

# Let’s Agree to Agree: Targeting Consensus for Incomplete Preferences through Majority Dynamics

Sirin Botan<sup>1</sup>, Simon Rey<sup>2</sup>, Zoi Terzopoulou<sup>3\*</sup>

<sup>1</sup>UNSW Sydney

<sup>2</sup>Institute for Logic, Language and Computation (ILLC), University of Amsterdam

<sup>3</sup>LAMSADE, Université Paris Dauphine-PSL

s.botan@unsw.edu.au, s.j.rey@uva.nl., zoi.terzopoulou@dauphine.psl.eu

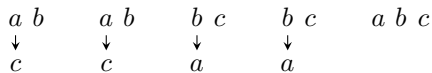
## Abstract

We study settings in which agents with incomplete preferences need to make a collective decision. We focus on a process of majority dynamics where issues are addressed one at a time and undecided agents follow the opinion of the majority. We assess the effects of this process on various consensus notions—such as the Condorcet winner—and show that in the worst case, myopic adherence to the majority damages existing consensus; yet, simulation experiments indicate that the damage is often mild. We also examine scenarios where the chair of the decision process can control the existence (or the identity) of consensus, by determining the order in which the issues are discussed.

## 1 Introduction

Groups of agents often need to make decisions by finding a consensus between different individual opinions: Among friends, hiring committees and teams of reviewers, as well as multiagent systems, reaching collective consensus is not always easy. Such situations become even less straightforward when agents hold incomplete opinions, which is often the case in real life, due to uncertainty, lack of knowledge, or reduced interest about the issues in question. Consider an example:

**Example 1.** Five organisers of an online conference must collectively decide about which video-conferencing app to use: *AppEar* (*a*), *Bridge* (*b*), or *C-nnect* (*c*)? Having never used *Bridge*, two organisers hold no opinion on it, but think that *AppEar* is better than *C-nnect* (denoted by  $a \rightarrow c$ ). Two organisers familiar with *AppEar* and *Bridge* find *Bridge* superior, while one organiser has never heard of any app:



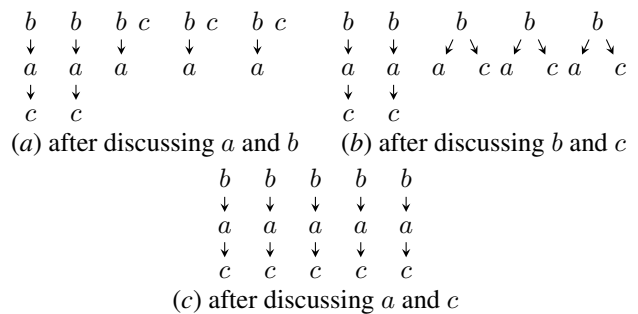
A consensus alternative for these opinions is not obvious.  $\triangle$

In two rapidly growing streams of literature within computational social choice, *opinion diffusion* and *liquid democracy*, agents adopt the opinions of their peers during collective decision processes—either directly by embracing them or indirectly by delegating their vote [Bredereck and Elkind, 2017;

Brill and Talmon, 2018; Botan *et al.*, 2019; Bloembergen *et al.*, 2019]. In our model, agents also consult their peers, but only about issues on which they initially hold no opinion. Supposing that the group discusses one issue at a time (here the comparison between two alternatives), as commonly happens in practice, all agents with a missing opinion on it will adopt the opinion of the majority. For agents who trust all their peers equally, following majority is indeed the best option, both if they aim at maximising agreement within the group, or at having the highest chances of making the correct decision [de Condorcet, 1785; May, 1952].<sup>1</sup>

Limited by constraints on time and energy, agents do their best to reach consensus *locally* for each single issue that is being discussed, but the group still targets consensus *globally*, after all issues have been addressed. *Majority dynamics* (MD) can help with that, as illustrated below.

**Example 1 (continued).** Suppose that the discussion starts with *AppEar* in comparison to *Bridge*. Since only two organisers have a relevant opinion (that *b* is better than *a*), everyone will adopt it. Those who find *a* better than *c* will also rank *b* above *c*, as a matter of consistency.<sup>2</sup> Next, suppose that *Bridge* and *C-nnect* are discussed. Note that only two organisers rank them, and both rank *b* above *c*, so this opinion will be adopted by everyone. The last issue addressed is the comparison between *AppEar* and *C-nnect*. Again everyone will adopt the only existing opinion: *a* above *c*.



An obvious consensus now exists: alternative *b*.<sup>3</sup>  $\triangle$

<sup>1</sup>Rawls [1971] also advocated the idea that agents hold twofold preferences—in the first level, their intrinsic preferences on the issues; in the second level, they wish to agree with their group.

<sup>2</sup>Formally, this consistency requirement is *transitivity*.

<sup>3</sup>In our specific example, the group reached a strong consensus,

\*Contact Author

Does MD assist with consensus beyond our example? Does the order of discussion matter? In this paper, we introduce a novel framework of majority dynamics for incomplete preferences, and use analytical and experimental tools to investigate its consequences with respect to a number of established consensus notions. Our main contributions show that (i) majority dynamics can damage even majority-based consensus alternatives such as the Condorcet winner (though not in some restricted preference domains); (ii) a chair can control the resulting consensus by determining the order of discussion, leading to controversial collective decisions; and (iii) the worst-case effects described in (i) and (ii) are in fact rare.

Earlier work has employed notions of consensus as a way to *rationalise* voting rules that minimise the distance from an ideal decision, given complete preferences [Elkind *et al.*, 2015]. Moreover, recent research has examined incomplete preferences from various static perspectives, e.g., axiomatic ones [Pini *et al.*, 2009; Terzopoulou and Endriss, 2019]. This paper’s angle is different. Instead of applying a voting rule, we search for direct agreement within the group in the form of a consensual alternative, which leads to outcomes that are less debatable and better explainable to the participating agents (note that choosing the best voting rule is a long-standing discussion in the social choice community). The process investigated by Eklund *et al.* [2007] bears a similar motivation, but assumes that agents evaluate alternatives with fixed criteria in mind, and that a chair has the (often unjustified) power to force the change of someone’s preference. Auletta *et al.* [2018] also study (majority-based) consensus in settings of opinion diffusion, but focus on social networks, and their results are limited to a single binary issue.

Moreover, many *consensus measures* have been defined to study the degree of discordance between the agents’ preferences in a group, without the goal of selecting a final alternative [Herrera-Viedma *et al.*, 2011]. These measures are usually based on distances and are associated with normative properties such as unanimity, suggesting that a group of agents with the same preferences has the highest possible consensus [Bosch, 2005; Alcalde-Unzu and Vorsatz, 2011].

Finally, manipulation by the chair has been explored under the name of *agenda manipulation* in judgment aggregation [Dietrich, 2016], and notions of *bribery* and *control* in voting [Faliszewski *et al.*, 2009; Faliszewski *et al.*, 2015]. Although related in flavour, these works are technically far from ours.

This paper proceeds as follows. Section 2 defines our model and the basic consensus notions. Section 3 studies the worst-case effects of the process of majority dynamics on consensus, including special cases of restricted domains. Section 4 solves the same exercise for the problem of consensus control by the chair, and Section 5 includes our experimental design and results. Finally, Section 6 concludes.

## 2 The Model

This section presents our model, terminology, and notation.

fully agreeing on the complete preference order. We will later discuss much weaker notions of consensus too.

### 2.1 Preliminaries

Let  $N = \{1, \dots, n\}$  be a finite set of agents, and  $A = \{a, b, c, \dots\}$  a finite set of  $m \geq 3$  alternatives.<sup>4</sup> Each agent  $i \in N$  reports a strict partial order  $\succ_i$  (an acyclic and transitive binary relation), which we call her *preference* and draw as a directed acyclic graph. We can also represent each preference as a set of pairwise rankings over alternatives. By  $ab$  we refer to the ordered pair of the alternatives  $a$  and  $b$ . For instance, if agent  $i$  prefers  $a$  to  $b$  and  $b$  to  $c$  (and thus also prefers  $a$  to  $c$  because of transitivity), her preference is the set  $\{ab, bc, ac\}$ .

A special kind of incomplete preferences are *strict weak orderings*: they represent alternatives ranked in different levels, with alternatives of the same level being incomparable. Concretely,  $\succ_i$  is a strict weak ordering if there exists some ordered partition  $S_i = (S_i^1, \dots, S_i^k)$  of  $A$  such that:

- (i) for any  $a, b \in S_i^g$  we have  $ab \notin \succ_i$  and  $ba \notin \succ_i$ , and
- (ii) for any  $S_i^g$  and  $S_i^h$  such that  $g < h$ , it must be the case that  $a \succ_i b$  for all  $a \in S_i^g$  and  $b \in S_i^h$ .

*Top-truncated* and *bottom-truncated* preferences are natural types of strict weak orderings, strictly ranking a subset of the alternatives and placing the remaining—incomparable—ones below or above them, respectively [Baumeister *et al.*, 2012].<sup>5</sup>

A vector *profile*  $\mathbf{P} = (\succ_1, \dots, \succ_n)$  of preference of size  $n$  is called a (possibly incomplete) profile. We will denote by  $N_{ab}^{\mathbf{P}} = |\{i \in N \mid a \succ_i b\}|$  the *support* of  $ab$ , i.e., the number of agents who prefer alternative  $a$  over  $b$  in the profile  $\mathbf{P}$ .

There are special structural characteristics exhibited by a preference profile that make the achievement of a collective decision an unequivocal task (recall Example 1). Each of the following notions of consensus generalises an existing one from the literature on complete preferences [Elkind *et al.*, 2015]. We start with a traditional concept:

- An alternative  $a$  is a **Condorcet winner** (CW) in  $\mathbf{P}$  if  $N_{ab}^{\mathbf{P}} > N_{ba}^{\mathbf{P}}$  for all  $b \in A \setminus \{a\}$ .

Next, we distinguish two notions that capture an alternative being on “top” of an incomplete preference, which are unambiguous in the complete case: An alternative  $a$  is *undominated* (abbreviated ‘UD’) in  $\succ$  if there is no  $b \in A$  such that  $b \succ a$ , and,  $a$  is *dominant* (abbreviated ‘Dom’) if  $a \succ b$  for all  $b \in A \setminus \{a\}$ . We define consensus for the former notion, and the latter is analogous.

- An alternative  $a$  is **unanimity undominated** (*UnanUD*) in  $\mathbf{P}$  if it is undominated in  $\succ_i$  for all  $i \in N$ .
- An alternative  $a$  is **majority undominated** (*MajUD*) in  $\mathbf{P}$  if  $|\{i \in N \mid a \text{ is undominated in } \succ_i\}| > |N|/2$ .
- An alternative  $a$  is **plurality undominated** (*PlurUD*) in  $\mathbf{P}$  if  $|\{i \in N \mid a \text{ is undominated in } \succ_i\}| > 0$  and  $a \in \operatorname{argmax}_{b \in A} |\{i \in N \mid b \text{ is undominated in } \succ_i\}|$ .

<sup>4</sup>Our counterexamples use the minimum number of needed alternatives, but extend to any larger number by including dummy ones.

<sup>5</sup>Our results for strict weak orderings hold for bottom-truncated preferences. Top-truncated preferences differ very little from complete ones in the context of consensus, and are thus less interesting.

| Quantifier Over Profiles |           | Preserving Consensus |                 | Losing Consensus |               | Generating Consensus     |                      | Preserving Absence of Consensus |
|--------------------------|-----------|----------------------|-----------------|------------------|---------------|--------------------------|----------------------|---------------------------------|
| Orders                   | Existence | Identity             | Existence       | Identity         | Existence     | Identity                 | Existence            | Identity                        |
| $\forall$                | $\forall$ | Def. 1               | Def. 2          | $\times$ co      | $\times$ co   | $\times$ co              | $\times$ $\emptyset$ | $\times$ $\emptyset$            |
| $\forall$                | $\exists$ | Def. 3               | —               | $\times$ co      | —             | $\times$ co              | —                    | $\times$ $\emptyset$            |
| $\exists$                | $\forall$ | $\checkmark$ co      | $\checkmark$ co | <u>Def. 3</u>    | <u>Def. 4</u> | $\checkmark$ $\emptyset$ | <u>Def. 5</u>        | $\checkmark$ co                 |
| $\exists$                | $\exists$ | $\checkmark$ co      | —               | <u>Def. 1</u>    | —             | $\checkmark$ $\emptyset$ | —                    | $\checkmark$ co                 |

Table 1: Effects on consensus. Each cell points to the answer source for the relevant question. For example, according to the cell at the first row and third (fourth) column, Def. 1 (Def. 2) captures the idea of preserving the existence (identity) of consensus for all preference profiles and all update orders. ‘—’ means that the question is meaningless: we can only inquire about different identities of consensus under more than one order; ‘ $\checkmark$ ’ and ‘ $\times$ ’ indicate a trivial positive or negative answer, respectively, by considering complete (co) or empty ( $\emptyset$ ) preference profiles. Overlined words denote the statement of a definition in the contrapositive.

More than one alternative in a profile  $\mathbf{P}$  may satisfy the undominated consensus definitions (and the plurality dominant one). Importantly, we only say that a profile exhibits consensus when there is a *unique* such alternative.

Given a consensus notion  $C \in \{\text{‘CW’}, \text{‘UnanUD’}, \text{‘UnanDom’}, \text{‘MajUD’}, \text{‘MajDom’}, \text{‘PlurUD’}, \text{‘PlurDom’}\}$  and a preference profile  $\mathbf{P}$ , we define  $C(\mathbf{P}) \in A \cup \{\perp\}$  to be the consensus alternative in  $\mathbf{P}$ , with respect to  $C$ ; if such a consensus alternative does not exist, we write  $C(\mathbf{P}) = \perp$ .

## 2.2 Majority Dynamics

Let  $\sigma = (p_1, \dots, p_\ell)$  be an ordering of pairs of alternatives, s.t. for any two alternatives  $a$  and  $b$ , exactly one of  $ab$  or  $ba$  are in  $\sigma$ . We define the process of *majority dynamics*  $\text{MD}_\sigma$  such that  $\text{MD}_\sigma(\mathbf{P})$  is the stable profile that results from the application of the dynamics on the initial profile  $\mathbf{P}$  where pairs are discussed following the update order  $\sigma$ . Let  $\mathbf{P}^0 := \mathbf{P}$ . When the pair  $p_t = ab$  is discussed, preferences are updated from  $\mathbf{P}^{t-1} = (\succ_i^{t-1})_{i \in N}$  to  $\mathbf{P}^t = (\succ_i^t)_{i \in N}$  such that:

$$\succ_i^t = \begin{cases} \succ_i^{t-1} & \text{if } ab \text{ or } ba \in \succ_i^{t-1} \\ \llbracket \succ_i^{t-1} \cup \{ab\} \rrbracket & \text{otherwise, and if } N_{ab}^{\mathbf{P}} \geq N_{ba}^{\mathbf{P}} \\ \llbracket \succ_i^{t-1} \cup \{ba\} \rrbracket & \text{otherwise.} \end{cases}$$

Here,  $\llbracket \succ \rrbracket$  is the transitive closure of  $\succ$ , ensuring we never violate transitivity. For example, an agent with preference set  $\{bc\}$  that adds  $ab$ , will also have to add  $ac$ . This means that individual agents’ hold (possibly incomplete) transitive preferences with no cycles at every step of the dynamics. The profile  $\text{MD}_\sigma(\mathbf{P})$  is thus a profile of complete preferences over the set of alternatives. Note that we always break ties in favour of the first alternative in the considered pair.

## 2.3 Effects on Consensus

All possible effects that MD may have on group consensus are categorised into four types: (i) preserving, (ii) losing, or (iii) generating consensus, and (iv) preserving the absence of consensus. Moreover, given a profile of incomplete preferences, different update orders may lead to different consensus alternatives. With this in mind, effects (i) to (iii) further distinguish between the existence of *any* consensus alternative, and the identity of a *specific* consensus alternative.

Table 1 shows that five definitions below capture all non-trivial cases regarding effects on consensus, meaning that together they are expressively complete. We begin with definitions that refer to effects regarding *all* possible update orders.

**Definition 1.** Given a consensus notion  $C$ , MD **preserves  $C$  existence** if for all profiles  $\mathbf{P}$ :

$$C(\mathbf{P}) \neq \perp \text{ implies that } C(\text{MD}_\sigma(\mathbf{P})) \neq \perp, \text{ for all orders } \sigma.$$

**Definition 2.** Given a consensus notion  $C$ , MD **preserves  $C$  identity** if for all profiles  $\mathbf{P}$  and alternatives  $a \in A$ :

$$C(\mathbf{P}) = a \text{ implies that } C(\text{MD}_\sigma(\mathbf{P})) = a, \text{ for all orders } \sigma.$$

We next weaken consensus preservation, by quantifying over *some* order instead of *all* orders. Positive control captures scenarios where MD enables the mechanism designer to select a suitable update order to preserve consensus.

**Definition 3.** Given a consensus notion  $C$ , MD enables **positive  $C$  existence control** if for all profiles  $\mathbf{P}$ :

$$C(\mathbf{P}) \neq \perp \text{ implies } C(\text{MD}_\sigma(\mathbf{P})) \neq \perp, \text{ for some order } \sigma.$$

**Definition 4.** Given a consensus notion  $C$ , MD enables **positive  $C$  identity control** if for all profiles  $\mathbf{P}$  and alt.  $a \in A$ :

$$C(\mathbf{P}) = a \text{ implies } C(\text{MD}_\sigma(\mathbf{P})) = a, \text{ for some order } \sigma.$$

While positive control enables the achievement of consensus, negative control prevents a specific consensus from forming. This may be done either by imposing no consensus at the end of the dynamic process, or by inflicting different consensus alternatives, depending on the update order.

**Definition 5.** Given a consensus notion  $C$ , MD enables **negative  $C$  control** if for all profiles  $\mathbf{P}$  with  $C(\mathbf{P}) = \perp$ , one of the following conditions hold:

- $C(\text{MD}_\sigma(\mathbf{P})) = \perp$ , for some order  $\sigma$ ;
- $C(\text{MD}_\sigma(\mathbf{P})) \neq C(\text{MD}_{\sigma'}(\mathbf{P}))$ , for some orders  $\sigma, \sigma'$ .

Table 2 summarises the results we will prove in Sections 3 and 4—notably, we will see that issues regarding existence and identity coincide for our consensus notions. Full proofs for all our results can be found in the supplementary material.

## 3 Preserving Consensus

We examine whether MD preserves consensus, starting with the notion of a Condorcet winner. For the special case of three alternatives, we can report some good news.

**Proposition 1.** For  $m = 3$ , MD *preserves CW existence, but not CW identity*.

|         | Pres. consensus | Pos. control | Neg. control |
|---------|-----------------|--------------|--------------|
| CW      | ✗ (✓)           | ✓            | ✓            |
| PlurUD  | ✗               | ✗            | ✗            |
| PlurDom | ✗               | ✗            | ✗            |
| MajUD   | ✗               | ✓            | ✗            |
| MajDom  | ✓               | ✓            | ✗            |
| UnanUD  | ✗ (✓)           | ✓            | ✓            |
| UnanDom | ✓               | ✓            | ✗            |

Table 2: Summary of effects on consensus. ‘✓’ means that the definition of the relevant effect holds for the given consensus notion, and ‘✗’ means that it is violated. The results for strict weak orderings, when they differ from the general case, are shown in a parenthesis.

*Proof.* Let  $A = \{a, b, c\}$ , and let  $\sigma = (ab, ac, bc)$  be the update order (for different orders the proof is analogous). Consider any profile  $\mathbf{P}^0$ . If alternative  $a$  is a Condorcet winner in  $\mathbf{P}^0$  we know that the support of  $ba$  or  $ca$  will not increase throughout  $\text{MD}_\sigma$ , and  $a$  will remain Condorcet winner in  $\text{MD}_\sigma(\mathbf{P}^0)$ . Thus, we only need to consider alternative  $b$  or  $c$  being the Condorcet winner. As the proofs proceed similarly, we only consider one of the two cases.

Suppose alternative  $b$  is the Condorcet winner in  $\mathbf{P}^0$ . When the pair  $ab$  is discussed at time  $t = 1$ , the agents who update their preference will support  $b$  over  $a$ , and transitivity requirements can lead them to support  $b$  over  $c$ , but never  $c$  over  $b$ .

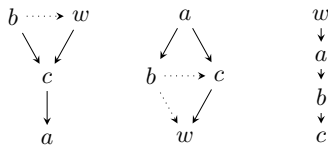
Now, if  $N_{ac}^{\mathbf{P}^1} \geq N_{ca}^{\mathbf{P}^1}$ , then the support of  $cb$  will not increase at time  $t = 2$ . We thus have  $N_{cb}^{\mathbf{P}^2} = N_{cb}^{\mathbf{P}^0} < N_{bc}^{\mathbf{P}^0} \leq N_{bc}^{\mathbf{P}^2}$ . Hence, after time  $t = 3$ ,  $bc$  will have more support than  $cb$  and  $b$  will still be the Condorcet winner in  $\text{MD}_\sigma(\mathbf{P}^0)$ .

Suppose instead that  $N_{ac}^{\mathbf{P}^1} < N_{ca}^{\mathbf{P}^1}$ . Then, at time  $t = 2$ , the support of  $ca$  will increase, and potentially also that of  $cb$ . At time  $t = 3$ , if  $N_{bc}^{\mathbf{P}^2} \geq N_{cb}^{\mathbf{P}^2}$ ,  $b$  will remain the Condorcet winner in  $\text{MD}_\sigma(\mathbf{P}^0)$ . If instead  $N_{bc}^{\mathbf{P}^2} < N_{cb}^{\mathbf{P}^2}$ , then  $c$  will be the Condorcet winner in  $\text{MD}_\sigma(\mathbf{P}^0)$ , as a majority of agents supported  $c$  over  $a$  at time  $t = 2$ . Note that this results in a change of the identity of the Condorcet winner.  $\square$

However, the situation is not as bright when  $m > 3$ .

**Proposition 2.** For  $m > 3$ , MD does not preserve CW existence (thus neither CW identity).

*Proof.* Consider the 3-agent profile  $\mathbf{P}$  described below (ignoring the dotted edges) where  $w$  is the Condorcet winner.

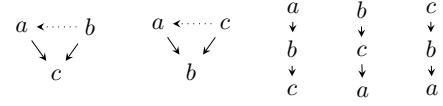


Let  $\sigma$  be any update order starting with  $bc$  and  $bw$ .  $\text{MD}_\sigma$  on this profile results in the inclusion of the dotted edges. Note that no Condorcet winner exists in  $\text{MD}_\sigma(\mathbf{P})$ .  $\square$

When it comes to undominated alternatives and *PlurDom*, we find that MD violates consensus preservation even for  $m = 3$ . The opposite holds for *UnanDom* and *MajDom*.

**Proposition 3.** Take  $C \in \{\text{‘PlurUD’}, \text{‘MajUD’}, \text{‘UnanUD’}, \text{‘PlurDom’}\}$ . Then MD does not preserve  $C$  existence (thus neither identity).

*Partial proof.* We present the proof for *PlurUD* and *MajUD*. Consider the profile below (ignoring the dotted edges) where  $a$  is the plurality and majority undominated consensus.



Let  $\sigma$  be an order that starts with the sequence  $(ba, ca)$ . As the majority of agents prefer both  $b$  and  $c$  to  $a$ , we will end up including the dotted edges. Overall, both  $b$  and  $c$  will be dominant in exactly two individual rankings.  $\square$

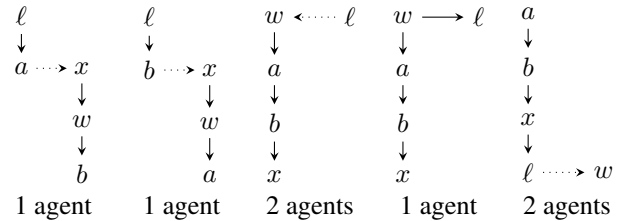
**Proposition 4.** Take  $C \in \{\text{‘UnanDom’}, \text{‘MajDom’}\}$ . Then MD preserves  $C$  identity (thus also existence).

*Proof sketch.* Suppose  $a \in A$  is the consensus alternative according to *UnanDom* or *MajDom*. Then, any update on a pair  $ab$  or  $ba$ , for any  $b \in A \setminus \{a\}$ , will be made in favour of  $a$ . For the same reason, no  $b \in A \setminus \{a\}$  can be preferred to  $a$  because of transitivity. The consensus will thus remain.  $\square$

**Quality of Consensus.** In cases where consensus is preserved, but the identity of the consensus alternative is not, we explore how “bad” the new consensus can be. Unfortunately, the answer is not very positive. Before diving in, we need the following definition: An alternative  $a$  is a *Condorcet loser* in  $\mathbf{P}$  if  $N_{ab}^{\mathbf{P}} < N_{ba}^{\mathbf{P}}$  for all  $b \in A \setminus \{a\}$ .

**Proposition 5.** For  $m \geq 5$ , MD can turn a Condorcet loser into a Condorcet winner.

*Proof.* Let  $A = \{w, \ell, a, b, x\}$ , and consider the 7-agent profile  $\mathbf{P}^0$  below (ignoring the dotted edges), where  $w$  is the Condorcet winner and  $\ell$  is the Condorcet loser.



Consider what happens if the update order  $\sigma$  starts with  $ax, bx, \ell w$ . After time  $t = 2$ , agents have updated their preferences on  $ax$  and  $bx$  and  $\ell$  is preferred to  $w$  by a majority. Updating on  $\ell w$  at time  $t = 3$  makes  $\ell$  the Condorcet winner, and that cannot change. The updated preferences after time  $t = 3$  are represented by the dotted edges.  $\square$

**Proposition 6.** For  $m \geq 8$ , MD can turn a plur. dominated (or undominated) alternative into a plur. dominant one.

A unanimity (majority) undominated alternative can become a dominated alternative in all agents’ preferences as this is a relatively weak consensus requirement. We will thus focus on *UnanDom* and *MajDom*. We say an alternative is *unanimity (majority) dominated* if it is dominated by all alternatives in the preferences of all (a majority of) of agents.

**Proposition 7.** *A unan. (majority) dominated alternative cannot become the unan. (majority) dominant alternative.*

**Restricted Preference Domains.** When reducing our scope to strict weak orderings, we find good news.

**Proposition 8.** *For profiles of strict weak orderings, MD is CW-identity preserving.*

*Proof.* Let  $w$  be the Condorcet winner in  $\mathbf{P}^0$ . Consider a given agent  $i$ , and some alternative  $a \neq w$ . We claim that it is not possible to have  $a \succ_i^t w$  at any time step  $t$  of the majority dynamics, unless  $a \succ_i^0 w$ . The proof is by induction.

Suppose our claim holds for all rounds  $t' < t$  and all alternatives  $z \neq w$ . Then  $w$  is a Condorcet winner in round  $t - 1$ . If agent  $i$  has already have expressed either that  $a \succ_i w$  or that  $w \succ_i a$  in a round before  $t$ , no update can be done on this pair, so we need only consider the case where  $a, w \in S_i^g$  for some  $g$ , where  $S_i^g$  is the  $g$ -th tier of  $i$ 's preferences. There are two cases we need to examine for the update in round  $t$ .

Suppose the update is performed on  $aw$  or  $wa$  directly. Then,  $i$  will support  $wa$  since  $w$  is a Condorcet winner.

Suppose instead we are updating on a pair that is neither  $aw$  nor  $wa$ , leading  $i$  to prefer  $a$  over  $w$  because of transitivity.

Then, there must exist some alternative  $b$  such that  $a \succ_i^{t-1} b$  and  $b \succ_i^{t-1} w$ . This, however, contradicts our assumption that  $\succ_i$  is a strict weak ordering, as this—in combination with our assumption that  $aw \notin \succ_i^{t-1}$  and our induction hypothesis—precludes the existence of an alternative that is dominated by one of  $a$  and  $w$ , while dominating the other. We conclude that if a Condorcet winner exists in  $\mathbf{P}^0$ , then that alternative must remain a Condorcet winner in  $\text{MD}_\sigma(\mathbf{P}^0)$ .  $\square$

The proof of Proposition 3 uses strict weak orderings, except for *UnanUD*, and that of Proposition 4 also holds with strict weak orderings. Only the case of *UnanUD* is left then.

**Proposition 9.** *For profiles of strict weak orderings, MD is UnanUD-identity preserving.*

*Proof.* Consider a profile  $\mathbf{P}^0$  that represents strict weak orderings. Let  $A' = \bigcup_{i \in N} S_i^1$  be the set of all alternatives that are in the top tier of at least one agent.

Suppose there exists a unique unanimity undominated alternative  $w$  in  $\mathbf{P}$ . Then for any  $a \in A' \setminus \{w\}$ , there exists at least one  $i \in N$  such that  $w \succ_i a$ , and that for no agent  $i' \in N$  we have  $a \succ_{i'} w$ . Hence, for any update order  $\sigma$ , when updating on the pair  $wa$  or  $aw$ , agents will update in favour of  $w$  and that for all  $a \in A' \setminus \{w\}$ . Since preferences are strict weak orderings, all alternatives in  $A \setminus A'$  are initially unanimously dominated by  $w$ . Overall, for any  $\sigma$ ,  $w$  will be the unanimity undominated consensus in  $\text{MD}_\sigma(\mathbf{P}^0)$ .  $\square$

## 4 Controlling Consensus

We now investigate control issues. Note that all our proofs in this section use profiles consisting only of strict weak orderings. We start by exploring positive control.

**Proposition 10.** *Take  $C \in \{\text{'CW'}, \text{'UnanUD'}, \text{'UnanDom'}, \text{'MajDom'}\}$ . Then MD enables positive  $C$  identity (and thus also existence) control.*

**Proposition 11.** *Take  $C \in \{\text{'PlurUD'}, \text{'PlurDom'}, \text{'MajUD'}\}$ . Then MD does not enable positive  $C$  existence (and thus neither identity) control.*

Negative control is more difficult to achieve.

**Proposition 12.** *Take  $C \in \{\text{'CW'}, \text{'UnanUD'}\}$ . Then MD enables negative  $C$  control.*

*Proof.* We first consider the case of *CW*. Let  $\mathbf{P}$  be a profile without a Condorcet winner. Then, for any alternative  $a \in A$ , there is another alternative  $b \in A \setminus \{a\}$  such that  $N_{ba}^{\mathbf{P}} \geq N_{ab}^{\mathbf{P}}$ . Now, for the update order  $\sigma$  starting with the pair  $ba$ , we must have  $\text{CW}(\text{MD}_\sigma(\mathbf{P})) \neq a$ . If there is no Condorcet winner in  $\text{MD}_\sigma(\mathbf{P})$ , we are done. In the other case, let  $w = \text{CW}(\text{MD}_\sigma(\mathbf{P}))$ . Because the previous also applies to  $w$ , there is  $\sigma'$  such that  $w$  is not the Condorcet winner in  $\text{MD}_{\sigma'}(\mathbf{P})$ . Now if  $\text{CW}(\text{MD}_{\sigma'}(\mathbf{P})) = \perp$ , we are done. Otherwise, we have two update orders  $\sigma$  and  $\sigma'$  such that  $\text{CW}(\text{MD}_\sigma(\mathbf{P})) \neq \text{CW}(\text{MD}_{\sigma'}(\mathbf{P}))$ , which proves the statement.

For the case of *UnanUD*, it is clear that for any profile  $\mathbf{P}$  with no unanimity undominated alternative, no such alternative can arise at the end of MD. If, on the other hand, there are two unanimity undominated alternatives  $a$  and  $b$  in  $\mathbf{P}$ , then for  $\sigma$  starting with  $ab$ , we will have  $\text{UnanUD}(\text{MD}_\sigma(\mathbf{P})) = a$ . However, for  $\sigma'$  starting with  $ba$ , we would obtain  $\text{UnanUD}(\text{MD}_{\sigma'}(\mathbf{P})) = b \neq \text{UnanUD}(\text{MD}_\sigma(\mathbf{P}))$ .  $\square$

**Proposition 13.** *Take  $C \in \{\text{'UnanDom'}, \text{'PlurUD'}, \text{'PlurDom'}, \text{'MajUD'}, \text{'MajDom'}\}$ . Then MD does not enable negative  $C$  control.*

## 5 Experimental Analysis

The previous two sections showed how diverse the outcomes of MD can be. We complement this formal analysis by an experimental one in order to quantify the different effects.<sup>6</sup>

For each experiment, only some representative results are displayed. There are overall very few distinctions between dominant and undominated-based consensus notions, so we only present the undominated case. Moreover, unanimity-based consensus imposes such a strong requirement that the relevant plots are uninformative, and have thus been omitted.

### 5.1 Quantifying the Effects on Consensus

We first explore the effects that MD has on consensus, including generation, preservation, and loss of consensus, as well as preservation of the absence of consensus. We applied MD on synthetic profiles and used a fixed update order.

We varied the number of agents from 1 to 25, only considering odd numbers to avoid effects that rely solely on ties. For each case, we generated 5 000 000 random profiles over five alternatives. To generate a partial preference, we considered each pair of alternatives  $\{a, b\}$  and decided with uniform probability whether  $a$  beats  $b$ ,  $b$  beats  $a$ , or,  $a$  and  $b$  are not compared. We repeated this process until the generated set of pairwise comparisons was transitive.

<sup>6</sup>The experiments have been coded in Python, run on a Debian machine with 16 cores and 16GB RAM. The code and the data is available at <https://github.com/Simon-Rey/Let-s-Agree-to-Agree-Majority-Dynamics-for-Incomplete-Preferences>.

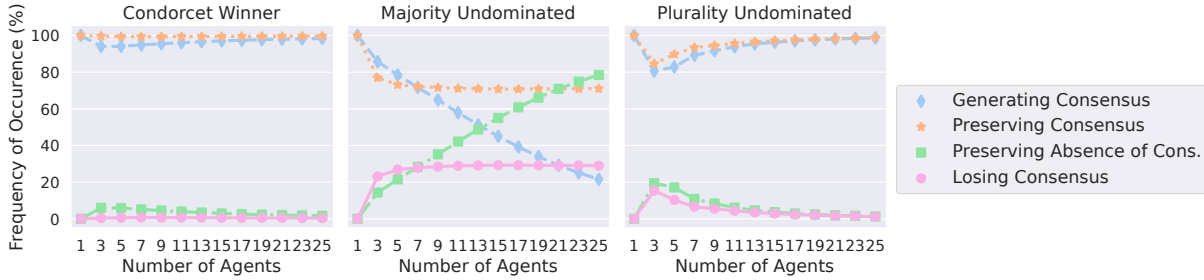


Figure 1: Frequency of each effect on consensus with respect to the number of agents. For preservation and loss of consensus, we normalized over the number of profiles with initial consensus, while for the other two effects over the number of profiles without initial consensus.

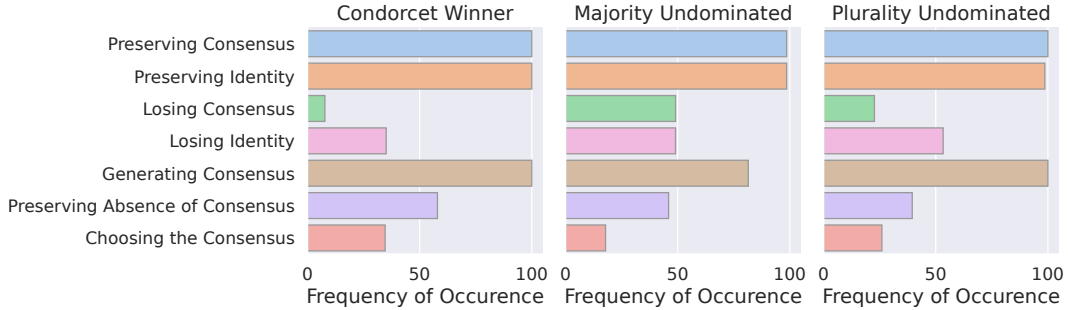


Figure 2: Frequency of each possible type of control. For the first four types, we normalized over the number of profiles with initial consensus; for the next two over the number of profiles without initial consensus; and for the last one over the total number of profiles. “Choosing the consensus” refers to the idea of making a specific alternative, say  $a$ , the consensus.

Figure 1 presents the effects on consensus, depending on the number of agents. Our first observation is that MD performs particularly well according to both Condorcet and plurality undominated consensus: consensus is almost always preserved if it was there to begin with, and generated otherwise. This trend is robust wrt. the number of agents. The picture is slightly different for majority undominated consensus. Even though MD often preserves consensus, the frequency of generating consensus drops significantly with the number of agents. This is likely due to the majority threshold increasing with the number of agents, making the consensus requirement more demanding. On the other hand, Condorcet and plurality-based requirements only depend on the profile and not on external parameters, hence their robustness.

We have also studied these effects depending on the level of completeness (how close to complete a profile is), and found that the level on completeness has little impact, except when it comes to majority-based consensus (for the same reasons as above). We present the figures in the supplementary material.

## 5.2 Quantifying the Opportunities for Control

Through experiments, we also explore the question of controlling the consensus. Our goal is to quantify the power that the chair is granted by choosing the update order. We generated 20 000 profiles of 11 agents over 4 alternatives and ran MD on each profile and for all of the 46 080 update orders. We counted for how many profiles at least one update order would allow the chair to achieve each type of control. The results are presented in Figure 2. We observed that MD per-

forms relatively well at preventing the chair from damaging consensus: Only on few occasions can the chair make consensus disappear or prevent it from being generated. A benevolent chair will however be much more successful; it is very often possible to preserve the consensus and to generate it.

## 6 Conclusion

We have studied an original process of majority dynamics for agents with incomplete preferences. We have asked whether consulting the majority to fill missing opinions assists group consensus, and have answered that in the worst case it does not—only alternatives that are dominant for at least half of the agents are safe. Countering this, we have provided some good news: Majority dynamics preserve a Condorcet winner within natural profiles of strict weak orderings, and also do so frequently within arbitrary ones. In addition, the chair always has the power to choose an order of discussion such that consensus is preserved (unless we care about plurality-based consensus), and she can very often generate a new one too (while she can rarely make an existing one disappear). Finally, our experiments indicate that the chair can rarely choose an order to make a specific alternative the consensus.

Yet, many questions remain open. For instance: What is the computational complexity of selecting a suitable order for control, or of minimising the number of updates until consensus is reached? In cases where consensus is not achieved, how far is the resulting profile from being consensual? These and other questions are left for future work.

## Acknowledgements

Sirin Botan is supported by the ARC Laureate Project FL200100204 (“Trustworthy AI”). Simon Rey is supported by the NWO Vici grant 639.023.811 (“Collective Information”). Zoi Terzopoulou is a fellow of the Paris Region Fellowship Programme, supported by the Paris Region. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945298-ParisRegionFP. We thank the reviewers of IJCAI-2022 for their useful feedback.

## References

- [Alcalde-Unzu and Vorsatz, 2011] Jorge Alcalde-Unzu and Marc Vorsatz. Measuring consensus: Concepts, comparisons, and properties. In *Consensual processes*, pages 195–211. Springer, 2011.
- [Auletta et al., 2018] Vincenzo Auletta, Diodato Ferraioli, and Gianluigi Greco. Reasoning about consensus when opinions diffuse through majority dynamics. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.
- [Baumeister et al., 2012] Dorothea Baumeister, Piotr Faliszewski, Jérôme Lang, and Jörg Rothe. Campaigns for lazy voters: Truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2012.
- [Bloembergen et al., 2019] Daan Bloembergen, Davide Grossi, and Martin Lackner. On rational delegations in liquid democracy. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, 2019.
- [Bosch, 2005] R Bosch. *Characterizations of Voting Rules and Consensus Measures*. PhD thesis, Tilburg University, 2005.
- [Botan et al., 2019] Sirin Botan, Umberto Grandi, and Laurent Perrussel. Multi-issue opinion diffusion under constraints. In *Proceedings of the 18th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2019.
- [Bredereck and Elkind, 2017] Robert Bredereck and Edith Elkind. Manipulating opinion diffusion in social networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.
- [Brill and Talmon, 2018] Markus Brill and Nimrod Talmon. Pairwise liquid democracy. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.
- [de Condorcet, 1785] Nicolas de Condorcet. *Essai sur l’Application de l’Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Imprimerie Royal, Paris, 1785.
- [Dietrich, 2016] Franz Dietrich. Judgment aggregation and agenda manipulation. *Games and Economic Behavior*, 95:113–136, 2016.
- [Eklund et al., 2007] Patrik Eklund, Agnieszka Rusinowska, and Harrie De Swart. Consensus reaching in committees. *European Journal of Operational Research*, 178(1):185–193, 2007.
- [Elkind et al., 2015] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Distance rationalization of voting rules. *Social Choice and Welfare*, 45(2):345–377, 2015.
- [Faliszewski et al., 2009] Piotr Faliszewski, Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. Llull and Copeland voting computationally resist bribery and constructive control. *Journal of Artificial Intelligence Research*, 35:275–341, 2009.
- [Faliszewski et al., 2015] Piotr Faliszewski, Yannick Reisch, Jörg Rothe, and Lena Schend. Complexity of manipulation, bribery, and campaign management in Bucklin and fallback voting. *Autonomous Agents and Multi-Agent Systems*, 29(6):1091–1124, 2015.
- [Herrera-Viedma et al., 2011] Enrique Herrera-Viedma, José Luis García-Lapresta, Janusz Kacprzyk, Mario Fedrizzi, Hannu Nurmi, and Sławomir Zadrozny, editors. *Consensual Processes*. Springer, 2011.
- [May, 1952] Kenneth O May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica: Journal of the Econometric Society*, 20(4):680–684, 1952.
- [Pini et al., 2009] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Aggregating partially ordered preferences. *Journal of Logic and Computation*, 19(3):475–502, 2009.
- [Rawls, 1971] John Rawls. *A Theory of Justice*. Belknap Press, 1971.
- [Terzopoulou and Endriss, 2019] Zoi Terzopoulou and Ulle Endriss. Aggregating incomplete pairwise preferences by weight. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.