing $\succsim_i$. For any preference profile $\succsim \in \mathcal{R}$, let $\mathcal{U}_{\succsim} = \prod_{i \in N} \mathcal{U}_{\succsim_i}$ be the set of all profiles of utility functions representing the profile of individual preferences $\succsim \in \mathcal{R}$. Then, $\boldsymbol{u} = (u_i)_{i \in N} \in \mathcal{U}_{\succsim}$ will be a profile of utility functions for each individual.

For each $i \in N$, let $\mathcal{C}^i \subsetneq [0,1]^n$ be the standard simplex on N: i.e., $\mathcal{C}^i = \{\boldsymbol{c}^i = (c_1^i, ..., c_{|N|}^i) \in \mathrm{R}_+^n \mid \sum_{j \in N} c_j^i = 1\}$. Note that $\mathcal{C}^i = \mathcal{C}^j$ for each $i, j \in N$. We refer to $\boldsymbol{c}^i \in \mathcal{C}^i$ as i's *consensus vector*. Each individual's consensus vector represents the degree to which s/he agrees or consents with the opinions/preferences of the other participants. That is, an individual i's consensus vector represents how much she is willing to transform her preferences towards the preferences of others through democratic deliberation.

One thing that we have already discussed is that deliberative democracy does not think that democratic deliberation will change people's preferences randomly; rather,

deliberative democracy assumes that, once people engage in democratic deliberation, people's preferences, through the exchange of reasoned arguments, will become closer to one another even if they do not arrive at full unanimous consensus.

We model this kind of directionality of deliberation-led preference change or transformation as the *convex combination* of each individual's utility function obtained via each individual's consensus vector. Specifically, for any utility function profile $\boldsymbol{u} \in \mathcal{U}_{\succsim}$ and consensus vector $\boldsymbol{c}^i \in \mathcal{C}^i$, let us define the convex combination $\boldsymbol{c}^i\boldsymbol{u} \in \mathcal{U}_i$ of $\boldsymbol{u}$ with $\boldsymbol{c}^i$ as follows: for each $x \in X$, $\boldsymbol{c}^i\boldsymbol{u}(x) = \sum_{j \in N} c^i_j u_j(x)$. Thus, $\boldsymbol{c}^i\boldsymbol{u}$ is a new utility function obtained as the weighted sum of all individuals' utility functions with weights given by the consensus vector $\boldsymbol{c}^i$. The new utility function $\boldsymbol{c}^i\boldsymbol{u}$ so defined will represent individual i's *post*-deliberation preferences that s/he obtains after appropriately listening to the arguments presented by the other participants with whom s/he gives positive considerations and, thereby, changes her/his preferences during the first deliberation stage.

Let $\mathcal{C} = \prod_N \mathcal{C}^i$. Thus, $C = (\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^{|N|}) \in \mathcal{C}$ denotes a consensus vector *profile*. For any $C = (\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^{|N|}) \in \mathcal{C}$, any $\succsim \in \mathcal{R}$, and any $\boldsymbol{u} \in \mathcal{U}_{\succsim}$, let $C\boldsymbol{u} = (\boldsymbol{c}^1\boldsymbol{u}, \boldsymbol{c}^2\boldsymbol{u}, \ldots, \boldsymbol{c}^{|N|}\boldsymbol{u}) \in \prod_{i \in N} \mathcal{U}_i$. Thus, $C\boldsymbol{u}$ is a profile of post-deliberation utility functions that have been obtained via the convex combinations of each individual's pre-deliberation utility functions contained in the profile $\boldsymbol{u}$ and each individual's (possibly different) consensus vectors $C = (\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^{|N|})$.

Note that although we assume that the individuals come into deliberation with their pre-deliberation preferences and leave with their potentially difference post-deliberation preferences, we have described the process of preference change/transformation that occurs during deliberation (not in terms of the individuals' preference, but) in terms of the individuals' utility functions and their convex combinations. The reason for this is to precisely define the acceptable range of preference change given successful deliberation, which we discuss in the next subsection.

3.2 Acceptable Range of Post-Deliberation Preference Change Given Successful Deliberation

Again, we do not think that democratic deliberation, when successful, changes people's preferences randomly; during democratic deliberation, people's preferences change and are transformed on the basis of the opinions/preferences and arguments of the other participants, and, as a result, they become *closer* to one another than what they were prior to deliberation. This gives us reason to think that we should appropriately restrict the range of acceptable preference changes from the individuals' pre-deliberation preferences

$\succsim^0$ to the indviduals' post-deliberation preferences $\succsim^1$ in successful deliberation.[2] The following situation is an example of what we consider to be an *unacceptable* preference change given that deliberation was successful:

Example 2: Implausible Post-Deliberation Preference Change

- $X = \{x, y, z\}$; As for $\succsim^0$, $x \succ_j^0 y \succ_j^0 z$ for every individual $j \in N$; As for $\succsim^1$, $x \succ_j^1 y \succ_j^1 z$ for all but i, and $z \succ_i^1 y \succ_i^1 x$ for i.

Example 2 illustrates a situation in which *all* individuals unanimously preferred x to y to z *before* deliberation, but *after* deliberation, a single individual, i, suddenly changed his/her mind and started to prefer the alternatives in opposite order, i.e., i now prefers z to y to x! The reason that this would be an implausible preference change is that, since everybody unanimously preferred x to y to z before deliberation, there simply existed no other individual who could have possibly persuaded and convinced i to think that z (which was unanimously considered to be the worst alternative) is now the best alternative and x (which was unanimously considered to be the best alternative) is now the worst alternative! Such a preference change would be implausible under successful deliberation.

Then, what sort of preference change would be acceptable under successful deliberation? Again, let $\succsim^0$ be the profile of the individuals' pre-deliberation preferences, and let $\succsim^1$ be the profile of the individuals' post-deliberation preferences. When deliberating with other people, it is common that one is persuaded and convinced, not by every other participant, but by the reasons and arguments presented by the members of a given subset of the participants of deliberation. Suppose $S \subseteq N$ denotes the subset of the participants whose opinions and arguments individual i finds persuasive. Then, after deliberation, we can expect that individual i's post-deliberation preferences would become more aligned with the preferences of the members of subgroup S. In other words, we can think of individual i's post-deliberation preferences to be a positive combination of his/her initial preference $\succsim_i$ and the profile $(\succsim_{s^1}, \succsim_{s^2}, \ldots, \succsim_{s^m})_{s^j \in S}$. If preferences were defined on a linear space, this concept would correspond to a convex combination of the preferences with a coefficient vector on $i \cup S$. However, currently, there is no canonical way to define a linear combination on preferences. Hence, we consider instead

[2] In this subsection, we treat the symbol $\succsim^0$ as denoting the *actual* profile of initial preferences of the participants. In the subsequent parts of our paper, for instance, in the definition of axioms and proofs, $\succsim^0$ will be interpretated as a *possible* profile of initial preference, and what the actual profile of initial preference is will not affect any conditions or results.

a convex combination of utility representations as an alternate way to capture this idea of changing one's preference toward others' through deliberation.

Since the set of cardinal utility functions is a linear space, we can define a convex combination of $\{i\} \cup S$'s utility functions as: $\boldsymbol{c}^i\boldsymbol{u}$ with $\boldsymbol{c}^i \in \mathcal{C}^i$ such that $c^i_j > 0$ for $j \in S$. We say that $\succsim^1_i$ is a *convex combination preference with utility representations of* $\succsim^0$ restricted to $i \cup S$ if and only if there exists some $\boldsymbol{u}^0$ representing $\succsim^0$ and u^1_i representing $\succsim^1_i$ such that u^1_i is a convex combination of $\boldsymbol{u}^0_{\{i\}\cup S}$ obtained through some consensus vector $\boldsymbol{c}^i \in \mathcal{C}^i$. Formally, we say that $\succsim^1_i$ is a convex combination preference with utility representations of $\succsim^0$ restricted to $i \cup S$ if there exist $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$, $u^1_i \in \mathcal{U}_{\succsim^1_i}$, and $\boldsymbol{c}^i \in \mathcal{C}^i$ satisfying $u^1_i = \boldsymbol{c}^i\boldsymbol{u}^0$ and $c^i_j > 0$ for $j \in S$.

So far, for ease of explanation, we have focused on just a single individual i's preference update. However, in most deliberative environments, preference change/update is not a one-way process, but rather, a multi-way process that occurs simultaneously among multiple participants. Furthermore, during deliberation, individual $j \neq i$ may find the opinions and arguments of a different subset $S' \subseteq N$ (where $S' \neq S$) of deliberative participants to be persuasive. We model this as individual i and individual j each having possibly different consensus vectors $\boldsymbol{c}^i$ and $\boldsymbol{c}^j$, where $\boldsymbol{c}^i \neq \boldsymbol{c}^j$. Hence, although the two individuals i and j use the same initial profile of pre-deliberation utility functions $\boldsymbol{u}^0$ as the basis of their preference change during deliberation, since each individual uses different consensus vectors $\boldsymbol{c}^i$ and $\boldsymbol{c}^j$, the resulting post-deliberation utility functions, which are obtained through the convex combinations $\boldsymbol{c}^i\boldsymbol{u}^0$ and $\boldsymbol{c}^j\boldsymbol{u}^0$, are different, and, hence, represent different post-deliberation preferences of individuals i and j. Let $C = (\boldsymbol{c}^1, \ldots, \boldsymbol{c}^{|N|}) \in \mathcal{C}$ be the profile of consensus vectors of the individuals. For any utility function profile $\boldsymbol{u}^0$, $C\boldsymbol{u}^0 = (\boldsymbol{c}^1\boldsymbol{u}^0, \ldots, \boldsymbol{c}^{|N|}\boldsymbol{u}^0)$ will represent the profile of the individuals' updated post-deliberation utility functions (representing the profile of each individual's post-deliberation preferences $\succsim^1$) obtained through the convex combinations of the initial utility functions contained in the profile $\boldsymbol{u}^0$ with the weights given by the profile of each individual's possibly different consensus vectors $C = (\boldsymbol{c}^1, \ldots, \boldsymbol{c}^{|N|})$.

Definition (Acceptable Range of Successful Deliberation) : We say that the profile of each individual's post-deliberation preferences $\succsim^1$ is within the *acceptable range of successful deliberation* if and only if $C\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^1}$ for some $C \in \mathcal{C}$ and some $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$ representing each individual's pre-deliberation preferences $\succsim^0$.

In other words, given a profile of initial pre-deliberation preferences $\succsim^1$, if there exists no $C \in \mathcal{C}$ and $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$ such that $C\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^1}$, then the post-deliberation preference profile $\succsim^1$ will be *outside* the acceptable range of successful deliberation and will be

considered implausible. This is because such a preference profile could not have possibly been generated by the convex combination of *any* profile of utility functions representing the initial profile of pre-deliberation preferences with weights given by *any* possible consensus vectors. In other words, a post-deliberation preference profile will be outside the acceptable range of successful deliberation if there is no possible way the individuals could have reached at such a profile of post-deliberation preferences by mutually convincing and persuading one another through reasoned deliberation.

Our definition of the acceptable range of successful deliberation easily explains why i's post-deliberation preference change in Example 2 was unacceptable. In that example, since everybody prior to deliberation unanimously preferred x to y to z, any utility representation of each individual's pre-deliberation preferences would have to assign the highest number to x, the second highest number to y, and the lowest number to z. Given this, there could be no convex combinations of the individuals' utility representations that would generate the highest number for z and the lowest number for x, which would be required to represent individual i's post-deliberation preferences of preferring z to y to x. We illustrate the notion of the acceptable range of successful deliberation with two more examples below. The first example illustrates a case in which only a single individual i updates and changes his/her preferences after deliberation. The second example illustrates a case where multiple individuals update and change their preferences simultaneously after deliberation.

But before we present the examples, it is important to understand that our approach does not require that the individuals have a particular profile of utility functions in their minds and actually compute and perform convex combinations with their predefined consensus vectors during the process of deliberation. Post-deliberation preference change can happen naturally and automatically without any individual being consciously aware of or understanding the concepts of utility functions, convex combinations, their consensus vectors, etc. These concepts are simply used to model preference change and define the acceptable range of successful deliberation, which we will later use to define our NNRD axiom.

Example 3: The Acceptable Range of Successful Deliberation of a Single Individual

Suppose that there are three individuals i, j and k, and three alternatives x, y, and z. Suppose that the initial pre-deliberation preference profile $\succsim^0$ is:

$$x \succ_i^0 y \succ_i^0 z, \quad z \succ_j^0 x \succ_j^0 y, \quad x \succ_k^0 z \succ_k^0 y$$

Suppose that, after engaging in successful democratic deliberation, only i's preferences changed, while j's and k's preferences remained unchanged: i.e., $\succsim_i^1 \neq \succsim_i^0$ and $(\succsim_j^1, \succsim_k^1) = (\succsim_j^0, \succsim_k^0)$. Given successful deliberation, what sort of preference change would it be admissible for i? Note that the three individuals' initial pre-deliberation preferences $\succsim^0$ can be represented by the following three utility functions:

$$(u_i^0(x), u_i^0(y), u_i^0(z)) = (2, 1, 0),$$
$$(u_j^0(x), u_j^0(y), u_j^0(z)) = (1, 0, 3),$$
$$(u_k^0(x), u_k^0(y), u_k^0(z)) = (3, 1, 2).$$

Such a profile of utility representations satisfies $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$. Suppose that, during deliberation, i gives equal considerations to his/her own opinions/preferences as well as those of j. This can be represented by i's consensus vector $\boldsymbol{c}^i = (c_i^i, c_j^i, c_k^i) = (1/2, 1/2, 0)$. Then, the convex combination $\boldsymbol{c}^i\boldsymbol{u}^0$ is

$$(\boldsymbol{c}^i\boldsymbol{u}^0(x), \boldsymbol{c}^i\boldsymbol{u}^0(y), \boldsymbol{c}^i\boldsymbol{u}^0(z)) = \frac{1}{2}(u_i^0(x), u_i^0(y), u_i^0(z)) + \frac{1}{2}(u_j^0(x), u_j^0(y), u_j^0(z)) = (\frac{3}{2}, \frac{1}{2}, \frac{3}{2}).$$

Thus, $\boldsymbol{c}^i\boldsymbol{u}^0$, which represents i's post-deliberation preferences, represents the preference: $x \sim_i' z \succsim_i' y$. Here, since $c_j^i > 0$ and $c_k^i = 0$, we can say that j was the only person who was able to convince i during deliberation, while k wasn't able to convince i at all. That is, we have: $S = j$. Since $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$, $\boldsymbol{c}^i = (c_i^i, c_j^i, c_k^i) = (1/2, 1/2, 0) \in \mathcal{C}$, and $\boldsymbol{c}^i\boldsymbol{u}^0 \in \mathcal{U}_{\succsim_i'}$, we can say that i's post-deliberation preference $\succsim_i'$ was within the acceptable range of successful deliberation.

Note that $\boldsymbol{u}^0$ above is not the only utility function that represents $\succsim^0$, and the consensus vector is not limited to $\boldsymbol{c}^i$ above. However, if some individual's post-deliberation preference can be represented by a convex combination of at least one pre-deliberation utility function profile and at least one consensus vector, then, according to our framework, such a post-deliberation preference is an acceptable convex combination preference derived from a valid utility representation of $\succsim^0$ and is thereby within the acceptable range of successful deliberation.

Under the current profile of initial pre-deliberation preferences $\succsim^0$, the entire set of i's acceptable post-deliberation preference $\succsim_i'$ can be summarized as follows: [3]

- xyz, $x(yz)$, xzy, $(xz)y$, and zxy when $S = \{j, k\}$ (i.e., i gives positive considerations to both j and k.)

[3]Here, $(xy)z$ means $x \sim_i y \succ_i z$.

- xyz, $x(yz)$, xzy, $(xz)y$, and zxy when $S = \{j\}$ (i.e., i gives positive considerations to only j but not k.)
- xyz, $x(yz)$, and xzy when $S = \{k\}$ (i.e., i gives positive considerations to only k but not j.)

All of these potential post-deliberation preferences of i are within the acceptable range of successful deliberation (with j and k.)

However, the post-deliberation preference $z \succ_i'' y \succ_i'' x$ cannot be represented by a convex combination of any pre-deliberation utility function profile and any consensus vector.[4] Therefore, a change from a pre-deliberation preference $\succsim_i^0$ to a post-deliberation preference $\succsim_i''$ will be considered *unacceptable* and will be ruled out from the acceptable range of successful deliberation.

Example 4: The Acceptable Range of Successful Deliberation for Multiple Agents Updating Simultaneously Consider again the same $\succsim^0$ as that of Example 3:

$$x \succ_i^0 y \succ_i^0 z, \quad z \succ_j^0 x \succ_j^0 y, \quad x \succ_k^0 z \succ_k^0 y$$

Let us now consider the case where multiple individuals update their preferences simultaneously during deliberation. As in Example 3, if a preference profile can be represented as a convex combination of at least one utility function profile and one consent vector profile, then it is a convex combination profile with a utility representation of $\succsim^0$ and is, thereby, within the acceptable range of post-deliberation preference change given successful deliberation. For example, let $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$ be:

$$\begin{aligned}(u_i^0(x), u_i^0(y), u_i^0(z)) &= (2, 1, 0),\\ (u_j^0(x), u_j^0(y), u_j^0(z)) &= (1, 0, 3),\\ (u_k^0(x), u_k^0(y), u_k^0(z)) &= (6, 1, 5).\end{aligned}$$

Let $C = (\boldsymbol{c}^i, \boldsymbol{c}^j, \boldsymbol{c}^k)$ be:

$$(c_i^i, c_j^i, c_k^i) = (\frac{1}{2}, \frac{1}{2}, 0), \quad (c_i^j, c_j^j, c_k^j) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}), \quad (c_i^k, c_j^k, c_k^k) = (0, \frac{1}{2}, \frac{1}{2}).$$

[4] This is because the pre-deliberation preferences of all three individuals $\succsim_i^0, \succsim_j^0$ and $\succsim_k^0$ prefer x to y while i's post-deliberation preference $\succsim_i''$ prefers y to x.

The set of consensus vectors imply that, during deliberation, i gives equal considerations only to his/her own views as well as those of j; j gives equal considerations to all three individuals; and k gives equal considerations only to his/her own views as well as those of i. Due to such differences in their respective consensus vectors (viz. the degree to which each individual agrees with the opinions of other individuals), each individual's post-deliberation preferences will be different from one another. Specifically, after deliberation, each individual's post-deliberation utility function becomes:

$$\begin{aligned}(\boldsymbol{c}^i\boldsymbol{u}^0(x), \boldsymbol{c}^i\boldsymbol{u}^0(y), \boldsymbol{c}^i\boldsymbol{u}^0(z)) &= (\frac{3}{2}, \frac{1}{2}, \frac{3}{2}),\\ (\boldsymbol{c}^j\boldsymbol{u}^0(x), \boldsymbol{c}^j\boldsymbol{u}^0(y), \boldsymbol{c}^j\boldsymbol{u}^0(z)) &= (3, \frac{2}{3}, \frac{8}{3}),\\ (\boldsymbol{c}^k\boldsymbol{u}^0(x), \boldsymbol{c}^k\boldsymbol{u}^0(y), \boldsymbol{c}^k\boldsymbol{u}^0(z)) &= (\frac{7}{2}, \frac{1}{2}, 4),\end{aligned}$$

which implies that the post-deliberation preference profile $\succsim^1$ after deliberation is:

$$x \sim_i^1 z \succ_i^1 y, \quad x \succ_j^1 z \succ_j^1 y, \quad z \succ_k^1 x \succ_k^1 y.$$

Thus, $\succsim^1$ is one of the convex combination profiles that can be obtained with utility representations of $\succsim^0$, and, hence, the change from the pre-deliberation profile $\succsim^0$ to the post-deliberation profile $\succsim^1$ is within the acceptable range of successful deliberation. Of course, there are other post-deliberation preference changes that will be in the acceptable range of successful deliberation as well. However, the following $\succsim^{1'}$ does not satisfy $C\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^{1'}}$ under any $\boldsymbol{u}^0 \in \mathcal{U}_{\succsim^0}$ and any $C \in \mathcal{C}$:

$$z \succ_i^{1'} y \succ_i^{1'} x, \quad y \succ_j^{1'} x \succ_j^{1'} z, \quad y \succ_k^{1'} z \succ_k^{1'} x$$

Thus, the change from the pre-deliberation profile $\succsim^0$ to the post-deliberation profile $\succsim^{1'}$ is not within the acceptable range of successful deliberation.

To sum up, a change from a pre-deliberation preference profile $\succsim^0$ to a post-deliberation profile $\succsim^1$ when i gives positive considerations to the opinions of the individuals in the set S is in the acceptable range of successful deliberation if and only if (i) $\succsim^1$ is a convex combination profile with a utility representation of $\succsim^0$ and, in particular, (ii) $\succsim_i^1$ is a convex combination profile of $\succsim^0$ restricted to $\{i\} \cup S$.

3.2.1 Two Types of (NNRD) Conditions

We now define our main axiom, Non-Negative Response toward Successful Deliberation (NNRD), formally:

(NNRD) : For any $i \in N$, any non-empty $S \subseteq N \setminus \{i\}$, and any $\succsim^0, \succsim^1 \in \mathcal{R}$ such that $C\boldsymbol{u} \in \mathcal{U}_{\succsim^1}$ for some $(C, \boldsymbol{u}) \in \mathcal{C} \times \mathcal{U}_{\succsim^0}$ with $S = \{j \neq i \mid c^i_j > 0\}$,

$$\exists j \in S, \;\; f(\succsim^1) \succsim^0_j f(\succsim^0_i, \succsim^1_{-i}).$$

In words, (NNRD) says that given that the preference change from the pre-deliberation preference profile $\succsim^0$ to the post-deliberation preference profile $\succsim^1$ is within the acceptable range of successful deliberation, there must exist at least one individual j among the individuals who have positively persuaded individual i during deliberation, who is not made worse-off by the resulting social choice arrived at the later aggregation stage. In other words, if everybody, who was able to positively persuade individual i to change his/her preferences through reasoned deliberation, is made worse-off by the final social decision generated by the SCF, then (NNRD) is violated, and we can say that our SCF has, thereby, failed to accommodate the positive effects of successful deliberation. In such cases, we can say that deliberation was futile; it would have been better for people to choose not to engage in democratic deliberation at all and spare themselves of trying to persuade individual i with reasoned arguments. That would defeat the primary purpose of why we wish to incorporate democratic deliberation into our democratic decision making process in the first place. Hence, if we wish our democratic institutions to achieve the ideals of deliberative democracy, it is important that the SCF that we employ in the aggregation stage satisfies the NNRD axiom.

Since non-emptiness of $S = \{j \neq i \mid c^i_j > 0\}$ is equivalent to $c^i_i \neq 1$, the above definition is mathematically equivalent to the next one:

(NNRD: Alternate Definition) : For any $i \in N$, any $\succsim^0 \in \mathcal{R}$, any $(C, \boldsymbol{u}) \in \mathcal{C} \times \mathcal{U}_{\succsim^0}$, and any $\succsim^1 \in \mathcal{R}$ with $C\boldsymbol{u} \in \mathcal{U}_{\succsim^1}$ and $c^i_i \neq 1$,

$$\exists j \neq i \text{ s.t. } c^i_j > 0, \;\; f(\succsim^1) \succsim^0_j f(\succsim^0_i, \succsim^1_{-i}).$$

Since this is clearer and also mathematically simpler than the first expression, we adopt this as the formalization of (NNRD) in many of our subsequent analyses and proofs.

Note that the definition of (NNRD) allows multiple individuals to simultaneously change their preferences during and after deliberation. We might consider a special case of (NNRD) in which only a single individual changes his/her preferences during and after deliberation, while everybody else's preference remains unchanged and is the same as his/her pre-deliberation preferences. We define this as *Weak* NNRD:

Weak NNRD (WNNRD) : For any $i \in N$, andy $\succsim^0 \in \mathcal{R}$, any $(\boldsymbol{c}^i, \boldsymbol{u}) \in \mathcal{C}^i \times \mathcal{U}_{\succsim^0}$, and any $\succsim_i^1 \in \mathcal{R}_i$ with $\boldsymbol{c}^i\boldsymbol{u} \in \mathcal{U}_{\succsim_i^1}$ and $c_i^i \neq 1$,

$$\exists j \neq i \text{ s.t. } c_j^i > 0, \quad f(\succsim_i^1, \succsim_{-i}^0) \succsim_j^0 f(\succsim^0).$$

(WNNRD) considers the special case of NNRD where $c_j^j = 1$ for each $j \neq i$, and, hence, $\succsim_{-i}^0 = \succsim_{-i}^1$. That is, it considers the special case in which the post-deliberation preference of everybody besides individual i's is the same as his/her pre-deliberation preference. It represents a situation in which only individual i has changed his/her preferences by being persuaded by other people's opinions and arguments during democratic deliberation. (WNNRD) requires that whenever this is so, there must exist at least one other person, whose opinions and arguments individual i has given positive weight during deliberation, to be made not worse-off by the resulting social choice generated by the post-deliberation aggregation rule after individual i has changed his/her preferences. In other words, if only individual i's preference has changed as a result of democratic deliberation, then (WNNRD) requires that there must exist at least one person, who has successfully persuaded i during democratic deliberation, to be at least as good as s/he would have been if s/he did not persuade i by the resulting social decision.

We believe that (NNRD) is not a strong normative requirement. It does not demand that *everybody*, who has successfully persuaded another individual during deliberation, should be made *better-off* by the resulting social choice generated by the post-deliberation aggregation rule. It merely requires that the post-deliberation aggregation rule should minimally respect the results of successful deliberation so that there should be *at least one person*, who has successfully persuaded another person during deliberation, who is not made worse-off by the resulting social choice generated by the post-deliberation aggregation rule. In this sense, (NNRD) is normatively weak. As (NNRD) implies (WNNRD), (WNNRD) is even weaker. In Section 4, we will show that the impossibility theorem is obtained even if we require the weaker NNRD axiom, (WNNRD).

3.3 Other Standard Axioms

We introduce two other categories of normative criteria besides (NNRD) that apply to our post-deliberation aggregation rule: one that concerns respecting deliberative democracy's ideal of *unanimous consensus* and the other that concerns respecting deliberative democracy's ideal of *democratic equality*.

As for the axioms that are designed to respect deliberative democracy's ideal of unanimous consensus, the following three axioms are stated in order of strength:

Pareto Optimality (PO) For any $\succsim \in \mathcal{R}$ and any $x, y \in X$, if $x \succsim_i y$ for all $i \in N$ and there exists $j \in N$ such that $x \succ_j y$, then $f(\succsim) \neq y$.

Weak Pareto Optimality (WPO) For any $\succsim \in \mathcal{R}$ and any $x, y \in X$, if $x \succ_i y$ for all $i \in N$, then $f(\succsim) \neq y$.

Top Unanimity (TU) For any $\succsim \in \mathcal{R}$, any $x \in X$, if $x \succ_i y$ for all $i \in N$ and all $y \neq x$, then $f(\succsim) = x$.

As explained previously, many early deliberative democratic theorists have regarded unanimous consensus as a regulative ideal of deliberative democracy. The intuitive appeal of these three axioms comes in part from our desire to respect unanimous consensus whenever we have successfully reached at one through democratic deliberation. (PO) requires that, for any two alternatives x and y, if everybody, after deliberation, thinks that x is at least as good as y and there exists at least one individual who thinks that x is strictly better than y, then our post-deliberation aggregation rule should not choose y as its social outcome. (WPO) requires that, for any two alternatives x and y, if everybody, after deliberation, thinks that x is strictly better than y, then our post-deliberation aggregation rule should not choose y as its social outcome.

Whereas the two Pareto axioms incorporate deliberative democracy's ideal of unanimous consensus by using unanimity as a basis to reject or eliminate universally dispreferred alternatives, (TU) incorporates deliberative democracy's ideal of unanimous consensus by requiring our second stage aggregative rule to actively choose the alternative that everybody considers to be the best whenever such an alternative exists and is identified through democratic deliberation. It is easy to see that (PO) is the strongest axiom, (WPO) is weaker, and (TU) is the weakest among the three axioms: that is, (PO) implies (WPO), which in turn implies (TU), but not the other way round.

We think that the normative appeal of (WPO) and (TU) are uncontroversial: both are very weak axioms in the sense that both operate only when there is a clear unanimity that

one alternative is strictly better than another. However, compared to (WPO) we think that (TU) is a bit too weak because of its very limited applicability. To see this, note that (TU) operates only when deliberative democracy practically achieves its regulative ideal to its fullest: that is, (TU) applies only when democratic deliberation leads to full unanimous agreement toward a single best alternative, which every participant agrees to be strictly better than any other alternative. But if that is the case, then there is no reason to rely on any post-deliberation aggregation rule to reach a democratic decision in the first place. The main reason that we are considering a two-stage democratic process, where aggregative voting occurs after democratic deliberation, is based on our practical understanding of the moral nature of modern pluralistic democracies that unanimous consensus toward a single best alternative can rarely be achieved through sustained democratic deliberation even under ideal circumstances. This means that (TU), which is an axiom designed to apply to aggregative voting rules, can only apply to situations where there is no need to rely on aggregative voting rules at all. Whenever we need to rely on an aggregative voting rule to reach a collective decision – that is, whenever democratic deliberation fails to reach full unanimous consensus toward a single best alternative – (TU) has no bite.

Moreover, unlike (WPO), there is an important sense in which (TU) fails to incorporate deliberative democracy's ideal of unanimous agreement. Suppose that after successful democratic deliberation, everybody unanimously agrees that despite there existing no clear winner that beats every other alternative, some alternatives are clearly worse than other alternatives. We can say that, here, we have made some progress through democratic deliberation because although we were not able to identify the uncontroversially best alternative, we were able to at least identify through democratic deliberation what alternatives we should *not* choose in the second aggregation stage. (WPO) is able to accommodate this kind of deliberative result by eliminating such dominated or inferior alternatives so that they do not get considered in the second aggregation stage. By contrast, (TU) is unable to eliminate such dominated or inferior alternatives and these alternatives will still be considered in the second aggregation stage. Hence, we can say that, unlike (WPO), (TU) lacks proper responsiveness to at least one particular way we wish our second stage aggregation rule to accommodate the results of the first stage democratic deliberation that has successfully achieved unanimous agreement; and that is by eliminating the unanimously agreed upon uncontroversially bad alternatives. This is why we think that (WPO) better accommodates deliberative democracy's ideal of unanimous agreement than (TU). Hence, we will impose (WPO) on our aggregation rule.

We now talk about the other regulative ideal of deliberative democracy: democratic equality. As a requirement of democratic equality, we would want our aggregation rule to treat the individuals fairly by not conferring any excess political decision-making power to any specific individual. The strongest axiom that requires the SCF to treat all individuals symmetrically would probably be the axiom known as 'Anonymity (AN)':

Anonymity (AN) For any $\succsim \in \mathcal{R}$ and any permutation π on N, $f(\succsim) = f((\succsim_{\pi(i)})_N)$.

(AN) requires the SCF to treat all individuals completely symmetrically. However, in many real life democratic decision making situations, requiring (AN) might be a bit too demanding. For instance, it might not only be permissible, but actually more in align with the aims of liberal democracies to confer a certain degree of veto power to each individual to block certain social outcomes, which can potentially violate their constitutionally guaranteed basic rights, from being socially imposed by the majority. In this sense, granting a certain degree of veto power to each individual, provided that it is within certain limits, is consistent with (and may even be required for) the protection of each person's basic liberal rights. The fact that such measures violate the (AN) axiom does not necessarily mean that a SCF that grants individuals a certain degree of veto power for the purpose of protecting their basic rights disrespects the ideals of democratic equality. Rather, what would violate our commitment to democratic equality is to arbitrarily extend and expand the veto power of a particular individual so that s/he has a rather unlimited power to block *almost any* social alternative from being socially chosen regardless of what everybody else prefers. Hence, even though (AN) seems a bit too restrictive and we have good reasons to relax (AN), we would at least want our SCF to avoid allowing the existence of such a *universal vetoer.*

Let us try to understand this a bit more formally. For each $Y \subseteq X$ with $|Y| \geq 2$, let $\mathcal{R}^Y$ be such that $\succsim \in \mathcal{R}^Y$ if and only if $y \succ_i x$ for each $i \in N$, for each $y \in Y$, and for each $x \in X \setminus Y$. That is, $\succsim \in \mathcal{R}^Y$ means that every alternative in $X \setminus Y$ is considered to be strictly worse than every alternative in Y by *everyone.* Thus, $\mathcal{R}^Y$ represents the situations where the alternatives in $X \setminus Y$ are unanimously considered to be "no longer worthy of further considerations." That is, we may think of $X \setminus Y$ as the set of alternatives that did not make the "first-cut" after democratic deliberation. Note that such a situation may occur quite frequently when the individuals engage in a prior stage of democratic deliberation; after exchanging reasoned arguments with one another, it is reasonable to expect that the individuals will be able to at least narrow down their options to a few viable choices even if they cannot fully reach agreement on which alternative would be the best.

For an arbitrary SCF f, we say that $v \in N$ is a *vetoer on* $Y \subseteq X$ if and only if, for every $y \in Y$, there exists $\succsim_v \in \mathcal{R}_v^Y$ such that $f(\succsim) \in Y \setminus \{y\}$ for every $\succsim_{-v} \in \mathcal{R}_{-v}^Y$. That is, $v \in N$ is a vetoer on Y if and only if when all the individuals, after engaging in democratic deliberation, unanimously agree that the alternatives in $X \setminus Y$ should no longer be considered, for any alternative $y \in Y$ (that is, for any alternative y that has made the "first-cut" after democratic deliberation and is currently under consideration), v can block the SCF from choosing y by submitting a specific preference in $\mathcal{R}_v^Y$ regardless of the others' preferences.

For an arbitrary SCF f, we say that $v \in N$ is a *universal vetoer* if and only if v is a vetoer on any $Y \subseteq X$ with $|Y| \geq 2$. The axiom *No Universal Vetoer* (NV) requires that our SCF must not allow anybody to be a universal vetoer:

No Universal Vetoer (NV) f does not have a universal vetoer.

A closely related concept to a universal vetoer is that of a *dictator*. We say that $i \in N$ is a *dictator* of the SCF f if and only if, for any $\succsim \in \mathcal{R}$ and $x \in X$, $f(\succsim) \succsim_i x$. The axiom *Non-dictatorship* (ND) requires that our SCF must not allow anybody to be a dictator:

Non-Dictatorship (ND) f does not have a dictator.

A dictator can unilaterally *impose* his/her preferred alternative despite everybody else's opposition *unconditionally*. By contrast, a universal vetoer can unilaterally *reject* any alternative s/he disprefers among any subset of alternatives that makes the "first-cut" despite everybody else's approval of that alternative. A universal vetoer is weaker than a dictator; in particular, a dictator is always a universal vetoer, yet a universal vetoer is not always a dictator. Hence, (NV) implies (ND); but (ND) does not always imply (NV).

Nevertheless, even though a universal vetoer is weaker than a dictator, a universal vetoer is very close to a dictator especially in the context of successful deliberation. As already explained, when democratic deliberation is successful, it can help narrow down the disagreement among the participants even when it falls short of generating full unanimous consensus toward a single best alternative. One way that this may occur is by the elimination of options that are unanimously agreed to be no longer worthy of consideration at the aggregation stage. As deliberation becomes more successful, we can reasonably expect that the participants will tend to converge toward a smaller set of considered options after deliberation, and as the post-deliberation set of considered

options become smaller, the more powerful the universal vetoer becomes. Once the post-deliberation set of considered options is reduced to a pair of alternatives, the universal vetoer will, in effect, have the same power as a dictator, as s/he can now *unilaterally impose* his/her preferred alternative as the collective choice even when everybody votes against it at the aggregation stage. In other words, a universal vetoer can socially impose his/her preferred alternative between *any* pair of alternatives despite everybody else's opposition *given that* there exists a prior agreement (established through successful democratic deliberation) that the specific pair of alternatives is better than all other alternatives.

We believe that despite not being a full dictator, the existence of such a universal vetoer goes against our commitment to democratic equality; it confers too much decision-making power to a single individual that goes far beyond what is necessary to protect the individual's basic rights. Note that the existence of a universal vetoer will violate (AN).[5] However, the existence of a universal vetoer does not violate (ND) and is compatible with the axiom. This is why we think that (ND) is too weak to guarantee that our SCF fully respects democratic equality.

In short, in terms of incorporating the ideals of democratic equality into our SCF at the aggregation stage, we believe that (AN) is a bit too strong, while (ND) is a bit too weak. Hence, we will work with (NV) to state and prove our impossibility theorem in the next section.

4 Results

4.1 The main theorem

In this section, we assume that people's initial pre-deliberation preferences are weak orders $\succsim \in \mathcal{R}$. We have the following impossibility result when $|N| \geq 4$.

Theorem 1. There exists no SCF f that satisfies (WPO), (NV) and (WNNRD).

The proof is in Appendix A.1. The overall structure of the proof shows that any SCF f that satisfies (WPO) and (WNNRD) will necessarily appoint some individual as a

[5] To see this, suppose that i is a universal vetoer of a SCF f. Then, i is also the vetoer on two distinctive alternatives a and b. That is, i can reject either a or b by reporting a specific preference in $\mathcal{R}_i^{\{a,b\}}$ when the others' preference profile is in $\mathcal{R}_{-i}^{\{a,b\}}$. If our SCF f satisfies (AN), then this would imply that any individual $j \neq i$ can also reject either a or b by reporting the same preferences that i may report to reject either alternative. But, then, if i vetoes a and j vetoes b and the other individuals' preferences are in $\mathcal{R}_k^{\{a,b\}}$, our SCF f will be unable to choose either a or b even though a and b are the only remaining viable alternatives. This also implies that the universal vetoer must be at most one.

universal vetoer, and, thereby, violate (NV): i.e., for any SCF f satisfying (WPO) and (WNNRC), there exists an individual $i \in N$ such that, for any $Y \subseteq X$ with $|Y| \geq 2$, $\exists \succsim_i \in \succsim_i^Y$ s.t. $f(\succsim) \neq x, \quad \forall (\succsim_{-i}, x) \in \succsim_{-i}^Y \times X$.

Here is the sketch of the proof. First, we assume that the individual preferences form linear orders. In this environment, we show that the combination of (WPO) and (WNNRD) implies Maskin monotonicity[6] under certain conditions (which we will not go into details). Maskin monotonicity (under the specific conditions) says that if a certain alternative x was socially chosen from the initial preference profile, and the alternative x does not go down in anybody's preference ranking in the updated preference profile, then x must continue to be socially chosen in the updated preference profile. By linking together the situations in which these conditions are satisfied, we show that, for each pair of distinct alternatives, there exists a vetoer. We then show that these vetoers must be a single individual. As a final step, we show that this pairwise vetoer is actually a *universal* vetoer, whose veto power is limited neither to pairs of alternatives nor to linear preferences, but also extends more generally to preferences that are weak orders.

Note that, between the two NNRD axioms, Theorem 1 uses the *weaker* WNNRD axiom (that says that if i's preference has changed as a result of successful deliberation, then there should be at least one person among those who have positively persuaded i during deliberation, who is not made worse-off by the resulting social choice generated by the post-deliberation aggregation rule.) Also, between the two Pareto optimality axioms, Theorem 1 uses the *weaker* WPO axiom (that says that if everybody thinks that x is strictly better than y, our aggregation rule should not choose y), which is normatively easier to justify than the stronger PO in terms of respecting deliberative democracy's ideal of unanimous consensus. Theorem 1 shows that there exists no post-deliberation aggregation rule that can simultaneously: (a) respect deliberative democracy's ideal of unanimous consensus (WPO); (b) respect our firm commitment to democratic equality by not conferring too much political decision-making to any specific individual (NV); and (c) non-negatively respond to the direction of people's positive preference change that occurs after successful deliberation (WNNRD).

We believe that our impossibility theorem derives from minimal number of relatively weak and uncontroversial axioms. For instance, we do not impose Arrow's condition of Independence of Irrelevant Alternatives (IIA), which many scholars have criticized both of its descriptive and normative force (see Mackie 2003: 156). Further note that (WNNRD) is necessary to derive our impossibility theorem; without (WNNRD), stan-

[6]Maskin (1999) introduced this condition, and Reny (2001) used to show the Gibbard–Satterthwaite theorem.

dard voting rules with tie-breakers, such as the plurality rule or the Borda rule, become possible voting rules that satisfy both (WPO) and (NV). (We will consider the plurality rule and the Borda rule in detail in Section 4.4.) Thus, given that we accept (WPO) and (NV), we can say that it is (WNNRD) that is responsible for generating the impossibility result. Yet, we want our aggregation rule to satisfy (WNNRD), because, otherwise, incorporating democratic deliberation into our democratic process becomes futile.

In the remainder of our paper, we will consider three potential escape routes to our impossibility theorem: the first is to assume $|N| = 3$; the second is to relax (WPO) and (NV); the third is to restrict the initial pre-deliberation preference domain. It will be argued that all of these potential escape routes face critical limitaions.

4.2 Possible Way Out 1: $|N| = 3$

Theorem 1 assumes four or more individuals. One way to escape the impossibility theorem is to assume that there is only three individuals who engage in democratic deliberation, and exactly three social alternatives under consideration.

To show this, suppose that we have exactly three individuals and three social alternatives: viz., $|N| = 3$, and $X = \{x, y, z\}$. Let $Mj : \mathcal{R} \to \{y, z\}$ be such that, for any $\succsim \in \mathcal{R}$,

$$|\{i \in N | y \succsim_i z\}| \geq |\{i \in N | z \succsim_i y\}| \quad \Leftrightarrow \quad Mj(\succsim) = y.$$

In words, Mj is an aggregation rule that chooses the pairwise majority winner between y and z, and, when there is a tie, it chooses y. Then, we may define what we call the x-priority rule, f^x, as follows:

Definition 1 (The x-priority rule, f^x)**.** For any $\succsim \in \mathcal{R}$, $f = f^x$ if and only if

$$f(\succsim) = \begin{cases} x & \text{if } \nexists x' \in \{y, z\} \text{ s.t. } \forall i \in N, x' \succ_i x, \\ y & \text{if } [\forall i \in N, y \succ_i x] \ \& \ [\exists i \in N, x \succsim_i z], \\ z & \text{if } [\forall i \in N, z \succ_i x] \ \& \ [\exists i \in N, x \succsim_i y], \\ Mj(\succsim) & \text{if } \forall i \in N, \ y \succ_i x \ \& \ z \succsim_i x. \end{cases}$$

According to the x-priority rule f^x: If neither y nor z, strictly Pareto dominates x, then x is chosen as the default social choice; if only one of either y or z strictly Pareto dominates x, then that alternative is socially chosen; finally, if both y and z strictly Pareto dominate x, then the pairwise majority winner between y and z is socially chosen, and when there is a tie, y is socially chosen.

Proposition 1 shows that the x-priority rule f^x satisfies (WP), (NV), and (WNNRC).

Proposition 1. SCF f^x satisfies (WPO), (NV), and (WNNRD).

The proof of proposition 1 is in Appendix A.2.

Although employing the x-priority rule with exactly 3 individuals and 3 social alternatives does provide an escape route to our impossibility theorem, we think that such an escape route is very limited for obvious reasons. First, the escape route that relies on the x-priority rule works only under the extremely limited situation in which there is exactly three individuals and exactly three social alternatives. Hence, we are unable to escape the impossibility theorem in more general democratic settings in which we have more than three individuals and more than three social alternatives under consideration. Second, another limitation of the x-priority rule is that it can be seen as giving a rather unfair priority to some alternative, in particular, x, which has not been justified either by democratic deliberation or by the individuals' preferences; this may seem arbitrary. Note that according to the x-priority rule, only the alternatives that *every* individual (after democratic deliberation) unanimously considers to be strictly better than x can beat x. If x is the status quo, then the x-priority rule gives a very strong bias toward retaining the status quo and hinder social change. This might be undesirable in situations where the status quo lacks a rational basis and does not deserve to have such a privileged status.

4.3 Possible Way Out 2: Relaxing the Axioms

Our impossibility theorem shows that (WPO) and (NV) together are incompatible with (WNRRD). Remember that (WPO) was justified in terms of respecting deliberative democracy's ideal of unanimous consensus and (NV) was justified in terms of respecting deliberative democracy's ideal of democratic equality. We now consider whether we can avoid the impossibility theorem if we used weaker axioms than either (WPO) or (NV). And it will turn out that we can indeed escape the impossibility theorem with such a strategy. This shows that our impossibility theorem is *tight*.

Throughout this section, we will introduce indices for alternatives: i.e., $X \stackrel{\text{def}}{=} \{x_1, \ldots, x_{|X|}\}$. For each non-empty $Y \subseteq X$, let

$$r(Y) = \min\{m \in \{1, \ldots, |X|\} \mid x_m \in Y\},$$

that is, $r(Y)$ denotes the smallest index of alternatives in Y. r provides the tie-breaking system, which chooses the alternative $x_{r(Y)} \in Y$ for each $Y \subseteq X$. The indices are used only to define the above tie-breaking system. (Rearranging indices has no effect on the conclusions.)

4.3.1 A SCF Satisfying (WNNRD), (NV), (TU)

Suppose we replace (WPO) with (TU). Recall:

Top Unanimity (TU) For any $\succsim \in \mathcal{R}$, any $x \in X$, if $x \succ_i y$ for all $i \in N$ and all $y \neq x$, then $f(\succsim) = x$.

Now, consider any alternative $d \in X$. We define the unanimity rule with default d in the following way:

Definition 2 (The Unanimity Rule with default d, f^d)**.** $f = f^d$ if and only if, for each $\succsim \in \mathcal{R}$,

$$f(\succsim) = \begin{cases} x & \text{if } \exists x \in X \text{ such that } x \succ_i y, \forall i \in N \text{ and } \forall y \neq x \\ d & \text{if otherwise.} \end{cases}$$

From Definition 2, f^d chooses the unanimously agreed best alternative if such an alternative can be identified after democratic deliberation, and chooses the "default alternative" d otherwise. We can easily show that f^d does not satisfy (WPO). Choose distinct $x, y \neq d$ and consider a $\succsim \in \mathcal{R}$ such that

$$\exists i \in N \text{ s.t. } x \succ_i y \succ_i z \;\; \& \;\; y \succ_j x \succ_j z \quad \forall j \neq i, \forall z \notin \{x, y\}.$$

Then, d is Pareto dominated by both x and y. However, since there exists no alternative that everybody considers to be the best, we have $f^d(\succsim) = d$. Hence, by replacing (WPO) with (TU), we obtain the following possibility result for f^d.

Proposition 2. f^d satisfies (TU), (NV), and (WNNRC).

The proof of this proposition is in Appendix A.3. The escape route described in Proposition 2 is not very attractive for reasons that we have already explained in the previous sections. The aggregation rule f^d gives a rather unfair advantage to some default alternative d, which is not justified through democratic deliberation or people's preferences. f^d also fails to fully incorporate deliberative democracy's regulative ideal of unanimous consensus because, given there exist no alternative that emerges as the clear winner, f^d will still choose the default alternative d even when everybody unanimously considers d to be uncontroversially bad after democratic deliberation. We get a possibility, but only by compromising our ability to respond positively toward deliberative consensus.

4.3.2 A SCF Satisfying (WNNRD), (WPO), and (ND)

Next, suppose that we drop (NV) and instead add (ND). Recall:

Non-Dictatorship (ND) f does not have a dictator. That is, there exists no $i \in N$ such that for any $\succsim \in \mathcal{R}$ and $x \in X$, $f(\succsim) \succsim_i x$.

Since we still impose both (WNNRD) and (WPO), by Theorem 1, the SCF must have a unique universal vetoer. The basic strategy of the escape route is to define an aggregation rule that appoints somebody as a universal vetoer, but who is, nonetheless, not fully a dictator. Throughout the section, suppose that $v \in N$ is the universal vetoer, whose existence is implied by both (WNNRD) and (WPO) according to Theorem 1.

For each $\succsim \in \mathcal{R}$, and for each non-empty $N' \subseteq N$, let

$$X^{top}_{\succsim_{N'}} \stackrel{\text{def}}{=} \{x \in X \mid x \succsim_i y \ \forall (y,i) \in X \times N'\}$$

$$X^{bot}_{\succsim_{N'}} \stackrel{\text{def}}{=} \{x \in X \mid y \succsim_j x \ \forall (y,j) \in X \times N'\}.$$

Intuitively, $X^{top}_{\succsim_{N'}}$ is the set of alternatives that everybody *in* N' considers to be the best, while $X^{bot}_{\succsim_{N'}}$ is the set of alternatives that everybody *in* N' considers to be the worst. As a notational convention, let $X^{top}_{\succsim} = X^{top}_{\succsim_N}$ and $X^{bot}_{\succsim} = X^{bot}_{\succsim_N}$ i.e., $X^{top}_{\succsim}$ is the set of alternatives that everybody considers to be the best and $X^{bot}_{\succsim}$ is the set of alternatives that everybody considers to be the worst.

For any preference $\succsim_v \in \mathcal{R}$ of the universal vetoer v such that $X^{top}_{\succsim_v} \neq X$, let

$$X^{sec}_{\succsim_v} \stackrel{\text{def}}{=} \{x \in X \setminus X^{top}_{\succsim_v} \mid x \succsim_v y \ \forall y \in X \setminus X^{top}_{\succsim_v}\}.$$

That is, $X^{sec}_{\succsim_v}$ is the set of universal vetoer v's second-best alternatives given $\succsim_v$. We now define what we call the *v-priority rule, f^v* in the following way:

Definition 3 (the v-priority rule, f^v)**.** [7] For each $\succsim \in \mathcal{R}$, $f = f^v$ if and only if

$$f(\succsim) = \begin{cases} x_{r(X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}})} & \text{if } X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}} \neq \emptyset \quad \text{(Case 1)} \\ x_{r(X^{sec}_{\succsim_v})} & \text{if } X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}} = \emptyset \ \& \ X^{bot}_{\succsim} \neq X \quad \text{(Case 2)} \\ x_1 & \text{if } X^{bot}_{\succsim} = X \quad \text{(Case 3).} \end{cases}$$

In words, the following explains how the v-priority rule f^v chooses the social alternative in the aggregation stage. For any $\succsim \in \mathcal{R}$,

[7] To confirm that the second line is well defined, we have to check that $X^{top}_{\succsim_v} \neq X$ holds under this condition. The proof is as follows: Suppose that $X^{bot}_{\succsim} \neq X$. Then, "$X^{bot}_{\succsim_v} \neq X$" or "$X^{bot}_{\succsim_v} = X$ and $X^{bot}_{\succsim_{-v}} \neq X$." Since $X^{bot}_{\succsim_v} \neq X \Leftrightarrow X^{top}_{\succsim_v} \neq X$, this means that "$X^{top}_{\succsim_v} \neq X$" or "$X^{top}_{\succsim_v} = X$ and $X^{bot}_{\succsim_{-v}} \neq X$." In other words, if $X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}} = \emptyset$, then $X^{top}_{\succsim_v} \neq X$.

Case 1 If there exist some alternatives in $X^{top}_{\succsim_v}$ that are not unanimously bottom-ranked for everybody other than v, then the rule chooses one among them,

Case 2 If all alternatives in $X^{top}_{\succsim_v}$ are unanimously bottom-ranked for everybody except v, then the rule chooses one of the second-best alternatives for $\succsim_v$, and

Case 3 If all alternatives in $X^{top}_{\succsim_v}$ are unanimously bottom-ranked for everybody except v, and there exists no second-best alternative for v because s/he is indifferent to every alternative in $\succsim_v$ (i.e., $X^{top}_{\succsim_v} = X$), then this rule chooses x_1.

We now state our possibility result:

Proposition 3. The v-priority rule f^v satisfies (WP), (ND) and (WNNRD).

The proof of the proposition is in Appendix A.4. Note that individual v designated by the v-priority rule f^v is not a full-fledge dictator because it is possible for the other individuals to "reject" one of v's best alternatives from being socially chosen by putting that alternative at the very bottom of their preference ranking. Even so, the universal vetoer v is guaranteed to get one of his/her second-best alternatives (provided that v is not completely indifferent to every alternative in X.) Hence, despite falling slightly short of being a full-fledge dictator, v, under the v-priority rule f^v, has enormous decision making power: specifically, v will be able to always impose one of his/her top-ranked alternatives as long as there exists at least one other person who does not rank at least one of v's top-ranked alternative at the very bottom (or equivalently, as long as it is not the case that every other individual ranks all of v's top-ranked alternatives at the very bottom); and even when everybody else unanimously ranks all of v's top-ranked alternatives at the very bottom, v will still be able to impose at least one of his/her second best alternatives. So, the main difference between a dictator and the universal vetoer v under the v-priority rule is that whereas a dictator is able to always impose one of his/her best alternatives, the universal vetoer v, while being able to impose one of his/her top-ranked alternative most of the time, can, at least fail and get one of his/her second-best alternatives (if everybody else can successfully coordinate through democratic deliberation to rank *all* of v's top-ranked alternatives at the very bottom of their preference rankings.) This is still an enormous undemocratic decision making power conferred to a single individual. Hence, we can say that this particular escape route comes at a significant democratic cost. We escape the impossibility theorem by paying the cost of compromising one of deliberative democracy's fundamental commitments to democratic equality.

4.4 Possible Way Out 3: Domain Restrictions

The third possible way to escape our impossibility theorem is by restricting our pre-deliberation or post-deliberation preference domain. Let $\mathcal{D}_i \subset \mathcal{R}_i$ denote the set of all *dichotomous* preference orders on X: specifically, $\succsim_i \in \mathcal{D}_i$ divides X into two groups $H_{\succsim_i}$ and $L_{\succsim_i}$ such that

$$a \sim_i a' \quad \forall a, a' \in H_{\succsim_i}; \quad b \sim_i b' \quad \forall b, b' \in L_{\succsim_i}; \quad a \succ_i b \quad \forall a \in H_{\succsim_i}, \quad \forall b \in L_{\succsim_i}.^{8}$$

In other words, an individual has dichotomous preferences if s/he is able to divide the set of all alternatives into two broad groups $H_{\succsim_i}$ and $L_{\succsim_i}$ and consider all alternatives in $H_{\succsim_i}$ as *equally good*, while considering all alternatives in $L_{\succsim_i}$ as *equally bad*. Let $\mathcal{D} = \prod_{i \in N} \mathcal{D}_i$ denote the set of all profiles of dichotomous preferences. The SCF *when the domain is restricted to* $\mathcal{D}$ is denoted by $f_{\mathcal{D}}$. Throughout section 4.4, we will consider possible escape routes when we restrict our initial pre-deliberation preference domain to $\mathcal{D}$.

Example 5 modifies our previous Example 1 so that the participants have dichotomous preferences both pre-deliberation and post-deliberation.

Example 5: Suppose that the initial profile, $\succsim \in \mathcal{D}$, is as follows:

Group A: six individuals have $H = \{x\}$, $L = \{y, z\}$;

Group B: five individuals have $H = \{y\}$, $L = \{x, z\}$; and

Group C: an individuals have $H = \{z\}$, $L = \{x, y\}$.

If the individuals do not engage in democratic deliberation and we derive the social choice from the individuals' pre-deliberation dichotomous preferences, any scoring rule or any approval voting rule will generate $f(\succsim) = x$ as its social outcome. Suppose, as it was the case for Example 1, after engaging in democratic deliberation, a single individual in Group A is persuaded by the arguments presented by the members of Group C and changes her preferences to '$H = \{z\}$, $L = \{x, y\}$.' Then, the post-deliberation preference profile, $\succsim' \in \mathcal{D}$, becomes

Group A': five individuals have $H = \{x\}$, $L = \{y, z\}$;

Group B': five individuals have $H = \{y\}$, $L = \{x, z\}$; and

[8] Strictly speaking, $\mathcal{D}_i \not\subset \mathcal{R}_i$ since $H_{\succsim_i}$ or $L_{\succsim_i}$ is able to be empty. However, we assume that i has the same preference that all alternatives are indifferent for i if $H_{\succsim_i} = 0$ or $L_{\succsim_i} = 0$. There is no effects of this setting on our results in Section 4.4.

Group C': two individuals have $H = \{z\}$, $L = \{x, y\}$.

By using any scoring or approval voting rule with a tie-breaker, we obtain either $f(\succsim') = x$ or $f(\succsim') = y$ as our social outcome in the aggregation stage. Since the individuals in Group C are indifferent between x and y, there exists no individual (in Group C) who has successfully persuaded another individual (i.e. an individual in Group A) through democratic deliberation that has been made worse-off by the social outcome generated by the aggregation rule. Hence, example 5 satisfies (NNRD).

Example 5 concerns a situation in which both the individuals' pre-deliberation and post-deliberation preferences are dichotomous; that is, the individuals all start with dichotomous pre-deliberation preferences before democratic deliberation and end up with dichotomous post-deliberation preferences after democratic deliberation as well. This can happen when each individual is either *completely persuaded* or *completely unpersuaded* by the opinions of other individuals during democratic deliberation.

Although it is perfectly possible for such situations (of either complete persuasion or complete inertia) to occur, a more common situation would be when the individuals change their preferences by being *partially persuaded* by other people's arguments and opinions during democratic deliberation. And when this occurs, it is possible for the domain of post-deliberation updated preferences to become $\mathcal{R}$ even when we start with dichotomous pre-deliberation preferences in $\mathcal{D}$.

For example, suppose that we have $X = \{x, y, z\}$ and i's initial pre-deliberation preferences are $x(yz)$, and j's initial pre-deliberation preferences are $(xy)z$. If we use $(u_i(x), u_i(y), u_i(z)) = (1, 0, 0)$ to represent i's pre-deliberation preference and $(u_j(x), u_j(y), u_j(z)) = (1, 1, 0)$ to represent j's pre-deliberation preference, and assume that i gives equal weight to his/her preference and j's preference during deliberation (i.e., $c_i^i = c_j^i = 0.5$), then i's post-deliberation preference becomes $(\boldsymbol{c}^i\boldsymbol{u}(x), \boldsymbol{c}^i\boldsymbol{u}(y), \boldsymbol{c}^i\boldsymbol{u}(z)) = (1, 0.5, 0)$, which implies that, after deliberation, i now strictly prefers x to y to z. This example shows that pre-deliberation preferences that start out to be dichotomous before deliberation can very well change into weak or linear orders after deliberation.

One practical way in which this may occur is for democratic deliberation to have an effect on *refining* people's preferences. That is, people, who previously lacked the ability to finely distinguish between similar but subtlely different alternatives may, after being presented, during democratic deliberation, with various reasons and considerations concerning the merits and demerits of different alternatives – considerations that were previously unknown or not salient – may start to develop a more sophisticated assessment over the various alternatives that allow them to now distinguish and finely rank those

alternatives, whose overall values were largely perceived to be roughly equivalent before they engaged in democratic deliberation.

In Section 4.4.1, we will investigate possible escape routes in such situations: that is, when people's pre-deliberation preferences are dichotomous while their post-deliberation preferences are weak orders. In Section 4.4.2, we will investigate possible escape routes in more restrictive cases where people's pre-deliberation preferences and their post-deliberation preference are *both* dichotomous.

4.4.1 Simple Scoring Rules with a Tie-Breaker: The Possibility of Plurality and Borda Rules when $\mathcal{D} \to \mathcal{R}$

Suppose that the participants start with dichotomous pre-deliberation preferences and end up with post-deliberation preferences that are weak orders. As explained above, this may happen when the participants update their pre-deliberation dichotomous preferences during deliberation by partially agreeing with the opinions and/or preferences of the other participants. Since we have now restricted the domain of pre-deliberation preferences to that of the set of dichotomous preferences $\mathcal{D}$, we redefine our (NNRD) axiom as follows:

(NNRD$_{\mathcal{D}\to\mathcal{R}}$) : For any $i \in N$, for any $\succsim^0 \in \mathcal{D}$, for any $(C, \boldsymbol{u}) \in \mathcal{C} \times \mathcal{U}_{\succsim^0}$, and for any $\succsim^1 \in \mathcal{R}$ with $C\boldsymbol{u} \in \mathcal{U}_{\succsim^1}$ and $c_i^i \neq 1$,

$$\exists j \neq i \text{ s.t. } c_j^i > 0 \ \& \ f(\succsim^1) \succsim_j^0 f(\succsim_i^0, \succsim_{-i}^1).$$

We now examine whether the two famous scoring rules[9] – namely, the plurality rule and the Borda rule – can jointly satisfy (NNRD$_{\mathcal{D}\to\mathcal{R}}$), (WPO) (or (PO)), and (NV). It turns out that they do, and to show this, we prove that the two simple scoring rules work even if we replace (NV) with a more stronger axiom (AN) (which implies (NV)).

Just as we explained in the beginning of 4.3., we will employ an arbitrary tie-breaking system by indexing each alternative in X (i.e., $X = \{x_1, ..., x_{|X|}\}$) and use $r(Y)$ to denote the smallest indexed alternative in Y for every $Y \subseteq X$. To define the plurality rule, for each non-empty $N' \subseteq N$, suppose

[9]When we consider the class of simple scoring rules that Young (1975) characterized, we often assume that the input of SCF is a profile of linear orders, $\succ \in \mathcal{P}$, as all individuals are assumed to use the same score vector: $\boldsymbol{s} = (s_1, ..., s_{|X|})$ such that $s_1 \geq \cdots \geq s_{|X|}$ and $s_1 > s_{|X|}$, where s_r is the score of the rth best alternative for each individual. Under the assumption that the input of SCFs is $\succ$, Llamazares and Peña (2015) shows that the plurality rule with $\boldsymbol{s} = (1, 0, ..., 0)$ and the Borda rule with $\boldsymbol{s} = (|X| - 1, |X| - 2, ..., 0)$ satisfy (WPO). However, the anti-plurality rule with $\boldsymbol{s} = (0, ..., 0, -1)$, the best-worst rule with $\boldsymbol{s} = (1, 0, ..., 0, -1)$, and the $k-$approval voting rules with $k > 1$ and $\boldsymbol{s} = \mathbb{1}_k^\frown \mathbb{0}_{|X|-k}$, where $\mathbb{1}$ includes only one, $\mathbb{0}$ includes only zero, violate (WPO).

$$n^{top}_{N'}(x, \succsim_{N'}) \stackrel{\text{def}}{=} |\{i \in N' \mid x \in X^{top}_{\succsim_i}\}|.$$

That is, $n^{top}_{N'}(x, \succsim_{N'})$ denotes the number of individuals who rank alternative x at the top within N'. Note that we omit N' if $N' = N$, $n^{top}_{\{i\}}(x, \succsim_i) = n^{top}_i(x, \succsim_i)$, and $n^{top}_{N'}(x, \succsim) = \sum_{i \in N'} n^{top}_i(x, \succsim_i)$. We now formally define the plurality rule defined over post-deliberation preferences that are weak orders as follows:

Definition 4 (The plurality rule, f^{pl})**.** $f = f^{pl}$ if and only if, for any $\succsim \in \mathcal{R}$,

$$f(\succsim) = x_{r(\text{argmax}_{x \in X} n^{top}(x, \succsim))}.$$

In words, the plurality rule f^{pl} chooses the alternative that is considered to be the best by the largest number of individuals than any other alternative (and if there are ties, chooses the alternative with the smallest index among such alternatives.)

To formally define the Borda rule, we first define the "(extended) Borda score"[10] of each alternative $x \in X$ for each non-empty $N' \subseteq N$, i.e., $br_{N'}(x, \succsim_{N'})$ as follows:

$$n_{N'}(x, y, \succsim_{N'}) \stackrel{\text{def}}{=} |\{i \in N' \mid x \succsim_i y\}| - |\{i \in N' \mid y \succsim_i x\}|;$$
$$br_{N'}(x, \succsim_{N'}) \stackrel{\text{def}}{=} \sum_{y \neq x} n_{N'}(x, y, \succsim_{N'}).$$

Note that we omit N' if $N' = N$, $n_{\{i\}}(x, y, \succsim_i) = n_i(x, y, \succsim_i)$, $br_{\{i\}}(x, \succsim_i) = br_i(x, \succsim_i)$, and $br_{N'}(x, \succsim) = \sum_{i \in N'} br_i(x, \succsim_i)$. In other words, the Borda score of an alternative $x \in X$ is calculated by: for each individual $i \in N$, adding the number of alternatives that are ranked above x and then subtracting from this the number of alternatives that are ranked below x, and, then, summing this number across all individuals. Further note that $\sum_{x \in X} br_i(x, \succsim_i) = 0$ for all $i \in N$, therefore, $\sum_{x \in X} br(x, \succsim) = 0$. That is, adding up the Borda score of every alternative has to sum to 0. (This is convenient to remember when we check whether we have calculated the Borda scores of each alternative correctly.) We now formally define the Borda rule defined over post-deliberation preferences that are weak orders as follows:

Definition 5 (The Borda rule, f^{br})**.** $f = f^{br}$ if and only if, for any $\succsim \in \mathcal{R}$,

$$f(\succsim) = x_{r(\text{argmax}_{x \in X} br(x, \succsim))}.$$

[10] There are several ways to define the Borda score with weak orders. We employ the extended Borda score since it is the most straightforward extension of the Borda score with linear orders (see Black, 1976; Coughlin, 1979; Gärdenfors, 1973; and Young, 1986 among others).

In other words, the Borda rule f^{br} chooses the alternative that receives the highest Borda score (and if there are ties, chooses the alternative with the smallest index among such alternatives.)[11]

Propositions 4 and 5 show that the Borda rule f^{br} is compatible with (AN), (PO), and (NNRD$_{\mathcal{D}\to\mathcal{R}}$) (and, therefore, with (NV), (WPO), and ((WNNRD$_{\mathcal{D}\to\mathcal{R}}$)) and that the plurality rule f^{br} is compatible with (AN), (WPO), and (NNRD$_{\mathcal{D}\to\mathcal{R}}$) (and, therefore, with (NV), (WPO), and (WNNRD$_{\mathcal{D}\to\mathcal{R}}$)).

Proposition 4. The Borda rule f^{br} satisfies (AN), (PO), and (NNRD$_{\mathcal{D}\to\mathcal{R}}$).

Proposition 5. The plurality rule f^{pl} satisfies (AN), (WPO), and (NNRD$_{\mathcal{D}\to\mathcal{R}}$).

The proofs of these proposition are in Appendices A.5 and A.6, respectively. Propositions 4 and 5 show that it is possible for us to escape the impossibility theorem and find post-deliberation aggregation rules that respect the results of successful deliberation if we restrict our pre-deliberation preference domain to that of dichotomous preferences. A major limitation of this escape route is that it requires every individual to enter democratic deliberation with dichotomous pre-deliberation preferences over the set of alternatives – that is, everybody must be able to partition the set of alternatives into two equivalence classes, viz., alternatives that are considered to be (relatively) 'good' and alternatives that are considered to be (relatively) 'bad.' Of course, the framework allows people to develop and eventually arrive at more refined and sophisticated preferences that are weak orders during the process of deliberation; but those kinds of refined and sophisticated preferences that form weak orders are not permitted to be used as *inputs* for democratic deliberation themselves. Hence, escaping the impossibility theorem by restricting the domain of pre-deliberation preferences to dichotomous preferences may go against deliberative democracy's ideal of "openness" that many deliberative democratic theorists regard as an essential feature for democratic deliberation to perform well. (Miller 1992: 55; Sunstein 2002: 194; Rawls 1997: 93; Mansbridge et al. 2010: 65-66)

Before ending this subsection, we show that not all aggregation rules survive this escape route of restricting the initial pre-deliberation preference domain to dichotomous preferences. In particular, we show that the Kemeny rules is incompatible with (NNRD$_{\mathcal{D}\to\mathcal{R}}$). The reason that we consider the Kemeny rule is because it is one of the standard Condorcet rules that chooses the pairwise majority rule winner if it exists.

[11] Here is how we can consider Borda rule as an instance of a scoring rule. The Borda score vector is denoted by $\boldsymbol{s} = (|X|-1, |X|-2, ..., 0)$ when $\succsim \in \mathcal{P}$. If we arrange the elements of $(br(x, \succ_i))_{x\in X}$ for any $i \in N$ in descending order, we obtain a vector $\boldsymbol{br} = (|X|-1, |X|-3, |X|-5, ..., 5-|X|, 3-|X|, 1-|X|) = (2s_1 - (|X|-1), ..., 2s_{|X|} - (|X|-1))$. Thus, the scoring rules with $\boldsymbol{s} = (|X|-1, |X|-2, ..., 0)$ and $\boldsymbol{br}$ are equivalent. See also Saari (1999, 2000a, 2000b).

To define the Kemeny rule, we must first define the Kemeny distance between two post-deliberation preferences (that are weak orders.) The *Kemeny distance* between two weak orders, $\succsim_\circ$ ($\in \mathcal{R}_\circ$) and $\succsim_i$ ($\in \mathcal{R}_i$), which was introduced by Kemeny (1959) and was characterized by Kemeny and Snell (1962) and Can and Storcken (2018), is defined as follows:

$$KD(\succsim_\circ, \succsim_i) \stackrel{\text{def}}{=} |\ \succsim_\circ \setminus \succsim_i\ | + |\ \succsim_i \setminus \succsim_\circ\ |.$$

Let us try to understand what this is saying. The Kemeny distance between two preferences is a measure of how close (or far apart) the two preferences are. Consider any $x, y \in X$ and any preference of individual $\circ$, $\succsim_\circ$, and any preference of individual i, $\succsim_i$. Here is how the Kemeny distance KD is calculated: (1) if both $\circ$ and i have the same preferences between x and y, for example, if $x \succ_\circ y$ and $x \succ_i y$, then we have $(x,y) \in \succsim_\circ$, $(x,y) \in \succsim_i$ and $(y,x) \notin \succsim_\circ$, $(y,x) \notin \succsim_i$. Hence, we have $(x,y), (y,x) \notin \succsim_\circ \setminus \succsim_i$ and $(x,y), (y,x) \notin \succsim_i \setminus \succsim_\circ$. So, with respect to the two alternatives x and y, the Kemeny distance $KD(\succsim_\circ, \succsim_i)$ will not increase; (2) if one of $\circ$ and i is indifferent between x and y and one has a strict preference between x and y, for example, if $x \sim_\circ y$ and $x \succ_i y$), then we have $(x,y), (y,x) \in \succsim_\circ$ and $(x,y) \in \succsim_i$ and $(y,y) \notin \succsim_i$. Hence, we have $(x,y) \notin \succsim_\circ \setminus \succsim_i$, $(y,x) \in \succsim_\circ \setminus \succsim_i$, and $(x,y), (y,x) \notin \succsim_i \setminus \succsim_\circ$. So, with respect to the two alternatives x and y, the Kemeny distance $KD(\succsim_\circ, \succsim_i)$ will increase by 1 distance; (3) if $\circ$ and i have strict preferences over x and y in the opposite direction, for example, if $x \succ_\circ y$ and $y \succ_i x$, then we have $(x,y) \in \succsim_\circ$, $(x,y) \notin \succsim_i$ and $(y,x) \notin \succsim_\circ$, $(y,x) \in \succsim_i$. Hence, we have $(x,y) \in \succsim_\circ \setminus \succsim_i$, $(y,x) \notin \succsim_\circ \setminus \succsim_i$, and $(y,x) \in \succsim_i \setminus \succsim_\circ$, $(x,y) \notin \succsim_i \setminus \succsim_\circ$. So, with respect to the two alternatives x and y, the Kemeny distance $KD(\succsim_\circ, \succsim_i)$ will increase by 2 distances. In short, the Kemeny distance KD between two identical preferences orders will be 0; it will increase by an increment of 1 for each pair of alternatives for which one individual is indifferent while the other individual has a strict preference; and it will increase by an increment of 2 for each pair of alternatives for which the two individuals have opposite strict preferences.

Utilizing the notion of Kemeny distance KD, we define the *Kemeny rule* as follows:

Definition 6 (The Kemeny rule, f_{kem})**.** For any $\succsim \in \mathcal{R}$, $f = f_{kem}$ if and only if

$$f(\succsim) \in \cup_{\succsim^* \in A} X^{top}_{\succsim^*},\ \text{where } A = \text{argmin}_{\succsim_\circ \in \mathcal{R}_\circ} \sum_{i \in N} KD(\succsim_\circ, \succsim_i).$$

Intuitively, the Kemeny rule constructs a social preference relation so that it minimizes the sum of the Kemeny distance between the social preference relation and each

individual's preference, and, then chooses the alternative that is top-ranked with respect to that newly constructed social preference relation.

The following example illustrates that the Kemeny rule f_{kem} violates (NNRD$_{\mathcal{D}\to\mathcal{R}}$).

Example 6: A Negative Result for the Kemeny Rule Suppose that there are 15 individuals and $X = \{x_1, x_2, x_3\}$. Suppose that individual 13's initial pre-deliberation preference was: $x_1 \sim_{13} x_2 \succ_{13} x_3$ (or $(x_1x_2)x_3$.) We consider two possible cases in which the 15 individuals update and change their preferences after engaging in democratic deliberation: (a) a case in which individual 13 does not change his/her preferences by being persuaded by nobody ($\succsim$), and (b) a case in which individual 13 does change his/her preferences by being positively persuaded by other people.

The following describes the 15 individuals' post-deliberation preferences where individual 13 retains his/her initial pre-deliberation preference after deliberation:

1. $x_1 \succ_i x_2 \succ_i x_3$, $i = 1, 2, 3, 4$;
2. $x_2 \succ_j x_3 \succ_j x_1$, $j = 5, 6, 7, 8$;
3. $x_3 \succ_k x_1 \succ_k x_2$, $k = 9, 10, 11, 12$;
4. $x_1 \sim_{13} x_2 \succ_{13} x_3$;
5. $x_2 \sim_{14} x_3 \succ_{14} x_1$; and
6. $x_3 \sim_{15} x_1 \succ_{15} x_2$.

The following describes the 15 individuals' post-deliberation preferences where individual 13 has also changed his/her preference :

1. $x_1 \succ'_i x_2 \succ'_i x_3$, $i = 1, 2, 3, 4$;
2. $x_2 \succ'_j x_3 \succ'_j x_1$, $j = 5, 6, 7, 8$;
3. $x_3 \succ'_k x_1 \succ'_k x_2$, $k = 9, 10, 11, 12$;
4. $x_1 \sim'_{13} x_2 \succ'_{13} x_3$;
5. $x_2 \sim'_{14} x_3 \succ'_{14} x_1$; and
6. $x_3 \sim'_{15} x_1 \succ'_{15} x_2$.

Note that the only difference between the two post-deliberation preference profiles $\succsim$ and $\succsim'$ is individual 13's preference.

If we apply the Kemeny rule to post-deliberation profile $\succsim$, then we discover that the three linear orders, $x_1 \succ_i x_2 \succ_i x_3$, $x_2 \succ_j x_3 \succ_j x_1$, and $x_3 \succ_k x_1 \succ_k x_2$ are all contained in $A = \text{argmin}_{\succsim_\circ \in \mathcal{R}_\circ} \sum_{i' \in N} KD(\succsim_\circ, \succsim_{i'})$.[12] Hence, we have: $\cup_{\succsim_* \in A} X^{top}_{\succsim_*} = \{x_1, x_2, x_3\}$. From our tie-breaking system r, we have: $r(\{x_1, x_2, x_3\}) = x_1$. Hence, the Kemeny rule chooses x_1 for $\succsim$: i.e., $f_{kem}(\succsim) = x_1$.

Now, if we apply the Kemeny rule to post-deliberation profile $\succsim'$, then we discover that *only* the linear order $x_3 \succ_k x_1 \succ_k x_2$ is contained in $A = \text{argmin}_{\succsim_\circ \in \mathcal{R}_\circ} \sum_{i' \in N} KD(\succsim_\circ, \succsim'_{i'})$.[13] Hence, we have: $\cup_{\succsim_* \in A} X^{top}_{\succsim_*} = \{x_3\}$. So, the Kemeny rule now chooses x_3 for $\succsim'$: i.e., $f_{kem}(\succsim') = x_3$.

In other words, the fact that individual 13 has been persuaded by other people and has changed his/her preference from $(x_1x_2)x_3$ to $x_1(x_2x_3)$ affects the social outcome to change from x_1 to x_3. However, in order for individual 13 to change his/her preferences from $(x_1x_2)x_3$ to $x_1(x_2x_3)$, s/he would have had to be persuaded by those whose preferences were either $x_1(x_2x_3)$ or $(x_1x_3)x_2$. If individual 13 was persuaded by those whose preferences were $x_1(x_2x_3)$, then since $f_{kem}(\succsim') = x_3$ is strictly worse for these people than $f_{kem}(\succsim) = x_1$, the fact that these people successfully persuaded individual 13 during deliberation made them all strictly worse-off. Thus, the Kemeny rule f_{kem} violates (NNRD$_{\mathcal{D}\to\mathcal{R}}$).

From Example 6, we can understand that restricting our pre-deliberation preferences to be dichotomous does not automatically make every aggregation rule compatible with (NNRD$_{\mathcal{D}\to\mathcal{R}}$).

4.4.2 The Possibility of The Approval Voting Rule with a Tie-Breaker When $\mathcal{D} \to \mathcal{D}$

Finally, we show that the possibility result does not go away even if we further restrict the post-deliberation preference profiles to be dichotomous. That is, in this framework, we assume that the individuals' initial *pre*-deliberation preferences and their *post*-deliberation preferences are *both* dichotomous. As explained previously, this may happen when each individual is either completely convinced or completely unconvinced by the opinions and preferences of other individuals during democratic deliberation.

As before, we consider an arbitrary tie-breaking system by assuming $X = \{x_1, ..., x_{|X|}\}$

[12] To see this, note that $\sum_{i' \in N} KD(\succsim_1, \succsim_{i'}) = \sum_{i' \in N} KD(\succsim_5, \succsim_{i'}) = \sum_{i' \in N} KD(\succsim_9, \succsim_{i'}) = 41$, that is, the minimum value.

[13] To see this, note that $KD(\succsim'_{13}, \succsim_1) - KD(\succsim_{13}, \succsim_1) = 0$, $KD(\succsim'_{13}, \succsim_5) - KD(\succsim_{13}, \succsim_5) = 2$, and $KD(\succsim'_{13}, \succsim_9) - KD(\succsim_{13}, \succsim_9) = -2$, $\text{argmin}_{\succsim_\circ \in \mathcal{R}_\circ} \sum_{i' \in N} KD(\succsim_\circ, \succsim'_{i'})$ includes only $\succsim_* = \succ_k$, $k = 9, 10, 11, 12$.

and using the function $r(Y)$ to denote the alternative in Y with the smallest index for every $Y \subseteq X$.

Since our aggregation rule is now applied to post-deliberation preferences, whose domain is the set of dichotomous orders, $\mathcal{D}$, we redefine our (NNRD), (PO), and (AN) axioms accordingly in the following way:

(NNRD$_{\mathcal{D}\to\mathcal{D}}$) : For any $i \in N$, for any $\succsim^0 \in \mathcal{D}$, for any $(C, \boldsymbol{u}) \in \mathcal{C} \times \mathcal{U}_{\succsim^0}$, and for any $\succsim^1 \in \mathcal{D}$ with $C\boldsymbol{u} \in \mathcal{U}_{\succsim^1}$,

$$\exists j \neq i \text{ s.t. } c^i_j > 0 \;\&\; f_{\mathcal{D}}(\succsim^1) \succsim^0_j f_{\mathcal{D}}(\succsim^0_i, \succsim^1_{-i}).$$

(PO$_{\mathcal{D}}$) For any $\succ \in \mathcal{D}$ and for any $x, y \in X$, if $y \succsim_i x$ for all $i \in N$, and there exists $j \in N$ such that $y \succ_i x$, then $f_{\mathcal{D}}(\succsim) \neq x$.

(AN$_{\mathcal{D}}$) : For any $\succsim \in \mathcal{D}$ and for any permutation pi on N, $f_{\mathcal{D}}(\succsim) = f_{\mathcal{D}}((\succsim_{\pi(i)})_{i\in N})$.

We then define the approval voting rule $f^{av}_{\mathcal{D}}$ as follows:

Definition 7 (The approval voting rule, $f^{av}_{\mathcal{D}}$)**.** $f_{\mathcal{D}} = f^{av}_{\mathcal{D}}$ if and only if, for any $\succsim \in \mathcal{D}$,

$$f_{\mathcal{D}}(\succsim) = x_{r(\text{argmax}_{x\in X}|\{i\in N|x\in H_{\succsim_i}\}|)}.$$

In words, the approval voting rule, $f^{av}_{\mathcal{D}}$, chooses the alternative that receives the highest number of approvals (where an alternative $x \in X$ is approved by individual $i \in N$ if and only if $x \in H_{\succsim_i}$) among all individuals (and when there is a tie, chooses the alternative with the smallest index among such alternatives.) Note that if the post-deliberation preferences are dichotomous, then the plurality and Borda rules are equivalent to the approval voting rule. Hence, based on Propositions 4 and 5, we immediately get a possibility for the approval voting rule when post-deliberation preferences are restricted to be dichotomous.

Proposition 6. The approval voting rule $f^{av}_{\mathcal{D}}$ satisfies (AN$_{\mathcal{D}}$), (PO$_{\mathcal{D}}$), and (NNRD$_{\mathcal{D}\to\mathcal{D}}$)

Again, such an escape route comes with costs. As explained previously, when we only allow dichotomous *pre*-deliberation preferences, that has an effect of precluding more fine-grained and sophisticated preferences or opinions from entering democratic deliberation. When we further require people's *post*-deliberation preferences to be dichotomous, we are, in effect, restricting the way in which people's preferences can potentially change or transform during democratic deliberation. As already explained, in

order for dichotomous pre-deliberation preferences to remain dichotomous after deliberation, each individual must either be completely convinced or completely unconvinced by the opinions/preferences of other deliberative participants. In other words, when we impose both pre-deliberation preferences and post-deliberation preferences to be dichotomous, this implies that total conversion or total inertia are the only kinds of preference change/transformation that are allowed during democratic deliberation. This significantly compromises and reduces the *deliberative role* of democratic deliberation itself. This means that the possible escape route considered in this subsection makes it possible for the second stage aggregation rule to respect the NNRD axioms and incorporate the results of successful democratic deliberation only by restricting the role of democratic deliberation and severely limiting the meaning of "successful democratic deliberation" itself.

5 Concluding Remarks

Throughout this paper, we assumed that the participants of the democratic process were sincere and participated in both the deliberative stage and the aggregation stage without any motivation to strategically manipulate the final outcome. Our impossibility theorem shows that even if people are sincerely committed to democratic deliberation and democratic deliberation itself runs successfully, there are few aggregation rules that can properly accommodate the results of such successful deliberation and at the same time respect deliberative democracy's ideal of unanimous consensus and democratic equality. Of course, there are potential escape routes to the impossibility result, but, as we have seen, each potential escape route compromises some core value of democracy.

In this paper, we interpreted the aggregation rules as social choice function that generates a unique social outcome, but we expect to obtain similar results even in the case of social choice correspondence if (NNRD) is properly defined. From the viewpoint of information invariance, the convex combination of utility representations is not ordinal. To be precise, we assume that during the process of democratic deliberation, individuals share a ratio-scale measurable utility when they update their preferences on the basis of the preferences of other participants. However, by applying the Kemeny distance, we can also define a convex combination based on ordinal preferences, which is different from the convex combination of utility representations we utilized in this paper. Under this ordinal notion of convex combinations, we get the same impossibility result. This is because under the ordinal convex combination framework, the scope of (NNRD) is less constrained and thus the NNRD axiom becomes more demanding. This shows that

our impossibility result does not crucially depend on the specific modeling assumptions employed in this paper.

We hope our paper can start a new line of research that investigates the proper normative relationship between democratic deliberation and aggregation.

References

[1] Arrow, Kenneth, J. (1951, 1963), *Social Choice and Individual Values*, Yale University Press.

[2] Austen-Smith, David and Feddersen, Timothy (2006), "Deliberation, Preference Uncertainty, and Voting Rules", *The American Political Science Review*, 100, 209 —217.

[3] Black, Duncan (1958), *The Theory of Committees and Elections*, Cambridge University Press.

[4] Black, Duncan (1976), "Partial justification of the Borda count", *Public Choice*, 28, 1–16.

[5] Bohman, James and Rehg, Williams, eds. (1997), *Deliberative Democracy: Essays on Reason and Politics* , The MIT Press.

[6] Can, Burak and Storcken, Ton (2018), "A re-characterization of the Kemeny distance", *Journal of Mathematical Economics*, 79, 112–116.

[7] Christiano, Thomas (1997), "The Significance of Public Deliberation", contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 243–278.

[8] Chung, Hun and Duggan, John (2020), "A Formal Theory of Democratic Deliberation", *The American Political Science Review*, 114, 14–35.

[9] Cohen, Joshua (1997a), "Deliberation and Democratic Legitimacy" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 67–92.

[10] Cohen, Joshua (1997b), "Procedure and Substance in Deliberative Democracy" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 407–438.

[11] Coughlan, Peter J. (2000), "In Defense of Unanimous Jury Verdicts: Mistrials, Communication and Strategic Voting", *The American Political Science Review*, 94, 375–393.

[12] Coughlin, Peter (1979), "A direct characterization of Black's first Borda count", *Economics Letters*, 4, 131–134.

[13] Dryzek, John S. (2000), *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*, Oxford University Press.

[14] Dryzek, John S. and List, Christian (2003), "Social Choice Theory and Deliberative Democracy: A Reconciliation", *British Journal of Political Science* 33, pp. 1–28.

[15] Elster, Jon (1997), "The Market and the Forum: Three Varieties of Political Theory" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 3–34.

[16] Fishkin, James S. (1995), *The Voice of the People: Public Opinion and Democracy*, Yale University Press.

[17] Gärdenfors, Peter (1973), "Positionalist voting functions", *Theory and Decision*, 4,1–24.

[18] Gaus, Gerald F. (1997), "Reason, Justification, and Consensus: Why Democracy Can't Have It All" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 205–242.

[19] Goodin, Robert. E. (2008), "First Talk, Then Vote" contained in *Innovating Democracy: Democratic Theory and Practice after the Deliberative Turn*, Oxford University Press, pp. 108–124.

[20] Gutman, Amy and Thomson, Dennis (2004), *Why Deliberative Democracy?*, Princeton University Press.

[21] Habermas, Jürgen (1997), "Popular Sovereignty as Procedure" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 35–66.

[22] Kemeny, John G. (1959), "Mathematics without numbers", *Daedalus*, 88, 577–591.

[23] Kemeny, John G. and Snell, James L. (1972), *Mathematical models in the social sciences*, The MIT Press.

[24] Knight, Jack and Johnson, James (1994), "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy", *Political Theory*, 22, 277–296.

[25] Knight, Jack and Johnson, James (1997), "What Sort of Equality Does Deliberative Democracy Require?" contained in *Deliberative Democracy: Essays on Reason and Politics* (edited by J. Bohman and W. Rehg), The MIT Press, pp. 279–320.

[26] Knight, Jack and Johnson, James (2007), "The Priority of Democracy: A Pragmatist Approach to Political-Economic Institutions and the Burdern of Justification", *The American Political Science Review*, 101, 47–61.

[27] List, Christian and Luskin, Robert C. and Fishkin, James S. and McLean, Iain (2013), "Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls", *The Journal of Politics*, 75, 80–95.

[28] Llamazares, Bonifacio and Peña, Teresa (2015), "Scoring rules and social choice properties: some characterizations", *Theory and Decision*, 78, 429–450.

[29] Macike, Gerry (2003), *Democracy Defended*, Cambridge University Press.

[30] Mackie, Gerry (2006), "Does democratic deliberation change minds?", *politics, philosophy & economics*, 5, 279–303.

[31] Mackie, Gerry (2018), "Deliberation and Voting Entwined" contained in *The Oxford Handbook of Deliberative Democracy* (edited by A. Bachtiger, J. S. Dryzek, J. Mansbridge, and M. Warren), Oxford University Press.

[32] Mansbridge, Jane and Bohman, James and Chambers, Simone and Estlund, David and Føllesdal, Andreas and Fung, Archon and Lafont, Cristina and Manin, Bernard and Martí, José L. (2010), "The Place of Self-Interest and the Role of Power in Deliberative Democracy", *The Journal of Political Philosophy*, 18, 64–100.

[33] Maskin, Eric (1999), "Nash Equilibrium and Welfare Optimality", *The Review of Economic Studies*, 66, 23–38.

[34] Mathis, Jérôme (2011), "Deliberation with Evidence", *The American Political Science Review*, 105, 516–529.

[35] McKelvey, Rechard D. (1976), "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control", *Journal of Economic Theory*, 12, 472–482.

[36] McKelvey, Rechard D. (1979), "General Conditions for Global Intransitivities in Formal Voting Models", *Econometrica*, 47, 1085–1112.

[37] Miller, David (1992), "Deliberative Democracy and Social Choice", *Political Studies*, 40, 54–67.

[38] Plott, Charles R. (1967), "A Notion of Equilibrium and its Possibility Under Majority Rule", *The American Economic Review*, 57, 787–806.

[39] Przeworski, Adam (1998), "Deliberation and Ideological Domination" contained in *Deliberative Democracy* (edited by J. Elster), Cambridge University Press.

[40] Rawls, John (1993), *Political Liberalism*, Columbia University Press.

[41] Rawls, John (2005), *A Theory of Justice: Original Edition*, Harvard University Press.

[42] Reny, Philip J. (2001), "Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach", *Economics Letters*, 70, 99–105.

[43] Riker, William (1982), *Liberalism Against Populism*, Waveland Press.

[44] Saari, Donald G. (1999), "Explaining all three-alternative voting outcomes", *Journal of Economic Theory*, 87, 313–355.

[45] Saari, Donald G. (2000a), "Mathematical structure of voting paradoxes I: Pairwise votes", *Economic Theory*, 15, 1–53.

[46] Saari, Donald G. (2000b), "Mathematical structure of voting paradoxes II: Positional voting", *Economic Theory*, 15, 55–102.

[47] Schofield, Norman (1978), "Instability of Simple Dynamic Games", *The Review of Economic Studies*, 45, 575–594.

[48] Sunstein, Cass R. (2002), "The Law of Group Polarization", *Journal of Political Philosophy*, 10, 175–195.

[49] Young, Hobart P. (1975), "Social Choice Scoring Functions", *SIAM Journal on Applied Mathematics*, 28, 824–838.

[50] Young, Hobart P. (1986), "Optimal ranking and choice from pairwise comparisons" contained in *Information Pooling and Group Decision Making* (edited by B. Grofman and G. Owen), JAI Press, pp. 113–122.

Appendix: Proofs

A.1 Proof of Theorem 1

Proof. Take any f satisfying (WP) and (WNNRD) and fix it. We denote the set of linear preferences for i and a linear preference profile by $\mathcal{P}_i$ and $\succ$, respectively. Let $\mathcal{P}^Y = \mathcal{P} \cap \mathcal{R}^Y$.

Suppose that $Adj(\succ_i) \subset X^2$ is a pair of outcomes whose ranks on $\succ_i$ is adjacent. For any $x, y \in X$, let $(x, y) \in Adj(\succ_i)$ if x is the m-th best and b is the $m+1$-th best on $\succ_i$, $m \in \{1, ..., |X| - 1\}$. For each $\succ_i$, and for each $(x, y) \in Adj(\succ_i)$, suppose that $\succ_i^{x \rightleftharpoons y}$ is a linear order obtained by reversing the order of (x, y) on $\succ_i$ (e.g. $x \succ_i^{x \rightleftharpoons y} y \succ_i^{x \rightleftharpoons y} z$ if and only if $y \succ_i x \succ_i z$).

We then prove five lemmas. Lemma 1.1 shows that (WNNRD) implies non-negative response for pairs in $Adj(\succ_i)$.

Lemma 1.1 (Local non-negative response, LNNR)**.** For each $\succ \in \mathcal{P}$, for each $i \in N$, and for each $(x, y) \in Adj(\succ_i)$, and for each $j \neq i$,

$$\begin{cases} f(\succ_i^{x \rightleftharpoons y}, \succ_{-i}) \succeq_j f(\succ) & \text{if } y \succ_j x, \\ f(\succ) \succeq_j f(\succ_i^{x \rightleftharpoons y}, \succ_{-i}) & \text{if } x \succ_j y, \end{cases}$$

where $\succeq$ indicates "$\succ$ or $=$."

Proof. Suppose that $y \succ_j x$. Let $c_k^i = 0$ for $k \neq i, j$, and let $c_j^i > 0$ be extremely small. Consider $u_i \in U_{\succ_i}$ such that $u_i(x) - u_i(y)$ is sufficiently small to make $\boldsymbol{c}^i \boldsymbol{u}$ be in $U_{\succ_i^{x \rightleftharpoons y}}$. By (WNNRD) and the anti-symmetry condition of linear orders, we have that $f(\succ_i^{x \rightleftharpoons y}, \succ_{-i}) \succeq_j f(\succ)$.

Next, suppose that $x \succ_j y$. Let $c_k^i = 0$ for $k \neq i, j$ and $c_j^i > 0$ be extremely small. Consider $u_i \in U_{\succ_i^{x \rightleftharpoons y}}$ such that $u_i(y) - u_i(x)$ is sufficiently small to make $\boldsymbol{c}^i \boldsymbol{u}$ be in $U_{\succ_i}$. By (WNNRD) and the anti-symmetry condition of linear orders, we have that $f(\succ) \succeq_j f(\succ_i^{x \rightleftharpoons y}, \succ_{-i})$. □

Lemma 1.2 shows that LNNR implies the following monotonicity condition.

Lemma 1.2 (Conditional monotonicity, CM)**.** For any non-empty $Y \subseteq X$, for any $\succ \in \mathcal{P}^Y$, for any $i \in N$, and for any $(y, x) \in Adj(\succ_i) \cap Y^2$, if (1) $f(\succ) = x$ and (2) there exists $j \neq i$ such that "$x \succ_j y'$ for each $y' \in Y \setminus \{x\}$," or "$y' \succ_j x$ for each $y' \in Y \setminus \{x\}$," then it follows that $f(\succ_i^{y \rightleftharpoons x}, \succ_{-i}) = x$.

Proof. Suppose that $y' \succ_j x$ for each $y' \in Y \setminus \{x\}$. By (WP), $f(\succ_i^{y \rightrightarrows x}, \succ_{-i}) \in Y$. Thus, $f(\succ_i^{y \rightrightarrows x}, \succ_{-i}) \succeq_j x = f(\succ)$. Since $y \succ_j x$ and $y \succ_i x$, $f(\succ) \succeq_j f(\succ_i^{y \rightrightarrows x}, \succ_{-i}) \neq f(\succ)$ by LNNR. Thus, we obtain that $f(\succ_i^{y \rightrightarrows x}, \succ_{-i}) = f(\succ) = x$.

Similarly, when $x \succ_j y'$ for each $y' \in Y \setminus \{x\}$, we obtain that $f(\succ_i^{y \rightrightarrows x}, \succ_{-i}) = f(\succ) = x$ by (WP) and LNNR. □

For each non-empty $Y \subseteq X$ and $x \in Y$, let $\mathcal{P}_i^{Y,x\cdots} \subset \mathcal{P}_i^Y$ and $\mathcal{P}_i^{Y,\cdots x} \subset \mathcal{P}_i^Y$ be

$$\begin{aligned} \succ_i \in \mathcal{P}_i^{Y,x\cdots} &\iff x \succ_i y \quad \forall y \in Y \setminus \{x\}, \\ \succ_i \in \mathcal{P}_i^{Y,\cdots x} &\iff y \succ_i x \quad \forall y \in Y \setminus \{x\}. \end{aligned}$$

Then, suppose that $\mathcal{P}_i^{Y,\cdots x\cdots} = \mathcal{P}_i^Y \setminus (\mathcal{P}_i^{Y,x\cdots} \cup \mathcal{P}_i^{Y,\cdots x})$.

For each $A \subseteq X$ and $\succ_i, \succ_i' \in \mathcal{P}_i$, $\succ_i =_{|-A} \succ_i'$ if and only if, for each $y, y' \in X \setminus A$, $y \succ_i y' \Leftrightarrow y \succ_i' y'$. Additionally, we use "$=_{|-x}$" instead of "$=_{|-\{x\}}$" for any $x \in X$.

Then, Lemma 1.3 shows that LNNR implies the following invariance condition:

Lemma 1.3 (Conditional invariance, CI)**.** For each non-empty $Y \subseteq X$, for each non-empty $N_1, N_2 \subseteq N$, and for each $x \in Y$, if there exists $\succ \in \mathcal{P}_{N_1}^{Y,x\cdots} \times \mathcal{P}_{N_2}^{Y,\cdots x} \times \mathcal{P}_{N_3}^{Y,\cdots x\cdots}$, where $N_3 = N \setminus (N_1 \cup N_2)$, such that $f(\succ) = x$, then $f(\succ_{N_1 \cup N_2}^*, \succ_{N_3}) = x$ for each $\succ_{N_1 \cup N_2}^* \in \mathcal{P}_{N_1}^{Y,x\cdots} \times \mathcal{P}_{N_2}^{Y,\cdots x}$.

Proof. Let $Y' = Y \setminus \{x\}$ and $\mathcal{P}^{Y,x} = \mathcal{P}_{N_1}^{Y,x\cdots} \times \mathcal{P}_{N_2}^{Y,\cdots x} \times \{\succ_{N_3}\}$. Choose and fix $(i_1, i_2) \in N_1 \times N_2$. We then prove the following two claims.

Claim 1.3.1. For each $x \in X$, for each $\succ \in \mathcal{P}^{Y,x}$, and for each $j \in (N_1 \cup N_2) \setminus \{i_1\}$, if $f(\succ) = x$, then $f(\succ_j^*, \succ_{-j}) = x$ for each $\succ_j^* \in \mathcal{P}_j^{Y,x}$ with $\succ_j^* =_{|-x} \succ_{i_1}$.

Proof. For each $\succ_j' \in \mathcal{P}_j^{Y'}$ with $\succ_{i_1} \neq_{|-x} \succ_j'$, there exists $(y, z) \in Adj(\succ_j') \cap (X \setminus \{x\})^2$ satisfying $z \succ_{i_1} y$. Since $(Y')^2$ is finite, there exists a finite sequence $\{\succ_j^t\}_{t=1}^{t^*}$ such that $\succ_j^1 = \succ_j$, $\succ_j^{t^*} = \succ_j^*$, and for each $t \in \{1, ..., t^* - 1\}$, $\succ_j^{t+1} = \succ_j^{t, y^t \rightleftharpoons z^t}$, where $(y^t, z^t) \in Adj(\succ_j^t) \cap (X \setminus \{x\})^2$ satisfying $z^t \succ_{i_1} y^t$.

Since x is the best for i_1, by LMMR, $f(\succ_j^t, \succ_{-j}) = x$ implies that $f(\succ_j^{t+1}, \succ_{-j}) = x$. Since $f(\succ_j^1, \succ_{-j}) = f(\succ) = x$, we have that $f(\succ_j^{t^*}, \succ_{-j}) = f(\succ_j^*, \succ_{-j}) = x$. □

Claim 1.3.2. For each $\succ \in \mathcal{P}^{Y,x}$ such that $\succ_{i_1} =_{|-x} \succ_{i_2}$, if $f(\succ) = x$, then $f(\succ_{i_1}^*, \succ_{-i_1}) = x$ for each $\succ_{i_1}^* \in \mathcal{P}_{i_1}^{Y,x}$.

Proof. If $\succ_{i_1}^* \neq \succ_{i_1}$, then $\succ_{i_1}^* \neq_{|-x} \succ_{i_1}$. Thus, there exists a finite sequence $\{\succ_{i_1}^t\}_{t=1}^{t^*}$ such that $\succ_{i_1}^1 = \succ_{i_1}^*$, $\succ_{i_1}^{t^*} = \succ_{i_1}$, and for each $t \in \{1, ..., t^* - 1\}$, $\succ_{i_1}^{t+1} = \succ_{i_1}^{t, y^t \rightleftharpoons z^t}$ where $(y^t, z^t) \in Adj(\succ_{i_1}^t) \cap (X \setminus \{x\})^2$ satisfying $z^t \succ_{i_2} y^t$.

Since x is the worst of Y for i_2 , by LNNR, $f(\succ_{i_1}^{t+1}, \succ_{-i_1}) = x$ implies that $f(\succ_{i_1}^{t}, \succ_{-i_1}) = x$. Since $f(\succ_{i_1}^{t^*}, \succ_{-i_1}) = f(\succ) = x$, $f(\succ_{i_1}^{1}, \succ_{-i_1}) = f(\succ_{i_1}^{*}, \succ_{-i_1}) = x$. □

From Claims 1.3.1 and 1.3.2, we have the conclusion as follows: for any $\succ \in \mathcal{P}^{Y,x}$, for any $j \in (N_1 \cup N_2) \setminus \{i_1, i_2\}$,

$$
\begin{aligned}
& f(\succ) = x \\
& \Rightarrow f(\succ_{i_1}, \succ_{i_2}^{i_1}, \succ_{-\{i_1,i_2\}}) = x, \text{ where } \succ_{i_2}^{i_1} \in \mathcal{P}_{i_2}^{Y,x},\ \succ_{i_2}^{i_1} =_{|-x} \succ_{i_1} && \text{(Claim 1.3.1)} \\
& \Rightarrow f(\succ_{i_1}^{*j}, \succ_{i_2}^{i_1}, \succ_{-\{i_1,i_2\}}) = x, \text{ where } \succ_{i_1}^{*j} \in \mathcal{P}_{i_1}^{*},\ \succ_{i_1}^{*j} =_{|-x} \succ_{j}^{*} && \text{(Claim 1.3.2)} \\
& \Rightarrow f(\succ_{i_1}^{*j}, \succ_{i_2}^{i_1}, \succ_{j}^{*}, \succ_{-\{i_1,i_2,j\}}) = x && \text{(Claim 1.3.1)} \\
& \Rightarrow f(\succ_{i_1}^{*j}, \succ_{i_2}^{*j}, \succ_{j}^{*}, \succ_{-\{i_1,i_2,j\}}) = x, , \text{ where } \succ_{i_2}^{*j} \in \mathcal{P}_{i_2}^{Y,x},\ \succ_{i_2}^{*j} =_{|-x} \succ_{i_1}^{*j} && \text{(Claim 1.3.1)} \\
& \Rightarrow f(\succ_{i_1}^{*i_2}, \succ_{i_2}^{*j}, \succ_{j}^{*}, \succ_{-\{i_1,i_2,j\}}) = x, , \text{ where } \succ_{i_1}^{*i_2} \in \mathcal{P}_{i_1}^{Y,x},\ \succ_{i_1}^{*i_2} =_{|-x} \succ_{i_2}^{*} && \text{(Claim 1.3.2)} \\
& \Rightarrow f(\succ_{i_1}^{*i_2}, \succ_{i_2}^{*}, \succ_{j}^{*}, \succ_{-\{i_1,i_2,j\}}) = x && \text{(Claim 1.3.1)} \\
& \Rightarrow f(\succ_{i_1}^{*}, \succ_{i_2}^{*}, \succ_{j}^{*}, \succ_{-\{i_1,i_2,j\}}) = x && \text{(Claim 1.3.2)}
\end{aligned}
$$

□

For each distinct $x, y \in X$, $i \in N$ is *top-two semi-decisive* for (x, y) if and only if

$$f(\succ) = x \quad \text{for } \succ \in \mathcal{P}^{\{x,y\}} \text{ s.t. } x \succ_i y \ \&\ y \succ_j x \quad \forall j \neq i,$$

Let $TSD(x, y) \subseteq N$ be the set of top-two semi-decisive individuals for (x, y).

Lemma 1.4. There exists $i^* \in N$ such that $i^* = \cap_{x,y \in X, x \neq y} TSD(x, y)$.

Proof. We prove that, for any $Y \subseteq X$ with $|Y| = 3$, there exists $i^Y = \cap_{x,y \in Y, x \neq y} TSD(x, y)$ since (1) if there exist i^Y and $i^{Y'}$ such that $|Y \cap Y'| = 2$, it follows that $i^Y = i^{Y'}$, and (2) for any $(x, y), (x', y') \in X^2$ with $x \neq y$ and $x' \neq y'$, we can construct a sequence of 3-subsets of X, $Y^1, Y^2, \ldots, Y^m \subseteq X$, such that $x, y \in Y^1$, $x', y' \in Y^m$, and for each $k \in \{1, 2, \ldots, m-1\}$, $|Y^k \cap Y^{k+1}| = 2$.

Hereafter, we choose any $Y \subseteq X$ with $|Y| = 3$ and fix it in this proof.

Claim 1.4.1. For any distinct $x, y \in Y$, $i \in TSD(x, y)$, $j \neq i$, and $\succ \in \mathcal{P}^Y$, if $x \succ_i y \succ_i z$ and $y \succ_j x \succ_j z$, then $f(\succ) = x$.

Proof. Without loss of generality, let $i = 1$ and $j = 2$. Let $\succ \in \mathcal{P}^Y$ be such that $x \succ_1 y \succ_1 z$ and $y \succ_k x \succ_k z$ for each $k \neq 1$. Since $1 \in TSD(x, y)$, $f(\succ) = x$.

For each $k \neq 1,2$, let $\succ_k^{yzx} \in \mathcal{P}_k^Y$ be such that $y \succ_k^{yzx} z \succ_k^{yzx} x$. For each $k \in \{3,\ldots,|N|-1\}$, if $f(\succ_{\{1,\ldots,k\}}, \succ_{\{k+1,\ldots,|N|\}}^{yzx}) = x$, by (WP) and LNNR, $f(\succ_{\{1,\ldots,k-1\}}, \succ_{\{k,\ldots,|N|\}}^{yzx}) = x$. We thus obtain that $f(\succ_{\{1,2\}}, \succ_{-\{1,2\}}^{yzx}) = x$.

By CM and CI, if any $k \neq 1,2$ changes his/her preference from $\succ_k^{yzx}$ to the others in $\mathcal{P}^Y$, the outcome will be x, therefore, for each $\succ'_{-\{1,2\}} \in \mathcal{P}_{-\{1,2\}}^Y$, $f(\succ_{\{1,2\}}, \succ'_{-\{1,2\}}) = x$. □

Claim 1.4.2. For any $x, y, x', y' \in Y$ with $x \neq x'$, if $TSD(x,y)$ and $TSD(x',y')$ are not empty, then there exists $i \in N$ satisfying $\{i\} \supseteq \bigcap_{v,w \in Y, v \neq w} TSD(v,w)$.

Proof. First, we prove that $\{i\} = TSD(x,y) = TSD(x',y')$. Assume that $i \in TSD(x,y)$, $j \in TSD(x',y')$, and $i \neq j$. Since $|N| \geq 4$, we can find distinct i', j' such that $\{i',j'\} \cap \{i,j\} = \emptyset$ and $\succ \in \mathcal{P}^Y$ with $x \succ_i y \succ_i z$, $y \succ_{i'} x \succ_{i'} z$, $x' \succ_j y' \succ_j z'$, and $y' \succ_{j'} x' \succ_{j'} z'$, where $z \in Y \setminus \{x,y\}$ and $z' \in Y \setminus \{x',y'\}$. From Claim 1.4.1, $f(\succ) = x$ and $f(\succ) = x'$, which contradicts $x \neq x'$.

Next, consider any $(k,v,w) \in N \times Y \times Y$ with $k \in TSD(v,w)$ and assume that $i \neq k$. Since $\{i\} = TSD(x,y) = TSD(x',y')$ and $v \neq x$ or $v \neq x'$, by the same argument as the previous paragraph, we have that $k = i$, that is, the desired result. □

Claim 1.4.3. For any $\{x,y,z\} = Y$, if there exists $\succ \in \mathcal{P}^Y$ such that $x \succ_i y \succ_i z$ and $y \succ_j x \succ_j z$ for each $j \neq i$, and $f(\succ) = x$, then $i \in TSD(x,y)$.

Proof. Choose any $\succ^* \in \mathcal{P}^{\{xy\}}$ with $x \succ_i^* y$ and $y \succ_j^* x$ for each $j \neq i$. Note that $\succ, \succ^* \in \mathcal{P}^{\{x,y\}}$ and $x \succ_k y \Leftrightarrow x \succ_k^* y$ for each $k \in N$. Since $f(\succ^*) = x$, $x \succ_i y$, and $y \succ_j x$ for $j \neq i$, by CI on $\{x,y\}$, we have that $f(\succ^*) = x$. □

Claim 1.4.4. For any distinct $x, y, z \in Y$ and any two N's partitions $N', N'' \subset N$ with $|N'|, |N''| \geq 2$, $N' \cap TSD(x,y) \neq \emptyset$, $N'' \subseteq TSD(x,z)$, $N'' \cap TSD(y,x)$, or $N' \subseteq TSD(y,z)$ holds.

Proof. Let $\succ \in \mathcal{P}^Y$ be such that $x \succ_i y \succ_i z$ and $\succ_i = \succ_j$ for each $i,j \in N'$, and $\succ_k = \succ_i^{x=y}$ for each $k \in N''$. By WP, $f(\succ) \in \{x,y\}$. We will prove that if $f(\succ) = x$, then $N' \cap TSD(x,y) \neq \emptyset$ or $N'' \subseteq TSD(x,z)$. (The proof of the fact that if $f(\succ) = y$, then $N'' \cap TSD(y,x)$, or $N' \subseteq TSD(y,z)$ is symmetric. The conclusion follows from the combination of them.)

Suppose that $f(\succ) = x$ and there exists $j' \in N'' \setminus TSD(x,z)$. We then prove that there exists $i' \in N' \cap TSD(x,y)$. Consider the following sequence starting from $\succ$ (see Table 1).

Table 1: A sequence from $\succ$

profile	label	N'	j'	$N'' \setminus \{j'\}$	f	$\because$
$\succ$		xyz	yxz	yxz	x	assumption
$\succ^{x \rightleftharpoons z}_{N'' \setminus \{j'\}}$		xyz	yxz	yzx	x	(WP) and LNNR
$\succ^{y \rightleftharpoons z}_{N'}$ & $(\succ^{x \rightleftharpoons z}_{N'' \setminus \{j'\}})^{y \rightleftharpoons z}$	$\succ^1$	xzy	yxz	zyx	x	CI
$(\succ^{y \rightleftharpoons z}_{N'})^{x \rightleftharpoons z}$	$\succ^2$	zxy	yxz	zyx	If the outcome is x, then ...	
$\succ^{y \rightleftharpoons x}_{j'}$		zxy	xyz	zyx	x	CM
$(\succ^{y \rightleftharpoons x}_{j'})^{y \rightleftharpoons z}$		zxy	xzy	zyx	x	CI
$((\succ^{x \rightleftharpoons z}_{N'' \setminus \{j'\}})^{y \rightleftharpoons z})^{y \rightleftharpoons x}$	$\succ^3$	zxy	xzy	zxy	x	CM

From Table 1, we obtain that $f(\succ^1) = x$. Additionally, if $f(\succ^2) = x$, then $f(\succ^3) = x$. By Claim 1.4.3, $j' \in \mathit{TSD}(x, z)$, which is a contradiction. Thus, $f(\succ^1) = x \neq f(\succ^2)$.

$f(\succ^1) \neq f(\succ^2)$ implies that there must exist $i' \in N'$ who changes the outcome from x to z (y cannot be the outcome from CI) when we change the preference of $i \in N'$ from $\succ^1_i$ to $\succ^2_i$ one by one since $\succ^1_{N''} = \succ^2_{N''}$. That is, there exists $i' \in N'$ and (possibly empty and) disjoint $N'_{i' \ldots}, N'_{\ldots i'} \subset N'$ with $N'_{i' \ldots} \cup \{i'\} \cup N'_{\ldots i'} = N'$ such that $f(\succ^1_{N'_{i' \ldots} \cup \{i'\}}, \succ^2_{N'_{\ldots i'}}, \succ^1_{N''}) = x \neq f(\succ^1_{N'_{i' \ldots}}, \succ^2_{\{i'\} \cup N'_{\ldots i'}}, \succ^1_{N''})$ (see Table 2).

Table 2: A sequence from $\succ^1$

profile	$N'_{i' \ldots}$	i'	$N'_{\ldots i'}$	j'	$N'' \setminus \{j'\}$	f	$\because$
$\succ^1$	xzy	xzy	xzy	yxz	zyx	x	
$\vdots$							
$\succ^{1, x \rightleftharpoons z}_{N'_{\ldots i'}} = \succ^2_{N'_{\ldots i'}}$	xzy	xzy	zxy	yxz	zyx	x	$f(\succ^1) \neq f(\succ^2)$
$\succ^{1, x \rightleftharpoons z}_{i'} = \succ^2_{i'}$	xzy	zxy	zxy	yxz	zyx	z	$f(\succ^1) \neq f(\succ^2)$ & CI
$\vdots$							

Finally, we check whether i' is the TSD individual for (x, y) when $f(\succ^x) \overset{\text{def}}{=} f(\succ^1_{N'_{i' \ldots} \cup \{i'\}}, \succ^2_{N'_{\ldots i'}}, \succ^1_{N''}) = x$ and $f(\succ^z) \overset{\text{def}}{=} f(\succ^1_{N'_{i' \ldots}}, \succ^2_{\{i'\} \cup N'_{\ldots i'}}, \succ^1_{N''}) = z$ (see Tables 3 and 4). From Tables 3 and 4, we have that $i' \in \mathit{TSD}(x, y)$.

Table 3: Sequences from $\succ^x$ and $\succ^z$

	$N'_{i'\ldots}$	i'	$N'_{\ldots i'}$	j'	$N'' \setminus \{j'\}$		$N'_{i'\ldots}$	i'	$N'_{\ldots i'}$	j'	$N'' \setminus \{j'\}$
$\succ^x$	xzy	xzy	zxy	yxz	zyx	$\succ^z$	xzy	zxy	zxy	yxz	zyx
$\succ^{x_1}$	xyz	xzy	zxy	yxz	zyx	$\succ^{z_1}$	xyz	zxy	zxy	yxz	zyx
$\succ^{x_2}$	yxz	xzy	zxy	yxz	zyx	$\succ^{z_2}$	yxz	zxy	zxy	yxz	zyx
$\succ^{x_3}$	yxz	xzy	zyx	yxz	zyx	$\succ^{z_3}$	yxz	zyx	zxy	yxz	zyx
$\succ^{x_4}$	yxz	xyz	yzx	yxz	yzx						
$\succ^*$	yxz	xyz	yxz	yxz	yxz						

Note: $f(\succ^x) = f(\succ^{x_1}) = \cdots = f(\succ^{x_4}) = f(\succ^*) = x$ and $f(\succ^z) = f(\succ^{z_1}) = \cdots = f(\succ^{z_3}) = z$.

Table 4: Some sequences from $\succ^x$ or $\succ^z$ to $\succ^*$

profile	label	f	$\because$		profile	label	f	$\because$
$\succ^x$		x			$\succ^z$		z	
$\downarrow$					$\downarrow$			
$\succ^{x,z\rightleftharpoons y}_{N'_{i'\ldots}}$	$\succ^{x_1}$	x	CI	$\rightarrow$	$\succ^{x,z\rightleftharpoons y}_{N'_{i'\ldots}}$ or $\succ^{x_1,x\rightleftharpoons z}_{i'}$	$\succ^{z_1}$	z	CI
$\downarrow$					$\downarrow$			
$\succ^{x_1,x\rightleftharpoons y}_{N'_{i'\ldots}}$ or $\succ^{z_2,z\rightleftharpoons x}_{i'}$	$\succ^{x_2}$	x	CI	$\leftarrow$	$\succ^{z_1,x\rightleftharpoons y}_{N'_{i'\ldots}}$	$\succ^{z_2}$	z	CI
$\downarrow$					$\downarrow$			
$\succ^{x_2,x\rightleftharpoons y}_{N'_{\ldots i'}}$ or $\succ^{z_3,z\rightleftharpoons x}_{i'}$	$\succ^{x_3}$	x	CI	$\leftarrow$	$\succ^{z_2,x\rightleftharpoons y}_{N'_{\ldots i'}}$	$\succ^{z_3}$	z	CI
$\downarrow$								
$\succ^{x_3,z\rightleftharpoons y}_{\{i'\}\cup N'_{\ldots i'}\cup(N''\setminus\{j'\})}$	$\succ^{x_4}$	x	CI					
$\downarrow$								
$\succ^{x_4,z\rightleftharpoons x}_{N'_{\ldots i'}\cup(N''\setminus\{j'\})}$	$\succ^*$	x	CM					

□

Let $\{a, b, c\} = Y$. Consider Claim 1.4.4 when $(x, y, z) = (a, b, c)$. Then, there exists $i \in N$ and $c' \in Y$ such that $i \in TSD(a, c') \cup TSD(b, c')$.

Suppose that $i \in TSD(a, c')$. Let $b' \in Y \setminus \{a, c'\}$. Consider Claim 1.4.4 when $(x, y, z) = (b', c', a)$. Then, there exists $i' \in N$ and $a' \in Y$ such that $i' \in TSD(b', a') \cup TSD(c', a')$. Since $b', c' \neq a$, by Claim 1.4.2, there exists $i^Y \in N$ such that $\{i^Y\} \supseteq \cap_{x',y' \in Y, x' \neq y'} TSD(x', y')$. For the case when $i \in TSD(b, c')$, we can obtain the same conclusion by considering Claim 1.4.4 when $(x, y, z) = (a'', c', b)$ where $a'' = Y \setminus \{c', b\}$.

Choose any $\{x', y', z'\} = Y$. Consider Claim 1.4.4 when $(x, y, z) = (x', y', z')$ and $i^Y \in N'$. Note that $N' \setminus \{i^Y\} \neq \emptyset$ since $|N'| \geq 2$. The conclusion of Claim 1.4.4 is consistent with by the property of i^Y only when $N' \cap TSD(x', y') = \{i^Y\}$. Thus,

$i^Y \subseteq TSD(x', y')$ for any distinct $x', y' \in Y$. □

Finally, from Lemmas 1.1–1.4, we complete the proof of Theorem 1.

Without loss of generality, let $n = i^*$ of Lemma 1.4. Choose any non-empty $Y \subseteq X$ with $|Y| \geq 2$, and $x \in X$. If $x \in X \setminus Y$, by (WP), $f(\succsim) \neq x$ for each $\succsim \in \mathcal{R}^Y$.

Suppose that $x \in Y$. Let $\succsim_n \in \mathcal{P}_n^Y \subset \mathcal{R}_n^Y$ be such that $x \succ_n y$ for each $y \in Y \setminus \{x\}$. By (WP), $f(\succsim) \in Y$. For any $\succsim_{-n} \in \mathcal{R}_{-n}^Y$, suppose that $f(\succsim) = x$. Choose any $\boldsymbol{u} \in \mathcal{U}_{\succsim}$. By choosing sufficiently small $\epsilon > 0$ such that

$$\frac{\epsilon}{1-\epsilon}(\max_{x' \in X} u_n(x) - \min_{x' \in X} u_n(x)) < \min_{x', x'' \in X,\ i \neq n \text{ s.t. } u_i(x') \neq u_i(x'')} |u_i(x') - u_i(x'')|,$$

for each $i \neq n$, $u_i' = (u_i - \epsilon u_n)/(1-\epsilon)$ satisfies that, for each $x', x'' \in X$,

$$[u_i'(x') = u_i'(x'') \Rightarrow x' = x''] \ \& \ [(x', x'') \in Y \times (X \setminus Y) \Rightarrow u_i'(x') > u_i'(x'')].$$

Thus, $\succsim_{-n}' \in \mathcal{P}_{-n}^Y$, where $\boldsymbol{u}_{-n}' \in \mathcal{U}_{\succsim_{-n}'}$.

For each $i \neq n$, suppose that $c_i^i = 1 - \epsilon$, $c_n^i = \epsilon$, and $c_j^i = 0$ for each $j \neq n, i$. Then, $\boldsymbol{c}^i \boldsymbol{u}' = u_i$. By (WNNRD),

$$(x =) f(\succsim) \succsim_n f(\succsim_i', \succsim_n, \succsim_{-\{i,n\}}) \succsim_n f(\succsim_i', \succsim_{i'}', \succsim_n, \succsim_{-\{i,i',n\}}) \succsim_n \cdots \succsim_n f(\succsim_n, \succsim_{-n}').$$

By (WP), $f(\succsim_n, \succsim_{-n}') \in Y$. By definition of $\succsim_n$, $f(\succsim_n, \succsim_{-n}') = x$.

Let $y \in Y$ be $y \succ_n x'$ for each $x' \neq y$. For each $i \neq n$, let $\succsim_i'^x, \succsim_i'^{xy} \in \mathcal{P}_i^Y \subset \mathcal{R}_i^Y$ be

$$\forall v, w \in X \setminus \{x\}, \quad [x \succ_i'^x v] \ \& \ [v \succsim_i'^x w \Leftrightarrow v \succsim_i' w]$$
$$\forall v, w \in X \setminus \{x, y\}, \quad [x \succ_i'^{xy} y \succ_i'^{xy} v] \ \& \ [v \succsim_i'^{xy} w \Leftrightarrow v \succsim_i' w].$$

By CM, $f(\succsim_n \succsim_{-n}'^x) = x$. By CI, $f(\succsim_n, \succsim_{-n}'^{xy}) = x$. Let $\succsim_n^{yx}$ be

$$\forall v, w \in X \setminus \{x, y\}, \quad [y \succ_n^{yx} x \succ_n^{yx} v] \ \& \ [v \succsim_n^{yx} w \Leftrightarrow v \succsim_n w].$$

By CM, $f(\succsim_n^{yx}, \succsim_{-n}'^{xy}) = x$, which contradicts the assumption that $n = i^*$ in Lemma 1.4 since $(\succsim_n^{yx}, \succsim_{-n}'^{xy}) \in \mathcal{P}$.

Thus, there exists $\succsim_n \in \mathcal{R}_n^Y$ such that $f(\succsim) \neq x$ for any $(\succsim_{-n}, x) \in \mathcal{R}_{-n}^Y \times X$. □

A.2 Proof of Proposition 1

Proof. (WP): Choose any $x', y' \in X$. Consider $\succsim \in \mathcal{R}$ such that $y' \succ_i x'$ for each $i \in N$ and $f^x(\succsim) = x'$. By definition, $x' \neq x$. Since $f^x(\succsim) = x'$, $(y' \succ_i) x' \succ_i x$ for each $i \in N$.

Thus, $f^x(\succsim) = Mj(\succsim) = y' \neq x'$, which is a contradiction.

(NV): Choose any $i \in N$ and let $Y = \{y, z\}$. Let $\succsim_{-i} \in \mathcal{R}_{-i}^Y$ be such that $z \succ_j y \succ_j x$ for each $j \neq i$. By definition, for each $\succsim_i \in \mathcal{R}_i^Y$, $f^x(\succsim) = Mj(\succsim) = z$.

(WNNRD): Choose any $i \in N$, $\succsim \in \mathcal{R}$, $\succsim_i' \in \mathcal{R}_i$, $\boldsymbol{u} \in \mathcal{U}_{\succsim}$, and $\boldsymbol{c}^i \in \mathcal{C}^i$ such that $\boldsymbol{c}^i u \in U_{\succsim_i'}$ and $\boldsymbol{c}_i^i \neq 1$.If $f^x(\succsim) = f^x(\succsim_i', \succsim_{-i})$, then $f^x(\succsim_i', \succsim_{-i}) \sim_j f^x(\succsim)$ for each $j \neq i$. If $f^x(\succsim) = x$, then $f^x(\succsim_i', \succsim_{-i}) \succ_j x = f^x(\succsim)$ for each $j \neq i$. For each case, the desired conclusion holds. In the following, suppose that $f^x(\succsim) \notin \{f^x(\succsim_i', \succsim_{-i}), x\}$. Let $x' = f^x(\succsim)$ and $x'' = f^x(\succsim_i', \succsim_{-i})$.

By definition of f^x, $x' \succ_k x$ for each $k \in N$. Since $\boldsymbol{u} \in \mathcal{U}_{\succsim}$ and $\boldsymbol{c}^i\boldsymbol{u} \in \mathcal{U}_{\succsim_i'}$, $x' \succ_i' x$. Thus, $(\succsim_i', \succsim_{-i})$ does not belong to the first case of f^x's definition. Thus, $x'' = f^x(\succsim_i', \succsim_{-i}) \neq x$. Since $x'' \neq x'$, then $(\succsim_i', \succsim_{-i})$ belongs to the fourth case of f^x's definition. That is, $x'' \succ_j x$ for each $j \neq i$ and $x'' = Mj(\succsim_i', \succsim_{-i})$.

Suppose $x'' \neq Mj(\succsim)$. Then, since $\succsim$ and $(\succsim_i', \succsim_{-i})$ are different only in i's preference, it follows from definition of Mj that $x' \succsim_i x''$ and $x'' \succsim_i' x'$. Since $\boldsymbol{u} \in \mathcal{U}_{\succsim}$, $\boldsymbol{c}^i\boldsymbol{u} \in \mathcal{U}_{\succsim_i'}$ and $\boldsymbol{c}_i^i \neq 1$, there exists $j \neq i$ with $\boldsymbol{c}_j^i > 0$ and $x'' \succsim_j x'$. That is, $f^x(\succsim_i', \succsim_{-i}) \succsim_j f^x(\succsim)$.

Suppose $x'' = Mj(\succsim)$. If $x'' \succ_i x$, since $x'' \succ_j x$ for each $j \neq i$, then $f^x(\succsim) = x'' = f^x(\succsim_i', \succsim_{-i})$, which is a contradiction. Thus, $x \succsim_i x''$. Since $x' \succ_i x$, we have $x' \succ_i x''$. Since $x'' = Mj(\succsim)$, it follows from definition of Mj that $x'' \succsim_j x'$ for each $j \neq i$. That is, $f^x(\succsim_i', \succsim_{-i}) \succsim_j f^x(\succsim)$ for each $j \neq i$. □

A.3 Proof of Proposition 2

Proof. (TU): It is obvious since $f^d(\succsim) = x$ for any $\succsim \in \mathcal{R}$ and $x \in X$ whenever $x \succ_i y$ for any $y \in X \setminus \{y\}$ and any $i \in N$.

(NV): Choose any $i \in N$ and $\succsim_{-i} \in \mathcal{R}$ such that $d \succ_j x$ for each $x \in X \setminus \{d\}$ and $j \in N \setminus \{i\}$. Then $f^d(\succsim) = d$ for any $\succsim_i \in \mathcal{R}_i$. Thus, any i is not the vetoer.

(WNNRD): Consider any $i \in N$, $\succsim \in \mathcal{R}$, $\succsim_i' \in \mathcal{R}_i$, $\boldsymbol{u} \in \mathcal{U}_{\succsim}$, $\boldsymbol{c}^i \in \mathcal{C}^i$ such that $u_i' \overset{\text{def}}{=} \boldsymbol{c}^i\boldsymbol{u} \in \mathcal{U}_{\succsim_i'}$ and $f^d(\succsim) \neq f^d(\succsim_i', \succsim_{-i})$. If $x = f^d(\succsim) \neq d$, then $x \succ_j y$ for each $j \in N$ and $y \neq x$.Since $u_i' = \boldsymbol{c}^i\boldsymbol{u}$, $u_i'(x) > u_i'(y)$ for any $y \in X$. Thus, $x \succ_i' y$ for any $y \neq x$. Therefore, $f^d(\succsim_i', \succsim_{-i}) = x = f^d(\succsim)$, which is a contradiction. Thus, $f^d(\succsim) = d$.

Since $f^d(\succsim_i', \succsim_{-i}) \neq f^d(\succsim) = d$, it follows that $f^d(\succsim_i', \succsim_{-i}) \succ_j f^d(\succsim)$ for each $j \neq i$. □

A.4 Proof of Proposition 3

Proof. (WP): Consider any $\succsim \in \mathcal{R}$. If $\succsim$ belongs to Case 1, $f^v(\succsim) \succsim_v x$ for each $x \in X$. If $\succsim$ belongs to Case 3, since all alternatives are indifferent for all individuals, $f^v(\succsim) \succsim_i x$

for any $i \in N$ and for any $x \in X$. Suppose that $\succsim$ belongs to Case 2. Since $f^v(\succsim) = x_{r(X^{sec}_{\succsim_v})}$, for each $x \in X$ such that $x \succ_v f^v(\succsim)$, $x \in X^{top}_{\succsim_v} \subseteq X^{bot}_{\succsim_{-v}}$. Thus, $f^v(\succsim) \succsim_i x$ for any $i \neq v$ and for any $x \in X^{top}_{\succsim_v} \subseteq X^{bot}_{\succsim_{-v}}$.

(ND): For each $i \in N$, let $\succsim^i \in \mathcal{R}$ be such that $x_1 \succ^i_i x_2 \succ^i_i \cdots \succ^i_i x_{|X|}$ and $x_{|X|} \succ^i_j x_{|X|-1} \succ^i_j \cdots \succ^i_j x_1$ for $j \neq i$. If $i = v$, then $f^v(\succsim) = x_{r(X^{sec}_{\succsim_v})} = x_2 \neq x_1$. If $i \neq v$, since $x_{|X|} \in X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}}$, it follows that $f(\succsim) = x_{|X|} \neq x_1$.

(WNNRD): Consider any $(i, \succsim) \in N \times \mathcal{R}$, any $(\boldsymbol{u}, \boldsymbol{c}^i) \in \mathcal{U}_{\succsim} \times \mathcal{C}^i$, and any $\succsim'_i \in \mathcal{R}_i$ such that $u'_i = \boldsymbol{c}^i \boldsymbol{u} \in \mathcal{U}_{\succsim'_i}$ and $f^v(\succsim) \neq f^v(\succsim'_i, \succsim_{-i})$. We will prove that there exists $j \neq i$ with $c^i_j > 0$ such that $f^v(\succsim'_i, \succsim_{-i}) \succsim_j f^v(\succsim)$.

If $X^{bot}_{\succsim_j} = X$ for some $j \neq i$ with $c^i_j > 0$, it is obvious that $f^v(\succsim'_i, \succsim_{-i}) \succsim_j f^v(\succsim)$. Then, suppose that $X^{bot}_{\succsim_j} \neq X$ for each $j \neq i$ with $c^i_j > 0$. This assumption implies that, for each $\succsim''_i \in \mathcal{R}$, $(\succsim''_i, \succsim_{-i})$ does not belong to Case 3. There are two possibilities: (a) $i = v$ or (b) $i \neq v$.

(a) Suppose that $i = v$. If $f^v(\succsim) \in X^{bot}_{\succsim_{-v}}$, then $f^v(\succsim'_v, \succsim_{-v}) \succsim_j f^v(\succsim)$ for each $j \neq v$, which implies the desired conclusion.

Consider that $f^v(\succsim) \notin X^{bot}_{\succsim_{-v}}$. By way of contradiction, suppose that $f^v(\succsim'_v, \succsim_{-v}) \in X^{bot}_{\succsim_{-v}}$, which implies that $(\succsim'_v, \succsim_{-v})$ belongs to Case 2. Thus, $u'_v(x) > u'_v(f^v(\succsim'_v, \succsim_{-v}))$ for some $x \in X^{bot}_{\succsim_{-v}}$, and $u'_v(x') \leq u'_v(f^v(\succsim'_v, \succsim_{-v}))$ for every $x' \notin X^{bot}_{\succsim_{-v}}$. Since $f^v(\succsim) \notin X^{bot}_{\succsim_{-v}}$, $u'_v(f^v(\succsim)) \leq u'_v(f^v(\succsim'_v, \succsim_{-v}))$. Since $\{f^v(\succsim'_v, \succsim_{-v}), x\} \subseteq X^{bot}_{\succsim_{-v}}$, for each $j \neq v$, $u_j(x) = u_j(f^v(\succsim'_v, \succsim_{-v})) \leq u_j(f^v(\succsim))$. From $u'_v = \boldsymbol{c}^v \boldsymbol{u}$ and $u'_v(x) > u'_v(f^v(\succsim'_v, \succsim_{-v}))$,

$$u_v(x) > u_v(f^v(\succsim'_v, \succsim_{-v})) \tag{1}$$

$$u'_v(f^v(\succsim)) - u'_v(f^v(\succsim'_v, \succsim_{-v}) \geq c^v_v[u_v(f^v(\succsim)) - u_v(f^v(\succsim'_v, \succsim_{-v}))] \tag{2}$$

By (1) and the definition of f^v, $f^v(\succsim)$ must satisfy $u_v(f^v(\succsim)) \geq u_v(f^v(\succsim'_v, \succsim_{-v}))$. From (2) and $u'_v(f^v(\succsim)) \leq u'_v(f^v(\succsim'_v, \succsim_{-v}))$, $u_v(f^v(\succsim)) = u_v(f^v(\succsim'_v, \succsim_{-v}))$ and $u'_v(f^v(\succsim)) = u'_v(f^v(\succsim'_v, \succsim_{-v}))$. Thus, the index of $f^v(\succsim'_v, \succsim_{-v})$ is smaller than that of $f^v(\succsim)$. By (1), $u_v(f^v(\succsim)) = u_v(f^v(\succsim'_v, \succsim_{-v}))$ implies that $u_v(x) > u_v(f^v(\succsim))$, therefore, $\succsim$ also belongs to Case 2. Thus, the index of $f^v(\succsim)$ is smaller than that of $f^v(\succsim'_v, \succsim_{-v})$, which is a contradiction.

From the above result, when $f^v(\succsim) \notin X^{bot}(\succsim_{-v})$, $f^v(\succsim'_v, \succsim_{-v}) \notin X^{bot}(\succsim_{-v})$ holds. By definition of f^v, it follows that $u_v(f^v(\succsim)) \geq u_v(f^v(\succsim'_v, \succsim_{-v}))$ and $u'_v(f^v(\succsim'_v, \succsim_{-v})) \geq u'_v(f^v(\succsim))$. Since $u'_v = \boldsymbol{c}^v \boldsymbol{u}$, for some $j \neq v$ such that $c^v_j > 0$, $u_j(f^v(\succsim'_v, \succsim_{-v})) \geq u_j(f^v(\succsim))$, therefore, $f^v(\succsim'_v, \succsim_{-v}) \succsim_j f^v(\succsim)$.

(b) Suppose that $i \neq v$. Then, we will prove that (b-1) $f^v(\succsim) \in X^{top}_{\succsim_v} \cap X^{bot}_{\succsim_{-\{i,v\}}}$ and $f^v(\succsim) \in X^{bot}_{\succsim'_i} \setminus X^{bot}_{\succsim_i}$, or (b-2) $f^v(\succsim'_i, \succsim_{-i}) \in X^{top}_{\succsim_v} \cap X^{bot}_{\succsim_{-\{i,v\}}}$ and $f^v(\succsim'_i, \succsim_{-i}) \in X^{bot}_{\succsim_i} \setminus X^{bot}_{\succsim'_i}$.

Before proving that, we introduce additional notations: Let $X^{\text{Case 1}} = X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim_{-v}}$ and $X^{\text{Case 1}'} = X^{top}_{\succsim_v} \setminus X^{bot}_{\succsim'_i, \succsim_{-\{i,v\}}}$. Note that (b-1) and (b-2) imply that $f^v(\succsim) \in X^{\text{Case 1}} \setminus X^{\text{Case 1}'}$ and $f^v(\succsim'_i, \succsim_{-i}) \in X^{\text{Case 1}'} \setminus X^{\text{Case 1}}$, respectively.

If $X^{\text{Case 1}} = X^{\text{Case 1}'} = \emptyset$, since $\succsim$ and $(\succsim'_i, \succsim_{-i})$ are not Case 3, then $f^v(\succsim) = f^v(\succsim'_i, \succsim_{-i}) = x_{r(X^{sec}_{\succsim_v})}$, which is a contradiction. Thus, $X^{\text{Case 1}} \neq \emptyset$ or $X^{\text{Case 1}'} \neq \emptyset$.

Note that $X^{\text{Case 1}} \neq \emptyset$ if and only if $f^v(\succsim) \in X^{\text{Case 1}}$. (Similarly, $X^{\text{Case 1}'} \neq \emptyset$ if and only if $f^v(\succsim'_i, \succsim_{-i}) \in X^{\text{Case 1}'}$).)

Suppose that $X^{\text{Case 1}} = \emptyset$, then $X^{\text{Case 1}'} \neq \emptyset$. Thus, $f^v(\succsim'_i, \succsim_{-i}) \in X^{\text{Case 1}'} \setminus X^{\text{Case 1}}$. In the same manner, if $X^{\text{Case 1}'} = \emptyset$, since $X^{\text{Case 1}} \neq \emptyset$, it follows that $f^v(\succsim) \in X^{\text{Case 1}} \setminus X^{\text{Case 1}'}$.

Suppose that $X^{\text{Case 1}} \neq \emptyset$ and $X^{\text{Case 1}'} \neq \emptyset$. Then, if the index of $f^v(\succsim'_i, \succsim_{-i})$ is smaller than that of $f^v(\succsim)$, it follows from the definition of $f^v(\succsim)$ that $f^v(\succsim'_i, \succsim_{-i}) \in X^{\text{Case 1}'} \setminus X^{\text{Case 1}}$. If the index of $f^v(\succsim)$ is smaller than that of $f^v(\succsim'_i, \succsim_{-i})$, from the definition of $f^v(\succsim'_i, \succsim_{-i})$, it follows that $f^v(\succsim) \in X^{\text{Case 1}} \setminus X^{\text{Case 1}'}$.

By using the above properties of $X^{\text{Case 1}}$ and $X^{\text{Case 1}'}$, we consider the cases of (b-1) and (b-2), respectively.

First, we consider (b-1). Since $f^v(\succsim) \in X^{bot}_{\succsim_{-\{i,v\}}}$, if $c^i_j > 0$ for some $j \neq v, i$, then $f^v(\succsim'_i, \succsim_{-i}) \succsim_j f^v(\succsim)$. Suppose that $c^i_v > 0$ and $c^i_j = 0$ for any $j \neq v, i$. If $c^i_i = 0$, then $X^{bot}_{\succsim'_i} = X^{bot}_{\succsim_v}$. Since $f^v(\succsim) \in X^{top}_{\succsim_v} \neq X$, $f^v(\succsim) \notin X^{bot}_{\succsim_v} = X^{bot}_{\succsim_i}$, which is a contradiction. Suppose $c^i_i > 0$. Since $f^v(\succsim) \notin X^{bot}_{\succsim_i}$, $u_i(f^v(\succsim)) > u_i(x)$ for any $x \in X^{bot}_{\succsim_i}$. Since $f^v(\succsim) \in X^{top}_{\succsim_v}$, $u_v(f^v(\succsim)) \geq u_v(x)$ for any $x \in X^{bot}_{\succsim_i}$. Since $u'_i = \boldsymbol{c}^i \boldsymbol{u}$, $u'_i(f^v(\succsim)) > u'_i(x)$ for any $x \in X^{bot}_{\succsim_i}$. Thus, $f^v(\succsim) \notin X^{bot}_{\succsim'_i}$, which is a contradiction.

Secondly, consider (b-2). Since $f^v(\succsim'_i, \succsim_{-i}) \in X^{top}_{\succsim_v}$, $f^v(\succsim'_i, \succsim_{-i}) \succsim_v f^v(\succsim)$. Thus, the desired conclusion holds if $c^i_v > 0$. Suppose that $c^i_v = 0$. Since $f^v(\succsim'_i, \succsim_{-i}) \in X^{bot}_{\succsim_{-v}}$, $u_j(x) \geq u_j(f^v(\succsim'_i, \succsim_{-i}))$ for any x and $j \neq v$. Since $u'_i = \boldsymbol{c}^i \boldsymbol{u}$, $u'_i(x) \geq u'_i(f^v(\succsim'_i, \succsim_{-i}))$. Thus, $f^v(\succsim'_i, \succsim_{-i}) \in X^{bot}_{\succsim'_i}$, which is a contradiction. □

A.5 Proof of Proposition 4

Proof. (AN): From the definition of br, for any $\succ \in \mathcal{R}$, for any $x \in X$, and for any permutation π on N, $br(x, \succsim) = br(x, (\succsim_{\pi(i)})_{i \in N})$. Thus, $f^{br}(\succsim) = f^{br}((\succsim_{\pi(i)})_{i \in N})$.

(PO): Suppose that there exist $\succsim \in \mathcal{R}$ and $x \in X$ such that (1) $x \succsim_i f^{br}(\succsim)$ for each $i \in N$, and (2) for some $j \in N$, $x \succ_j f^{br}(\succsim)$.

By (1), for each $x' \in X$ and for each $i \in N$, $f^{br}(\succsim) \succsim_i x'$ implies that $x \succsim_i x'$, and $x' \succsim_i x$ implies that $x' \succsim_i f^{br}_{\mathcal{R}}(R)$. By (2), $x \succ_j f^{br}(\succsim)$, therefore, $f^{br}(\succsim) \succsim_j x$ does not hold. Thus, $br(x, \succsim) > br(f^{br}(\succsim), \succsim)$, which contradicts the definition of f^{br}.

(NNRD$_{\mathcal{D}\to\mathcal{R}}$): Take any $(\succsim, i, C) \in \mathcal{D} \times N \times \mathcal{C}$ such that $\boldsymbol{c}^k\boldsymbol{u} \in \mathcal{U}_{\succsim'_k}$ and $\succsim'_k \in \mathcal{R}_k$ for all $k \in N$, and $\boldsymbol{u} \in \mathcal{U}_{\succsim^0}$. Suppose that $f^{br}(\succsim_i, \succsim'_{-i}) = x$, and $f^{br}(\succsim') = x'$.

By way of contradiction, assume that $x \neq x'$ and $x \succ_j x'$ for all $j \neq i$ such that $c^i_j > 0$. Since $\succsim \in \mathcal{D}$,

$$\forall j \neq i \text{ s.t. } c^i_j > 0, \quad (x, x') \in H_{\succsim_j} \times L_{\succsim_j}.$$

We will show that

$$br(x, (\succsim_i, \succsim'_{-i})) = br(x', (\succsim_i, \succsim'_{-i})) \quad \& \quad br(x', \succsim') = br(x, \succsim'), \tag{3}$$

which contradicts $x \neq x'$ since we have the fixed tie-breaker.

From Definition 5,

$$br(x, (\succsim_i, \succsim'_{-i})) \geq br(x', (\succsim_i, \succsim'_{-i})) \quad \& \quad br(x', \succsim) \geq br(x, \succsim'). \tag{4}$$

We consider the following two cases:

Case 1: Suppose that $c^i_i = 0$. From the assumption, $x \succ'_i x'$ and $x \succsim'_i y \succsim'_i x'$ for all $y \in X$. Since $\succsim_i \in \mathcal{D}_i$,

$$br_i(x, \succsim'_i) - br_i(x', \succsim'_i) \geq |X| \geq br_i(x, \succsim_i) - br_i(x', \succsim_i).$$

This result and (4) imply that

$$0 \geq br(x, \succsim') - br(x', \succsim') \geq br(x, (\succsim_i, \succsim'_{-i})) - br(x', (\succsim_i, \succsim'_{-i})) \geq 0,$$

therefore, we obtain (3).

Case 2: Suppose that $c^i_i > 0$. From the assumption, $\sum_{j\neq i, c^i_j>0} c^i_j u_j(x) \geq \sum_{j\neq i, c^i_j>0} c^i_j u_j(y)$, and $\sum_{j\neq i, c^i_j>0} c^i_j u_j(y) \geq \sum_{j\neq i, c^i_j>0} c^i_j u_j(x')$ for any $y \in X$. Thus,

$$\forall y \in X, \quad \boldsymbol{c}^i\boldsymbol{u}(x) - c^i_i u_i(x) \geq \boldsymbol{c}^i\boldsymbol{u}(y) - c^i_i u_i(y), \tag{5}$$

$$\forall y \in X, \quad \boldsymbol{c}^i\boldsymbol{u}(x') - c^i_i u_i(x') \leq \boldsymbol{c}^i\boldsymbol{u}(y) - c^i_i u_i(y). \tag{6}$$

Since (5) implies that $br_i(x, \succsim_i) \leq br_i(x, \succsim'_i)$, and (6) implies $br_i(x', \succsim'_i) \leq br_i(x', \succsim_i)$, (4)–(6) show that

$$br(x, (\succsim_i, \succsim'_{-i})) \leq br(x, \succsim') \leq br(x', \succsim') \leq br(x', (\succsim_i, \succsim'_{-i})) \leq br(x, (\succsim_i, \succsim'_{-i})).$$

Thus, we obtain (3). □

A.6 Proof of Proposition 5

Proof. We can show that f^{pl} satisfies (AN) and (NNRD$_{\mathcal{D}\to\mathcal{R}}$) by replacing f^{br} and br in the proof of Proposition 4 with f^{pl} and n^{top}, respectively.

(WPO): For any $\succsim \in \mathcal{R}$, if there exist $x, y \in X$ such that $x \succ_i y$ for all $i \in N$, $n^{top}(y, \succsim) = 0$ and there exists $z_i \in X \setminus \{y\}$ such that $n_i^{top}(z_i, \succsim_i) = 1$ for all $i \in N$. Thus, $y \notin \operatorname{argmax}_{x' \in X} n^{top}(x', \succsim)$. □