United for Change: Deliberative Coalition Formation to Change the Status Quo

Edith Elkind, Davide Grossi, Ehud Shapiro, and Nimrod Talmon

Abstract

We study a setting in which a community wishes to identify a strongly supported proposal from a space of alternatives, in order to change the status quo. We describe a deliberation process in which agents dynamically form coalitions around proposals that they prefer over the status quo. We formulate conditions on the space of proposals and on the ways in which coalitions are formed that guarantee deliberation to succeed, that is, to terminate by identifying a proposal with the largest possible support. Our results provide theoretical foundations for the analysis of deliberative processes, in particular in systems for democratic deliberation support, such as, for instance, LiquidFeedback or Polis.[a]

[a]A preliminary version of this paper appears in AAAI '21 Elkind et al. [2021]; the main change here is the addition of Section 7 regarding hypercubes.

1 Introduction

Democratic decision-making requires equality in voting on alternatives, and social choice theory provides us with a wealth of tools for preference aggregation [see, e.g., Zwicker, 2016]. However, another important dimension, which has received considerably less attention in the literature, is equality in deciding on which alternatives to vote upon. In practice, agents engaging in group decisions do not just vote on an externally determined set of choices: rather, they make proposals, deliberate over them, and join coalitions to push their proposals through. Understanding these processes is critical for digital democracy applications [Brill, 2018] that can provide support for online voting and deliberation to self-governing communities. Notable examples of this kind are the LiquidFeedback[1] [Behrens et al., 2014] and the Polis[2] platforms.

In this paper, we aim to provide a plausible model of deliberative processes in self-governed (online) systems. We abstract away from the communication mechanisms by means of which deliberation may be concretely implemented, and focus instead on the coalitional effects of deliberation, that is, on how coalitions in support of various proposals may be formed or broken in the face of new suggestions. In our model, agents and alternatives are located in a metric space, and there is one distinguished alternative, which we refer to as the *status quo*. We assume that the number of alternatives is large (possibly infinite), so that the agents cannot be expected to rank the alternatives or even list all alternatives they prefer to the status quo. Rather, during the deliberative process, some agents formulate new proposals, and then each agent can decide whether she prefers a given proposal p to the status quo, i.e., whether p is closer to her location than the status quo alternative; if so, she may join a coalition of agents supporting p.

Importantly, such coalitions are dynamic: an agent may move to another coalition if she thinks that its proposal is more likely to attract a large number of supporters; or, two coalitions may merge, possibly leaving some members behind. We assume that agents are driven by a desire to get behind a winning proposal; thus, they may prefer a larger coalition with a less appealing proposal to a smaller coalition with a more appealing proposal, as long

[1]`https://liquidfeedback.org`
[2]`https://pol.is/home`

as the former proposal is more attractive to them than the status quo. The deliberation succeeds if it identifies a most popular alternative to the status quo.

This process is democratic in that each participant who is capable of formulating a new proposal is welcome to do so; furthermore, participants who do not have the time or sophistication to work out a proposal can still take part in the deliberation by choosing which coalition to join. Also, participants could be assisted by AI-enabled tools, to help them search for suitable proposals that may command larger support.

To flesh out the broad outline of the deliberative coalition formation process described above, we specify rules that govern the dynamics of coalition formation. We explore several such rules ranging from single-agent moves to more complex transitions where coalitions merge behind a new proposal, possibly leaving some dissenting members behind. Our aim is to understand whether the agents can succeed at identifying alternatives to the status quo that obtain largest support in the group, if they conduct deliberation in a certain way. We concentrate on how the properties of the underlying abstract space of proposals and the coalition formation operators available to the agents affect our ability to provide guarantees on the success of the deliberative coalition formation process: that is on whether the process terminates with a coalition having formed around a proposal with largest support. We show that, as the complexity of the proposal space increases, more sophisticated forms of coalition formation are required in order to assure success. Intuitively, this seems to suggest that complex deliberative spaces require more sophisticated coalition formation abilities on the side of the agents.

Paper contribution We study four ways in which agents can form coalitions:

1. by *deviation*, when a single agent from one coalition moves to another, weakly larger coalition;
2. by *following*, when a coalition joins another coalition in supporting the second coalition's proposal;
3. by *merging*, when two coalitions join forces behind a new proposal;
4. by *compromising*, when agents belonging to two or more coalitions form a new larger coalition, possibly leaving dissenting agents behind.

We refer to these types of coalition formation operations as *transitions*.

We show that, for each class of transitions, the deliberation process is guaranteed to terminate; in fact, for single-agent, follow, and merge transitions the number of steps until convergence is at most polynomial in the number of agents. Furthermore, follow transitions are sufficient for deliberation to succeed if the set of possible proposals is a subset of the 1-dimensional Euclidean space, but this is no longer the case in two or more dimensions; in contrast, compromise transitions that involve at most two coalitions are sufficient in sufficiently rich subsets of $\mathbb{R}^d$ for each $d \geq 1$. The 'richness', however, is essential: we provide an example where the space of proposals is a finite subset of $\mathbb{R}^3$, but 2-coalition compromises are not capable of identifying a most popular proposal.

While our primary focus is on deliberation in Euclidean spaces, we also provide results about two classes of non-Euclidean spaces. The first one is the class of weighted trees rooted on the status quo, with the metric given by the standard path length. We show that, in these spaces, merge transitions suffice for successful deliberation. The second class of spaces arises naturally in combinatorial domains as well as in multiple referenda settings. These spaces, which we refer to as d-hypercubes, are defined by the set of all binary opinions on d issues with the discrete metric known as Hamming or Manhattan distance. We show that deliberative success in d-hypercubes requires compromises involving at least d coalitions,

Transition	Termination	1-Eucl.	2-Eucl.	d-Eucl.	Trees	d-Hyperc.
Single-ag.	n^2 (Pr. 1)	✗ (Ex. 2)	✗	✗	✗	✗
Follow	n^2 (Pr. 2)	✓ (Th. 1)	✗ (Ex. 3)	✗	✗	✗
Merge	n^2 (Pr. 3)	✓ (Th. 1)	✗ (Ex. 5)	✗	✓ (Th. 2)	✗
Compr.	n^n (Pr. 4)	✓	✓	✓ (Th. 3)	✓	✗ (Ex. 7)
d-Compr.	n^n (Co. 1)	✓	✓	✓	✓	✓ ($d \leq 3$, Pr. 8)

Table 1: Summary of our main results. Each row corresponds to a transition type, for which we show bounds on the length of deliberations for it, and whether success—when using the corresponding transition type—is guaranteed for various metric spaces.

and that $(2^d + \frac{d+1}{2})$-coalition compromises are sufficient. Table 1 provides an overview of our findings, with pointers to the relevant results in the paper.

We view our work as an important step towards modeling a form of pre-vote deliberation usually not studied within the social choice literature: how voters can identify proposals with large support. Such a theory provides formal foundations for the design and development of practical systems that can support successful deliberation, for instance by helping agents to identify mutually beneficial compromise positions (cf. [di Fenizio and Velikanov, 2016]).

Related Work Group deliberation has been an object of research in several disciplines, from economics to political theory and artificial intelligence. A wealth of different approaches to deliberation can be identified, reflecting the complexity of the concept. Work has focused on axiomatic approaches to deliberation, viewed as an opinion transformation function [List, 2011]. Experimental work—involving lab experiments [List et al., 2013] or simulations [Roy and Rafiee Rad, 2020]—has then tried to assess the effects of deliberations on individual opinions (e.g., whether deliberation facilitates individual opinions to become single-peaked). A prominent approach in modeling deliberation has been the game-theoretic one [Landa and Meirowitz, 2009]. With the tools of game theory, deliberation has been studied under several angles: as a process involving the exchange of arguments for or against positions [Chung and Duggan, 2020, Patty, 2008]; as a process of persuasion [Glazer and Rubinstein, 2001, 2004, 2006]; as the pre-processing of inputs for voting mechanisms [Austen-Smith and Feddersen, 2005, Perote-Pea and Piggins, 2015, Karanikolas et al., 2019]; as a mechanism enabling preference discovery [Hafer and Landa, 2007]. Under a more algorithmic lens, deliberation has been studied as a form of distributed protocol for large-scale decision-making involving only sequential local interaction in small groups [Goel and Lee, 2016, Fain et al., 2017].

Our work sets itself somewhat apart from the above literature by focusing exclusively on the consensus-seeking aspect of deliberation. Deliberation is studied here as a distributed process whereby widely supported proposals can be identified in a decentralized way. In doing so we abstract away from: the concrete interaction mechanisms (e.g., argumentation) that agents may use to communicate and assess proposals; strategic issues that may arise during communication; as well as how deliberation might interact with specific voting rules. Instead, we concentrate on the results of such interactions, as they are manifested by changes in the structure of coalitions supporting different proposals. This focus also sets our work apart from influential opinion dynamics models (e.g., [De Groot, 1974]) where deliberation is driven by social influence rather than by coalition formation via, for instance, compromise.

The fact that we abstract from strategic issues that agents may be confronted with in deciding to join or leave coalitions differentiates our work from the related literature on dynamic coalition formation [Dieckmann and Schwalbe, 2002, Chalkiadakis and Boutilier, 2008]. In particular, Chalkiadakis and Boutilier [2008] concentrate on uncertainties that the agents may experience; and Dieckmann and Schwalbe [2002] focus on agents that receive payoffs, which depend on the coalitions they are in.

From a technical point of view, our work is closely related to work on spatial voting and coalition formation. Many social choice settings are naturally embedded in a metric space, and there is a large literature that considers preference aggregation and coalition formation in such scenarios [Coombs, 1964, Merrill and Grofman, 1999, de Vries, 1999]. In this context we mention, in particular, the work of Shahaf et al. [2019], which considers a framework for voting and proposing, instantiated to several metric spaces, modeling a broad range of social choice settings, and anticipating their use for deliberative decision making. In the model of Shahaf et al. [2019], the approach includes an agent population where each agent is associated with an ideal point, and several general aggregation methods are evaluated based on computational and normative properties. The current paper naturally extends that framework to accommodate the processes of coalition formation.

2 Formal Model

We view deliberation as a process in which agents aim to find an alternative preferred over the status quo by as many agents as possible. Thus, we assume a (possibly infinite) domain X of *alternatives*, or *proposals*, which includes the *status quo*, or *reality*, $r \in X$. We also assume a set $V = \{v_1, \ldots, v_n\}$ of n *agents*. For each proposal $x \in X$, an agent v is able to articulate whether she (strictly) prefers x over the status quo r (denoted as $x >_v r$); when $x >_v r$ we say that v *approves* x. For each $v \in V$, let $X^v = \{x \in X : x >_v r\}$; the set X^v is the *approval set* of v. Conversely, given a subset of agents $C \subseteq V$ and a proposal $x \in X$, let $C^x := \{v \in C : x >_v r\}$; the agents in C^x are the *approvers* of x in C.

Throughout this paper, we focus on the setting where X and V are contained in a metric space (M, ρ), i.e., (1) $X, V \subseteq M$, (2) for every $x \in X$ and every $v \in V$ we have $x >_v r$ if and only if $\rho(v, x) < \rho(v, r)$, and (3) the mapping $\rho : M \times M \to \mathbb{R}^+ \cup \{0\}$ satisfies (i) $\rho(x, y) = 0$ if and only if $x = y$, (ii) $\rho(x, y) = \rho(y, x)$, and (iii) $\rho(x, y) + \rho(y, z) \leq \rho(x, z)$ for all $x, y, z \in M$. E.g., if $M = \mathbb{R}^2$ and the metric is the usual Euclidean metric in $\mathbb{R}^2$, then the approval set of v consists of all points in X that are located inside the circle with center v and radius $\rho(v, r)$ (see Figure 1), whereas the set of approvers of a proposal p in C consists of all agents $v \in C$ such that $\rho(v, p) < \rho(v, r)$ (geometrically, consider the line ℓ that passes through the midpoint of the segment $[p, r]$ and is orthogonal to it; then C^p consists of all agents $v \in C$ such that v and p lie on the same side of ℓ).

Thus, an instance of our problem can be succinctly encoded by a 4-tuple (X, V, r, ρ); we will refer to such tuples as *deliberation spaces*.

Remark 1. *Note that we do not require that $X = M$. By allowing X to be a proper subset of M we can capture the case where the space of proposals is, e.g., a finite subset of $\mathbb{R}^d$ for some $d \geq 1$. Moreover, in our model it need not be the case that $V \subseteq X$, i.e., we do not assume that for each agent there exists a 'perfect' proposal. Furthermore, while for each agent v the quantities $\rho(v, x)$ are well-defined for each $x \in X$, we do not expect the agents to compare different proposals based on distance; rather, the distance only determines which proposals are viewed as acceptable (i.e., preferred to the status quo).*

Agents proceed by forming coalitions around proposals. Thus, at each point in the deliberation, agents can be partitioned into coalitions, so that each coalition C is identified with a proposal p_C and all agents in C support p_C. Agents may then move from one coalition to another as well as select a proposal from X that is not associated with any of the existing coalitions and form a new coalition around it. We consider a variety of permissible moves, ranging from single-agent transitions (when an agent abandons her current coalition and joins a new one), to complex transitions that may involve agents from multiple coalitions and a new proposal. In each case, we assume that agent v is unwilling to join a coalition if

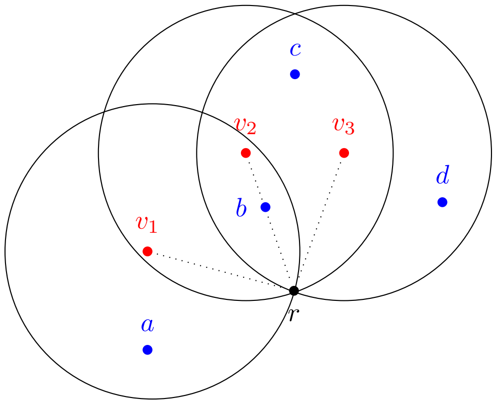

Figure 1: A deliberation space with $X = \{a, b, c, d, r\}$. The circle with center v_i, $i \in [3]$, contains all proposals approved by v_i.

this coalition advocates a proposal p such that v prefers the status quo r to p. To reason about the coalition formation process, we introduce additional notation and terminology.

Definition 1 (Deliberative Coalition)**.** *A* deliberative coalition *is a pair* $\mathbf{d} = (C, p)$*, where* $C \subseteq V$ *is a set of agents,* $p \in X$ *is a proposal, and either (i)* $p = r$ *and* $x \not>_v r$ *for all* $v \in C, x \in X \setminus \{r\}$*, or (ii)* $p \neq r$ *and* $p >_v r$ *for all* $v \in C$*. We say that agents in* C support p*. The set of all deliberative coalitions is denoted by* $\mathcal{D}$*.*

When convenient, we identify a coalition $\mathbf{d} = (C, p)$ with its set of agents C and write $\mathbf{d}^p := C^p$, $|\mathbf{d}| := |C|$.

Remark 2. *While we allow coalitions that support the status quo, we require that such coalitions consist of agents who weakly prefer the status quo to all other proposals in* X*. We discuss a relaxation of this constraint in Section 8.*

A partition of the agents into deliberative coalitions is called a *deliberative coalition structure*.

Definition 2 (Deliberative Coalition Structure)**.** *A* deliberative coalition structure *(*coalition structure *for short) is a set* $\mathbf{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_m\}$*,* $m \geq 1$*, such that:*

- $\mathbf{d}_i = (C_i, p_i) \in \mathcal{D}$ *for each* $i \in [m]$*;*
- $\bigcup_{i \in [m]} C_i = V$*,* $C_i \cap C_j = \emptyset$ *for all* $i, j \in [m]$ *with* $i \neq j$*.*

The set of all deliberative coalition structures over V *and* X *is denoted by* $\mathfrak{D}$*.*

Note that a deliberative coalition structure may contain several coalitions supporting the same proposal; also, for technical reasons we allow empty deliberative coalitions, i.e., coalitions (C, p) with $C = \emptyset$.

Example 1. *Consider a set of agents* $V = \{v_1, v_2, v_3\}$ *and a space of proposals* $X = \{a, b, c, d, r\}$*, where* r *is the status quo. Suppose that* $X^{v_1} = \{a, b\}$*,* $X^{v_2} = \{b, c\}$*, and* $X^{v_3} = \{b, c, d\}$*. Then for* $C = \{v_1, v_2\}$ *we have* $C^a = \{v_1\}$ *and* $C^b = \{v_1, v_2\}$*. Furthermore, let* $C_1 = \{v_1, v_2\}$*,* $C_2 = \{v_3\}$*, and let* $\mathbf{d}_1 = (C_1, b)$*,* $\mathbf{d}_2 = (C_2, c)$*. Then* $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$ *is a deliberative coalition structure; see Figure 1 for an illustration.*

Remark 3. *An important feature of our model is that agents do not need to explicitly list all proposals they approve of, or reason about them. At any time during the deliberation, each agent supports a single proposal in her approval set. Thus, this model can be used even if the agents have limited capability to reason about the available proposals.*

We model deliberation as a process whereby deliberative coalition structures change as their constituent coalitions evolve, i.e., we view deliberation as a transition system. Formally, a *transition system* is characterized by a set of *states* S, a subset $S_0 \subseteq S$ of *initial states*, and a set of *transitions* T, where each transition $t \in T$ is represented by a pair of states $(s, s') \in S \times S$; we write $t = (s, s')$ and $s \xrightarrow{t} s'$ interchangeably. We use $s \xrightarrow{T} s'$ to denote that $s \xrightarrow{t} s'$ for some $t \in T$. A *run* of a transition system is a sequence $s_0 \xrightarrow{T} s_1 \xrightarrow{T} s_2 \cdots$ with $s_0 \in S_0$. The last state of a finite run is called its *terminal state*.

Definition 3 (Deliberative Transition System)**.** *A* deliberative transition system *over a set of proposals X and a set of agents V is a transition system that has $\mathfrak{D}$ as its set of states, a subset of states $\mathfrak{D}_0 \subseteq \mathfrak{D}$ as its set of initial states, and a set of transitions $\mathcal{S}$.*

Maximal runs over a deliberative transition system can be seen as outcomes of the deliberative process.

Definition 4 (Deliberation)**.** *A* deliberation *is a maximal run of a deliberative transition system, that is, a run that does not occur as a prefix of any other run.*

A *successful* deliberation is one that identifies some of the most popular proposals in $X \setminus \{r\}$, with all agents who approve that proposal forming a coalition that supports it. In particular, if there is a majority-approved proposal, a successful deliberation process allows the agent population to identify some such proposal; this can then result in a majority-supported change to the status quo.

Definition 5 (Successful Deliberation)**.** *For a set of agents V and a set of proposals X, the set of* most-supported alternatives *is $M^* = argmax_{x \in X \setminus \{r\}} |V^x|$ and the* maximum support *is $m^* = max_{x \in X \setminus \{r\}} |V^x|$. A deliberative coalition $\mathbf{d} = (C, p)$ is* successful *if $p \in M^*$ and $|C| = m^*$. A coalition structure is* successful *if it contains a successful coalition, and* unsuccessful *otherwise. A deliberation is* successful *if it is finite and its terminal coalition structure is successful.*

Note that if all approval sets are non-empty, $M^* \neq \emptyset$.

Below we consider specific types of transitions; each can be thought of as a *deliberation operator* that might be available to the agents. For each transition type, we aim to determine if a deliberation that only uses such transitions is guaranteed to terminate, and, if so, whether the final coalition structure is guaranteed to be successful. We show that the answer to this question depends on the underlying metric space: simple transition rules guarantee success in simple metric spaces, but may fail in richer spaces.

3 Single-Agent Transitions

The simplest kind of transition we consider is a deviation by a single agent. As we assume that agents aim to form a successful coalition and are not necessarily able to distinguish among approved proposals, it is natural to focus on transitions where an agent moves from a smaller group to a larger group; of course, this move is only possible if the agent approves the proposal supported by the larger group.

Definition 6 (Single-Agent Transition)**.** *A pair $(\mathbf{D}, \mathbf{D}')$ of coalition structures forms a* single-agent transition *if there exist coalitions $\mathbf{d}_1, \mathbf{d}_2 \in \mathbf{D}$ and $\mathbf{d}'_1, \mathbf{d}'_2 \in \mathbf{D}'$ such that $|\mathbf{d}_2| \geq |\mathbf{d}_1|$, $\mathbf{D} \setminus \{\mathbf{d}_1, \mathbf{d}_2\} = \mathbf{D}' \setminus \{\mathbf{d}'_1, \mathbf{d}'_2\}$, and there exists an agent $v \in \mathbf{d}_1$ such that $\mathbf{d}'_1 = \mathbf{d}_1 \setminus \{v\}$, and $\mathbf{d}'_2 = \mathbf{d}_2 \cup \{v\}$.*[3] *We refer to v as the* deviating agent.

[3] Note that $\mathbf{d}'_1$ may be empty; we allow such 'trivial' coalitions as it simplifies our definitions.

Since $\mathbf{D}'$ is a deliberative coalition structure, agent v must approve the proposal supported by $\mathbf{d}_2$. As a consequence, no agent can deviate from a coalition that supports r to a coalition that supports some $p \in X \setminus \{r\}$ or vice versa.

An easy potential function argument shows that a sequence of single-agent transitions necessarily terminates after polynomially many steps.

Proposition 1. *A deliberation that consists of single-agent transitions can have at most n^2 transitions.*

However, a deliberation consisting of single-agent transitions is not necessarily successful, even for very simple metric spaces. The next example shows that such a deliberation may fail to identify a majority-supported proposal even if the associated metric space is the 1D Euclidean space.

Example 2. *Suppose that $X, V \subseteq \mathbb{R}$, and $X = \{r, a, b, c\}$ with $r = 0$, $a = 1$, $b = 5$, $c = -1$. There are three agents v_1, v_2, v_3 located at a, four agents v_4, v_5, v_6, v_7 located at b, and three agents v_8, v_9, v_{10} located at c. Observe that a majority of the agents prefer a to r. Consider the deliberative coalition structure $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ with $\mathbf{d}_1 = (C_1, a)$, $\mathbf{d}_2 = (C_2, b)$, $\mathbf{d}_3 = (C_3, c)$, and such that $C_1 = \{v_1, v_2, v_3\}$, $C_2 = \{v_4, v_5, v_6, v_7\}$, and $C_3 = \{v_8, v_9, v_{10}\}$. There are no single-agent transitions from this coalition structure: in particular, the agents in $\mathbf{d}_2$ do not want to deviate to $\mathbf{d}_1$ because $|\mathbf{d}_1| < |\mathbf{d}_2|$ and the agents in $\mathbf{d}_1$ do not want to deviate to $\mathbf{d}_2$, because they do not approve b.*

Note that this argument still applies to any proposal space X' with $X \subseteq X'$; e.g., we can take $X' = \mathbb{R}$.

Thus, to ensure success, we need to consider more powerful transitions.

4 Follow Transitions

Instead of a single agent we focus now on the case in which the entirety of a coalition decides to join another coalition in supporting that coalition's current proposal. We call such transitions 'follow' transitions.

Definition 7 (Follow Transition)**.** *A pair of coalition structures $(\mathbf{D}, \mathbf{D}')$ forms a* follow transition *if there exist non-empty coalitions $\mathbf{d}_1, \mathbf{d}_2 \in \mathbf{D}$ and $\mathbf{d}'_2 \in \mathbf{D}'$ such that $\mathbf{d}_1 = (C_1, p_1)$, $\mathbf{d}_2 = (C_2, p_2)$, $\mathbf{D} \setminus \{\mathbf{d}_1, \mathbf{d}_2\} = \mathbf{D}' \setminus \{\mathbf{d}'_2\}$, and $\mathbf{d}'_2 = (C_1 \cup C_2, p_2)$.*

Note that each follow transition reduces the number of coalitions by one, so a deliberation that consists only of follow transitions converges in at most n steps. Also, it can be shown that follow transitions increase the potential function defined in the proof of Proposition 1. This implies the following bound.

Proposition 2. *A deliberation that consists of single-agent transitions and follow transitions can have at most n^2 transitions.*

However, in contrast to single-agent transitions, follow transitions are sufficient for successful deliberation in a one dimension Euclidean space.

Theorem 1. *Consider a deliberation space (X, V, r, ρ), where $X, V \subseteq \mathbb{R}$ and $\rho(x, y) = |x-y|$. Then any deliberation that consists of follow transitions only, or of a combination of follow transitions and single-agent transitions, is successful.*

Unfortunately, Theorem 1 does not extend beyond one dimension. The following examples show that in the Euclidean plane a deliberation that only uses single-agent transitions and follow transitions is not necessarily successful.

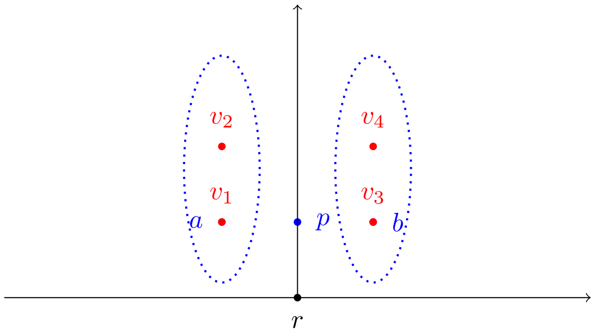

Figure 2: Illustration of Example 3: single-agent and follow transitions are not sufficient to identify a most-supported proposal.

Example 3. *Consider a space of proposals $\{a, b, p, r\}$ embedded into $\mathbb{R}^2$, where r is located at $(0,0)$, p is located at $(0,3)$, a is located at $(-3,3)$, and b is located at $(3,3)$. There are four agents v_1, v_2, v_3, v_4 located at $(-3,3)$, $(-3,4)$, $(3,3)$, and $(3,4)$, respectively. Note that all agents prefer p to r. Let $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$, where $\mathbf{d}_1 = (\{v_1, v_2\}, a)$, $\mathbf{d}_2 = (\{v_3, v_4\}, b)$. Then no agent in $\mathbf{d}_1$ approves b, and no agent in $\mathbf{d}_2$ approves a, so there are no follow transitions and no single-agent transitions from $\mathbf{D}$. See Figure 2.*

5 Merge Transitions

So far, we have focused on transitions that did not introduce new proposals. Example 3, however, shows that new proposals may be necessary to reach success: indeed, none of the proposals supported by existing coalitions in this example had majority support. Thus, next we explore transitions that identify new proposals.

As a first step, it is natural to relax the constraint in the definition of the follow transitions that requires the new coalition to adopt the proposal of one of the two component coalitions, and, instead, allow the agents to identify a new proposal that is universally acceptable.

We do not specify how the compromise proposal p is identified. In practice, we expect that the new proposal would be put forward by one of the agents in $\mathbf{d}_1, \mathbf{d}_2$ or by an external mediator whose goal is to help the agents reach a consensus.

Definition 8 (Merge Transition)**.** *A pair of coalition structures $(\mathbf{D}, \mathbf{D}')$ forms a* merge transition *if there exist non-empty coalitions $\mathbf{d}_1, \mathbf{d}_2 \in \mathbf{D}$ and $\mathbf{d}'_2 \in \mathbf{D}'$ such that $\mathbf{d}_1 = (C_1, p_1)$, $\mathbf{d}_2 = (C_2, p_2)$, $\mathbf{D} \setminus \{\mathbf{d}_1, \mathbf{d}_2\} = \mathbf{D}' \setminus \{\mathbf{d}'_2\}$, and $\mathbf{d}'_2 = (C_1 \cup C_2, p)$ for some proposal p.*

One can verify that in Example 3 the agents have a merge transition to the deliberative coalition $(\{v_1, v_2, v_3, v_4\}, p)$; indeed, merge transitions are strictly more powerful than follow transitions. Moreover, our potential function argument shows that a deliberation that consists of single-agent transitions and merge transitions can have at most n^2 steps, implying the following bound.

Proposition 3. *A deliberation that consists of single-agent transitions and merge transitions can have at most n^2 transitions.*

We will now describe a class of metric spaces where merge transitions guarantee convergence, but follow transitions do not.

Trees Consider a weighted tree T with a set of vertices U, a set of edges E, and a weight function $\omega : E \to \mathbb{R}^+$. This tree defines a metric space (M_T, ρ_T), where $M_T = U$, and for a pair of vertices $x, y \in U$ the distance $\rho_T(x, y)$ is the length of the path between x and y in T, where the length of an edge e is given by $\omega(e)$. Consequently, a tree T defines a family of deliberation spaces (X, V, r, ρ), where $X \subseteq M_T$, $V \subseteq M_T$, $r \in X$ and $\rho = \rho_T$. It turns out that for any such deliberation space merge transitions guarantee successful termination.

Theorem 2. *Consider a deliberation space (X, V, r, ρ) that corresponds to a tree $T = (U, E, \omega)$ with root r. Then any deliberation that consists of merge transitions only, or of a combination of merge transitions and single-agent transitions, is successful.*

Note that, for trees, we indeed need the full power of merge transitions: follow or single-agent transitions are insufficient for successful deliberation. This holds even if all edges have the same weight and is demonstrated by the following example.

Example 4. *Consider a tree T with vertex set $U = \{r, u_1, u_2, u_3\}$ and edge set $E = \{\{r, u_1\}, \{u_1, u_2\}, \{u_1, u_3\}\}$; the weight of each edge is 1. Let $X = U$, and suppose that there are two agents with ideal points in u_2 and u_3, respectively. Consider a coalition structure where each of these agents forms a singleton coalition around her ideal point. Note that $\rho(u_2, r) = \rho(u_3, r) = \rho(u_2, u_3)$ and hence the agents do not approve each other's positions. Therefore, there are no follow transitions or single-agent transitions from this coalition structure (even though u_1 is approved by both agents).*

However, the next example shows that for Euclidean spaces, merge and single-agent transitions are insufficient for successful deliberation.

Example 5. *We can modify Example 3 by adding an agent v_5 at $(-4, 0)$ to $\mathbf{d}_1$ and an agent v_6 at $(4, 0)$ to $\mathbf{d}_2$; let $\mathbf{D}'$ be the resulting coalition structure. Note that v_5 approves a, but does not approve p, thereby preventing a merge transition between $\mathbf{d}_1$ and $\mathbf{d}_2$ at p.*

6 Compromise Transitions

Example 5 illustrates that, to reach a successful outcome, coalitions may need to leave some of their members behind when joining forces. Next we formalize this idea.

Definition 9 (Compromise Transition). *A pair of coalition structures $(\mathbf{D}, \mathbf{D}')$ forms a* compromise transition *if there exist coalitions $\mathbf{d}_1, \mathbf{d}_2 \in \mathbf{D}$, $\mathbf{d}'_1, \mathbf{d}'_2, \mathbf{d}' \in \mathbf{D}'$ and a proposal p such that $\mathbf{D} \setminus \{\mathbf{d}_1, \mathbf{d}_2\} = \mathbf{D}' \setminus \{\mathbf{d}'_1, \mathbf{d}'_2, \mathbf{d}'\}$, $\mathbf{d}_1 = (C_1, p_1)$, $\mathbf{d}_2 = (C_2, p_2)$, $\mathbf{d}' = (C_1^p \cup C_2^p, p)$, $\mathbf{d}'_1 = (C_1 \setminus C_1^p, p_1)$, $\mathbf{d}_2 = (C_2 \setminus C_2^p, p_2)$, and $|C_1^p \cup C_2^p| > |C_1|, |C_2|$.*

Intuitively, under a compromise transition some of the agents in $\mathbf{d}_1$ and $\mathbf{d}_2$ identify a suitable proposal p, and then those of them who approve p move to form a coalition that supports p, leaving the rest of their old friends behind; a necessary condition for the transition is that the new coalition should be larger than both $\mathbf{d}_1$ and $\mathbf{d}_2$.

Example 6. *In Example 5 there is a compromise transition from $\mathbf{D}'$ to the coalition structure $(\mathbf{d}'_1, \mathbf{d}'_2, \mathbf{d}')$, where $\mathbf{d}'_1 = (\{v_5\}, a)$, $\mathbf{d}'_2 = (\{v_6\}, b)$, and $\mathbf{d}' = (\{v_1, v_2, v_3, v_4\}, p)$. The new coalition has size 4, so the resulting coalition structure is successful.*

Importantly, we assume that all agents in $\mathbf{d}_1$ and $\mathbf{d}_2$ who support p join the compromise coalition; indeed, this is what we expect to happen if the agents myopically optimize the size of their coalition.

An important feature of compromise transitions is that they ensure termination.

Proposition 4. *A deliberation that consists of compromise transitions can have at most n^n transitions.*

In contrast to the case of single-agent transitions and follow/merge transitions, we are unable to show that a deliberation consisting of compromise transitions always terminates after polynomially many steps; it remains an open problem whether this is indeed the case. We note that a compromise transition does not necessarily increase the potential function defined in the proof of Proposition 1.

We say that a deliberation space (X, V, r, ρ) is a *Euclidean deliberation space* if $X = \mathbb{R}^d$, $V \subseteq \mathbb{R}^d$ for some $d \geq 1$, and ρ is the standard Euclidean metric on $\mathbb{R}^d$. The main result of this section is that, in every Euclidean deliberation space, every maximal run of compromise transitions is successful.

Theorem 3. *In every Euclidean deliberation space, every maximal run of compromise transitions is a successful deliberation.*

The argument in the proof of Theorem 3 relies on the underlying metric space being sufficiently rich. The argument goes through if we replace $\mathbb{R}^d$ with $\mathbb{Q}^d$; however, it does not extend to the case where X is an arbitrary finite subset of $\mathbb{R}^d$. Intuitively, for deliberation to converge, at least some agents should be able to spell out nuanced compromise proposals.

Further, a close inspection of the proof of Theorem 3 shows for every Euclidean deliberation space there exists a successful deliberation that consists of compromise transitions and has length at most $n^2 + 1$.

Given our positive results for Euclidean deliberation spaces, it is natural to ask whether compromise transitions are sufficient for convergence in other metric spaces. However, in the next section we will see that this is not always the case.

7 Beyond Two-Way Compromises?

In this section we step away from the Euclidean framework, and focus on spaces that naturally arise in decision-making on combinatorial domains and multiple referenda [Lang and Xia, 2016]: d-dimensional hypercubes. These consist of the set $\{0, 1\}^d$ of binary vectors of length d (d being a positive integer), endowed with the discrete metric h, defined as $h((x_1, \ldots, x_d), (y_1, \ldots, y_d)) = \sum_{i=1}^{d} |x_i - y_i|$. This metric is known as Hamming, or Manhattan, distance. Intuitively, each element of $\{0, 1\}^d$ denotes one possible position on a set of d binary issues. For any positive integer d, we refer to the space $(\{0, 1\}^d, h)$ as the *d-dimensional hypercube* or simply *d-hypercube*. In what follows, for readability, we will often write vectors in $\{0, 1\}^d$ as strings of 1s and 0s of length d: e.g., we will write 0110 instead of $(0, 1, 1, 0)$. Also, without loss of generality we will assume throughout that the status quo r is the all-0 vector.

We start with a simple observation.

Proposition 5. *In d-hypercubes, with $d \leq 2$, any deliberation consisting of compromise transitions is successful. However, single-agent, follow, or merge transitions may be insufficient for convergence.*

In higher dimensions, however, compromise transitions no longer guarantee success.

Example 7 (Compromise failure in the 3-hypercube)**.** *Consider the* 3*-hypercube* $\{0, 1\}^3$ *and let N consist of the following agents: v_1 at* 100*, v_2 at* 010*, v_3 at* 110*, v_4 at* 001*, and v_5 at* 101 *(see Figure 3). Consider then the deliberative coalition structure consisting of the following three coalitions: $(\{v_1\}, 100)$, $(\{v_2, v_3\}, 010)$, and $(\{v_4, v_5\}, 001)$. This coalition structure is terminal. It is not successful, however, as v_1, v_3 and v_5 approve* 100.

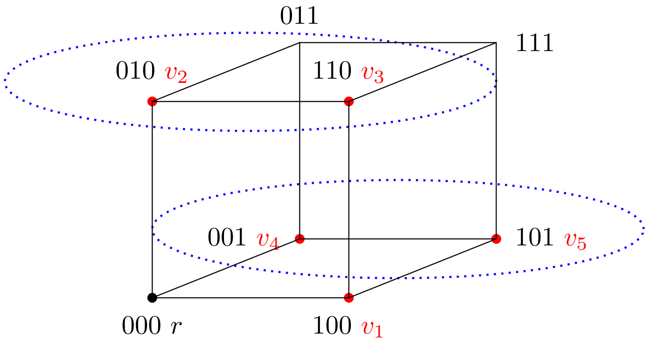

Figure 3: The deliberative space in Example 7.

In the example above, a successful outcome would be within reach if compromises that involve three coalitions were allowed. This motivates us to study a natural multi-party generalization of compromise transitions.

Definition 10 (ℓ-Compromise Transition)**.** *A pair of coalition structures* $(\mathbf{D}, \mathbf{D}')$ *forms an* ℓ-compromise transition*, with* $2 \leq \ell$ *integer, if there exist coalitions* $\mathbf{d}_1, \ldots, \mathbf{d}_\ell \in \mathbf{D}$*,* $\mathbf{d}'_1, \ldots, \mathbf{d}'_\ell, \mathbf{d}' \in \mathbf{D}'$ *and a proposal* p *such that, for all* $1 \leq i \leq \ell$ *it holds that* $\mathbf{D} \setminus \{\mathbf{d}_1, \ldots, \mathbf{d}_\ell\} = \mathbf{D}' \setminus \{\mathbf{d}'_1, \ldots, \mathbf{d}'_\ell, \mathbf{d}'\}$*,* $\mathbf{d}_i = (C_i, p_i)$*,* $\mathbf{d}' = (\bigcup_{1 \leq i \leq \ell} C_i^p, p)$*,* $\mathbf{d}'_i = (C_i \setminus C_i^p, p_i)$*, and* $|\bigcup_{1 \leq i \leq \ell} C_i^p| > |C_i|$*.*

Indeed, a compromise transition (Definition 9) is a 2-compromise transition. Notice also that, in an ℓ-compromise, some coalitions C_i may actually contribute an empty set C_i^p of agents to the newly formed coalition (although this cannot occur for $\ell = 2$). So, for every ℓ, ℓ' with $2 \leq \ell < \ell' \leq |\mathbf{D}|$, it holds that an ℓ-compromise transition $(\mathbf{D}, \mathbf{D}')$ is also an ℓ'-compromise transition, as any coalition C not involved in the ℓ-compromise trivially contributes $\emptyset$ to the newly formed coalition.

The argument in the proof of Proposition 4 establishes that a sequence of ℓ-compromises has length at most n^n.

Corollary 1. *For each* $\ell \geq 2$*, a deliberation that consists of* ℓ*-compromise transitions can have at most* n^n *transitions.*

At the extreme, if the agents can engage in n-compromises, then a successful coalition structure can be reached in one step, by having all agents who approve a proposal $p \in M^*$ form a coalition around p. However, we are interested in outcomes that can be achieved when each interaction only involves a few coalitions. Thus, given a deliberation space, we are interested in finding the smallest value of ℓ such that ℓ-compromise transitions can reach a successful outcome. Formally, let $\ell^*(d)$ be the smallest value of ℓ such that in a d-hypercube a deliberation that consists of ℓ-compromise transitions is necessarily successful.

We will now establish some lower and upper bounds on $\ell^*(d)$, which hold for all $d \geq 3$.

Proposition 6. *For each* $d \geq 3$ *we have* $\ell^*(d) \geq d$*.*

It is worth contrasting Proposition 6 with the earlier Theorem 3: While d-Euclidean spaces are 'rich' enough for 2-compromises to guarantee success in any dimension, the 'sparsity' of d-hypercubes is such that we need to facilitate interactions among at least d coalitions to guarantee success.

For the upper bound, an easy observation is that $\ell^*(d) \leq 2^d$: in a terminal coalition structure there cannot be two distinct coalitions supporting the same proposal. We will now show how to strengthen this bound by (almost) a factor of two.

Proposition 7. *For each $d \geq 2$ we have $\ell^*(d) \leq 2^{d-1} + \frac{d+1}{2}$.*

The gap between the lower bound of Proposition 6 and the upper bound of Proposition 7 is very large. In what follows, we compute the exact value of $\ell^*(d)$ for $d = 3$ and $d = 4$; we hope that the intuition developed in the analysis of these cases will be useful for tackling the general case. For $d = 3$, the lower bound of Proposition 6 turns out to be tight; however, for $d = 4$ this is no longer the case.

Proposition 8. *We have $\ell^*(3) = 3$ and $\ell^*(4) = 5$.*

8 Conclusions and Future Work

We proposed a formal model of deliberation for agent populations forming coalitions around proposals in order to change the status quo. We identified several natural modes of coalition formation capturing the many coalitional effects that deliberative processes could support. We studied sufficient conditions for them to succeed in identifying maximally supported proposals. We intend our model as a foundation for the study of mechanisms and systems allowing communities to self-govern.

To the best of our knowledge, ours is the first model of deliberation focusing on iterative processes of coalition formation. The model lends itself to several avenues for future research. First of all, there are two technical challenges left unanswered by our work: (1) does a sequence of compromise transitions always converge in polynomial time? (2) what are the values of $\ell^*(d)$ for $d > 4$? Then, natural extensions of the model could be considered involving other transition types as well as other types of metric spaces. In particular, it will be important to understand exactly how 'rich' a space should be in order to support success for a specific type of deliberation. One can also ask whether our positive results are preserved if we impose additional conditions on the structure of the deliberative process, e.g., require new proposals to be 'close' to the original proposals. We may also revisit our approach to modeling coalitions that support the status quo: we now assume that each agent v with $X^v \neq \emptyset$ is capable of identifying some proposal in X^v, and this assumption may be too strong for many deliberation scenarios. It is perhaps more realistic to assume that some agents start out by supporting the status quo, and then learn about a new proposal p that they prefer to the status quo by observing a coalition that supports p, and then move to join this coalition. Investigating a model that permits such transitions is another direction for future research we consider promising.

Further afield, it would be interesting to consider a stochastic variant of our model, in which each transition from a state is assigned a certain probability. A Markovian analysis of such systems might shed further light on the behavior of deliberative processes.

Crucially, in our current model agents are not strategic: they truthfully reveal whether they support a given proposal. Yet, revealed support for proposals may well be object of manipulation. Such a game-theoretic extension is also an obvious direction for future research. Another ambitious direction for future work is to design a practical tool for deliberation and self-government of an online community, building on top of our model and analysis. Such a tool could take the form of an AI bot over existing on-line deliberation platforms such as the LiquidFeedback and Polis platforms mentioned earlier. The bot would suggest proposals to agents in order to support compromise, hopefully fostering successful deliberations.

References

David Austen-Smith and Timothy J. Feddersen. Deliberation and voting rules. In *Social Choice and Strategic Decisions*, Studies in Social Choice and Welfare. Springer, 2005.

Jan Behrens, Axel Kistner, Andreas Nitsche, and Björn Swierczek. *Principles of Liquid Feedback*. Interaktive Demokratie, 2014.

Markus Brill. Interactive democracy. In *Proceedings of AAMAS '18*, pages 1183–1187, 2018.

Georgios Chalkiadakis and Craig Boutilier. Sequential decision making in repeated coalition formation under uncertainty. In *Proceedings of AAMAS '08*, pages 347–354, 2008.

Hun Chung and John Duggan. A formal theory of democratic deliberation. *American Political Science Review*, 114(1):1435, 2020.

Clyde H. Coombs. *A theory of data*. Wiley, 1964.

Morris De Groot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

Miranda W. M. de Vries. *Governing with your closest neighbour : an assessment of spatial coalition formation theories*. Radboud University Nijmegen, 1999.

Pietro Speroni di Fenizio and Cyril Velikanov. System-generated requests for rewriting proposals. *CoRR*, abs/1611.10095, 2016. URL `http://arxiv.org/abs/1611.10095`.

Tone Dieckmann and Ulrich Schwalbe. Dynamic coalition formation and the core. *Journal of Economic Behavior & Organization*, 49(3):363–380, 2002.

Edith Elkind, Davide Grossi, Ehud Shapiro, and Nimrod Talmon. United for change: Deliberative coalition formation to change the status quo. In *Proceedings of AAAI '21*, 2021.

Brandon Fain, Ashish Goel, Kamesh Munagala, and Sukolsak Sakshuwong. Sequential deliberation for social choice. In *Proceedings of WINE '17*, pages 177–190, 2017.

Jacob Glazer and Ariel Rubinstein. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior*, 36(2):158–173, 2001.

Jacob Glazer and Ariel Rubinstein. On optimal rules of persuasion. *Econometrica*, pages 1715–1736, 2004.

Jacob Glazer and Ariel Rubinstein. A study in the pragmatics of persuasion: A game theoretical approach. *Theoretical Economics*, 1:395–410, 2006.

Ashish Goel and David T Lee. Towards large-scale deliberative decision-making: Small groups and the importance of triads. In *Proceedings of EC '16*, pages 287–303, 2016.

Catherine Hafer and Dimitri Landa. Deliberation as self-discovery and institutions for political speech. *Journal of Theoretical Politics*, 19(3):329–360, 2007.

Nikos Karanikolas, Pierre Bisquert, and Christos Kaklamanis. A voting argumentation framework: Considering the reasoning behind preferences. In *Proceedings of ICAART '19*, pages 42–53, 2019.

Dimitri Landa and Adam Meirowitz. Game theory, information, and deliberative democracy. *American Journal of Political Science*, 53(2):427–444, 2009.

Jérôme Lang and Lirong Xia. Voting in combinatorial domains. In *Handbook of Computational Social Choice*, pages 197–222. Cambridge University Press, 2016.

Christian List. Group communication and the transformation of judgments: An impossibility result. *The Journal of Political Philosophy*, 19(1):1–27, 2011.

Christian List, Robert Luskin, James Fishkin, and Iain McLean. Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. *The Journal of Politics*, 75(1):80–95, 2013.

Samuel Merrill and Bernard Grofman. *A unified theory of voting: Directional and proximity spatial models*. Cambridge University Press, 1999.

John W. Patty. Arguments-based collective choice. *Journal of Theoretical Politics*, 20(4): 379–414, 2008.

Juan Perote-Pea and Ashley Piggins. A model of deliberative and aggregative democracy. *Economics and Philosophy*, 31(1):93121, 2015.

Olivier Roy and Soroush Rafiee Rad. Deliberation, single-peakedness, and coherent aggregation. *American Political Science Review*, 2020.

Gal Shahaf, Ehud Shapiro, and Nimrod Talmon. Aggregation over metric spaces: Proposing and voting in elections, budgeting, and legislation. In *Proceedings of ADT '19*, 2019.

William S. Zwicker. Introduction to the theory of voting. In *Handbook of Computational Social Choice*, pages 23–56. Cambridge University Press, 2016.

A Proofs for Sections 3–6

Proof of Proposition 1. Given a coalition structure $\mathbf{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_m\}$ such that $\mathbf{d}_i = (C_i, p_i)$ for each $i \in [m]$, let $\lambda(\mathbf{D}) = \sum_{i \in [m]} |C_i|^2$; we will refer to $\lambda(\mathbf{D})$ as the *potential* of $\mathbf{D}$. Consider a single-agent transition where an agent moves from a coalition of size x to a coalition of size y; note that $1 \leq x \leq y$. This move changes the potential by $(y+1)^2 + (x-1)^2 - y^2 - x^2 = 2 + 2(y-x) \geq 2$. Now, we claim that for every deliberative coalition structure $\mathbf{D}$ over n agents we have $\lambda(\mathbf{D}) \leq n^2$.

Indeed, for the coalition structure $\mathbf{D}_0$ where all agents are in one coalition we have $\lambda(\mathbf{D}_0) = n^2$. On the other hand, if a coalition structure contains two non-empty coalitions $\mathbf{d}_1$, $\mathbf{d}_2$ with $|\mathbf{d}_1| \leq |\mathbf{d}_2|$, then the calculation above shows that we can increase the potential by moving one agent from $\mathbf{d}_1$ to $\mathbf{d}_2$. Further, every coalition structure can be transformed into $\mathbf{D}_0$ by a sequence of such moves, and hence $\lambda(\mathbf{D}_0) \geq \lambda(\mathbf{D})$ for each $\mathbf{D} \in \mathfrak{D}$. As every single-agent transition increases the potential by at least 2, and the potential takes values in $\{1, \dots, n^2\}$, the claim follows. □

Proof of Theorem 1. Assume for convenience that $r = 0$. Consider a deliberation that consists of single-agent transitions and follow transitions. By Proposition 2 we know that it is finite; let $\mathbf{D}$ be its terminal state, let p be some proposal in M^*, and assume without loss of generality that $p > 0$.

Suppose that $\mathbf{D}$ contains two deliberative coalitions (C_1, p_1) and (C_2, p_2) with $p_1, p_2 \in \mathbb{R}^+$; assume without loss of generality that $p_1 \leq p_2$. Note that $C_1, C_2 \subseteq \mathbb{R}^+$: all agents in $\mathbb{R}^- \cup \{0\}$ prefer r to p_1, p_2. Furthermore, every agent in C_2 approves p_1: indeed, if $v \in C_2$ does not approve p_1 then $|v - r| \leq |v - p_1|$, i.e., $v \leq |v - p_1|$. Since $v, p_1 > 0$, this would imply $2v \le p_1 \le p_2$, in which case v would not approve p_2 either. Hence there is a follow transition in which C_2 joins C_1, a contradiction with $\mathbf{D}$ being a terminal coalition structure.

Thus, $\mathbf{D}$ contains at most one coalition, say, (C^+, q^+), that supports a proposal in $\mathbb{R}^+$; by the same argument, it also contains at most one coalition, say, (C^-, q^-), that supports a proposal in $\mathbb{R}^-$ and at most one coalition, say, (C^0, r), that supports r. Since agents in $\mathbb{R}^-$ prefer r to p, we have $C^- \cap V^p = \emptyset$; also, by definition all agents in C^0 weakly prefer r to p. Hence $V^p \subseteq C^+$ and therefore $|C^+| = m^*$. □

Proof of Theorem 2. It will be convenient to think of T as a rooted tree with root r and, for each $x \in U$, denote by T_x the subtree of T that is rooted in x.

Assume for convenience that $X = M_T$, i.e., every vertex of T is a feasible proposal. Let p be some proposal in M^*, and let q be the child of r such that the path from p to r passes through q (it is possible that $p = q$). We claim that $q \in M^*$. Indeed, an agent approves q if and only if her ideal point is located at T_q. On the other hand, if an agent's ideal point is not located in T_q, then the path from her ideal point to p passes through the root r and hence she does not approve p. Hence, the set of agents who approve p forms a subset of the set of agents who approve q, establishing our claim.

This argument also shows that there are exactly m^* agents whose ideal point is located in T_q. Each such agent can only be in a deliberative coalition that supports a proposal in T_q: indeed, if an agent's ideal point is in T_q, then she does not approve proposals outside of T_q.

Now, consider a terminal coalition structure $\mathbf{D}$ and suppose that there does not exist a deliberative coalition of size m^* supporting a proposal in T_q. Then, there are multiple deliberative coalitions that support proposals in T_q and consist of agents whose ideal points are in T_q. Let (C, p') and (C, p'') be two such coalitions. But then there is a merge transition in which coalition $(C' \cup C'', q)$ forms, a contradiction with $\mathbf{D}$ being a terminal coalition structure.

If $X \neq M_T$, then we can use a similar argument; the difference is that we have to define q by considering the ancestors of p (other than r) that belong to X, and, among these, pick one that is closest to r. □

Proof of Proposition 4. Given a coalition structure $\mathbf{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_m\}$ such that $\mathbf{d}_i = (C_i, p_i)$ for each $i \in [m]$, $|C_1| \ge \dots \ge |C_m|$, and there exists an $\ell \in [m]$ such that $|C_i| > 0$ for $i \in [\ell]$, $|C_i| = 0$ for $i = \ell+1, \dots, m$, let $\gamma(\mathbf{D}) = (|C_1|, \dots, |C_\ell|)$. Note that $\gamma(\mathbf{D})$ is a non-increasing sequence of positive integers. Given two non-increasing sequences $(a_1, \dots, a_s)$, $(b_1, \dots, b_t)$ of positive integers we write $(a_1, \dots, a_s) <_{\text{lex}} (b_1, \dots, b_t)$ if either (a) there exists a $j \le \min\{s, t\}$ such that $a_i = b_i$ for all $i < j$ and $a_j < b_j$, or (b) $s < t$, and $a_i = b_i$ for all $i \in [s]$. Now, observe that if $(\mathbf{D}, \mathbf{D}')$ is a compromise transition, then for the respective coalitions $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}'$ we have $|\mathbf{d}'| > \max\{|\mathbf{d}_1|, |\mathbf{d}_2|\}$, and hence

$$\gamma(\mathbf{D}) <_{\text{lex}} \gamma(\mathbf{D}') \tag{1}$$

Now, consider a sequence of deliberative coalition structures $\mathbf{D}_1, \mathbf{D}_2, \dots$ such that for each $i \ge 1$ the pair $(\mathbf{D}_i, \mathbf{D}_{i+1})$ is a compromise transition. Then it follows from Equation 1 that $\gamma(\mathbf{D}_i) \neq \gamma(\mathbf{D}_j)$ for each $i < j$. Since each $\gamma(\mathbf{D}_i)$ is a sequence of at most n numbers, with each number taking values between 1 and n, our claim follows. □

Proof of Theorem 3. To prove Theorem 3, we need two auxiliary lemmas. In what follows, for a coalition structure $\mathbf{D}$, $|\mathbf{D}|$ denotes the number of non-empty coalitions in $\mathbf{D}$ that do not support r.

Lemma 1. *In every deliberation space, a deliberation that consists of compromise transitions and has a coalition structure $\mathbf{D}$ with $|\mathbf{D}| = 2$ as its terminal state is successful.*

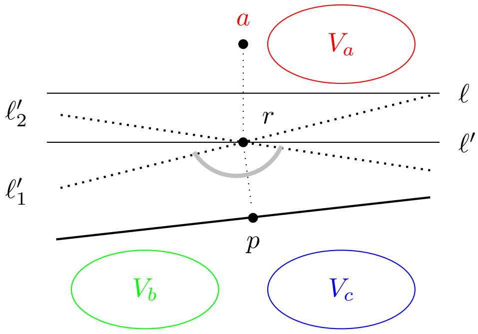

Figure 4: Proof of Lemma 2.

Proof. Consider a coalition $\mathbf{D}$ with $|\mathbf{D}| = 2$ that is not successful. Let $\mathbf{d}_1$, $\mathbf{d}_2$ be the two coalitions in $\mathbf{D}$ that do not support r; we have $|\mathbf{d}_1| < m^*$, $|\mathbf{d}_2| < m^*$. For each $p \in M^*$ we have $V^p = \mathbf{d}_1^p \cup \mathbf{d}_2^p$ and hence $|\mathbf{d}_1^p \cup \mathbf{d}_2^p| = m^* > |\mathbf{d}_1|, |\mathbf{d}_2|$. Thus, there exists a compromise transition from $\mathbf{D}$ in which agents in $\mathbf{d}_1^p \cup \mathbf{d}_2^p$ form a coalition around p. □

Lemma 2. *In every Euclidean deliberation space, a deliberation that consists of compromise transitions and has a coalition structure* $\mathbf{D}$ *with* $|\mathbf{D}| \geq 3$ *as its terminal state is successful.*

Proof. For the case $d = 1$ our claim follows from the proof of Theorem 1. We will now provide a proof for $d = 2$; it generalizes straightforwardly to $d > 2$.

Fix a coalition structure $\mathbf{D}$ that contains at least three coalitions, all of which are not successful and none of which supports r; we will show that $\mathbf{D}$ is not terminal. Let $\mathbf{d}^a = (C_a, a)$ be a maximum-size coalition in $\mathbf{D}$ that does not support r. Let ℓ be the line that passes through the middle of the a–r segment and is orthogonal to it; this line separates $\mathbb{R}^2$ into two open half-planes so that r lies in one of these half-planes while all points in $\mathbf{d}^a$ lie in the other half-plane (see Figure 4). Let ℓ' be the line that passes through r and is parallel to ℓ. For a positive α, let ℓ_1' be the line obtained by rotating ℓ' about r clockwise by α, and let ℓ_2' be the line obtained by rotating ℓ' counterclockwise by α. The line ℓ_1' (resp. ℓ_2') partitions $\mathbb{R}^2$ into open half-planes H_1 and H_1' (resp. H_2 and H_2'). We can choose α to be small enough that $\mathbf{d}^a \subset H_1, \mathbf{d}^a \subset H_2$ and so that no agent lies on ℓ_1' or on ℓ_2'.

Now, if there exists a coalition $\mathbf{d}^b = (C_b, b) \in \mathbf{D}$, $\mathbf{d}^b \neq \mathbf{d}^a$, $b \neq r$, such that $v \in H_1$ or $v \in H_2$ for some $v \in C_b$, then r is not in the convex hull of C_a and v, and hence there is a line that separates $C_a \cup \{v\}$ from r; by projecting r onto this line, we obtain a proposal r' that is approved by v and all agents in C_a. Thus, there is a compromise transition in which a non-empty subset of agents in $C_b^{r'}$ joins C_a to form a deliberative coalition around r'.

Otherwise, $\mathbf{d}^x \subseteq H_1' \cap H_2'$ for all $\mathbf{d}^x \in S \setminus \{\mathbf{d}^a\}$. Consider two distinct coalitions $\mathbf{d}^b, \mathbf{d}^c \in \mathbf{D}$ with $b, c \neq r$. As $H_1' \cap H_2'$ is bounded by two rays that start from r, and the angle between these rays is $2\pi - 2\alpha < 2\pi$, there is a line ℓ^* that divides $\mathbb{R}^2$ into two open half-planes so that r is in one half-plane and $\mathbf{d}^b \cup \mathbf{d}^c$ are in the other half-plane; thus, all agents in $\mathbf{d}^b \cup \mathbf{d}^c$ approve the proposal p obtained by projecting r onto ℓ^*, and hence there is a merge transition. □

We are now ready to finish the proof of Theorem 3. Consider a maximal run of compromise transitions, and let $\mathbf{D}$ be its terminal state. Suppose for the sake of contradiction that $\mathbf{D}$ is not successful. Note that this implies that $|\mathbf{D}| > 1$. If $|\mathbf{D}| = 2$ there is a transition from $\mathbf{D}$ by Lemma 1 and if $|\mathbf{D}| \geq 3$, then there is a transition from $\mathbf{D}$ by Lemma 2. □

B Notation and Proofs for Section 7

In what follows, given a proposal $p = (i_1, \dots, i_d) \in \{0,1\}^d$ and a coalition structure $\mathbf{D}$ we denote the deliberative coalition in $\mathbf{D}$ that supports p by $\mathbf{d}_p = (C_p, p)$; also, if an agent's ideal point is p, we will say that this agent is of *type* p. We will refer to the quantity $i_1 + \cdots + i_d$ as the *weight* of proposal p.

Proof of Proposition 5. For $d = 1$, the claim holds trivially as any deliberative coalition structure on $\{0,1\}^1$ is terminal (as there are only two elements in the space).

For $d = 2$, assume towards a contradiction that there exists a terminal deliberative coalition structure $\mathbf{D}$ that is not successful. Hence there exists a most supported alternative p, $p \neq r$, but there is no coalition in $\mathbf{D}$ supporting p with maximum support m^*. Observe that $\mathbf{D}$ cannot contain two coalitions that support the same proposal, as these two coalitions could then join forces. Hence, $\mathbf{D}$ contains at most three coalitions that do not support r. Suppose first that $\mathbf{D}$ contains a deliberative coalition $(C, 11)$. Note that all agents in C are of type 11, as agents of type 01 or 10 do not approve 11. Since $|C| < m^*$, there must exist a coalition (C', q) with $q \in \{01, 10\}$. But then there would be a follow transition in which C joins C', so $\mathbf{D}$ would not be terminal. Hence, $\mathbf{D}$ contains at most two coalitions. But then there is a compromise transition in which the members of these coalitions that support p form a coalition around p. □

Proof of Proposition 6. Fix some $d \geq 3$. We prove the claim by constructing a coalition structure in the d-hypercube that is terminal for $(d-1)$-compromise transitions, but not successful. The construction generalizes Example 7.

For each $i \in [d]$, let e_i be the point in the hypercube that has 1 in the i-th coordinate and 0 in all other coordinates. Also, for each $i \in [d] \setminus \{1\}$, let f_i be the point in the hypercube that has 1 in coordinates 1 and i, and 0 in all other coordinates.

We create one agent of type e_1, and place him into a singleton coalition that supports e_1. For each $i = 2, \dots, d$ we create $d-2$ agents of type e_i and one agent of type f_i, and place them into a coalition that supports e_i; this coalition has $d-1$ members. Let $\mathbf{D}$ be the resulting coalition structure.

Note that agents of type e_i, $i \in [d]$, do not approve any proposals other than e_i. An agent of type f_i for some $i \in [d] \setminus \{1\}$ approves f_i, e_1 and e_i. Hence, e_1 is the most-supported proposal, as it has d supporters. However, there are no ℓ-compromise transitions from $\mathbf{D}$ for $\ell < d$: if fewer than $d-1$ agents move to e_1, then the resulting coalition would have at most $d-1$ members. □

The proofs of Propositions 7 and 8 make use of the two lemmas stated and proved below.

Lemma 3. *In every deliberation space, any deliberation that consists of ℓ-compromise transitions and terminates in a coalition structure $\mathbf{D}$ with $|\mathbf{D}| = \ell$, is successful.*

Proof. Consider a coalition $\mathbf{D}$ with $|\mathbf{D}| = \ell$ that is not successful. Either $\mathbf{D}$ contains a coalition that supports r or it does not. Suppose it does not. Let $\mathbf{d}_1, \dots, \mathbf{d}_\ell$ be the d coalitions in $\mathbf{D}$. We have $|\mathbf{d}_i| < m^*$ for $1 \leq i \leq \ell$. For each $p \in M^*$ we have $V^p = \bigcup_{1 \leq i \leq \ell} \mathbf{d}_i^p$ and hence $|\bigcup_{1 \leq i \leq \ell} \mathbf{d}_i^p| = m^* > |\mathbf{d}_i|$. Thus, by Definition 10, there exists an ℓ-compromise transition from $\mathbf{D}$ in which agents in $\bigcup_{1 \leq i \leq \ell} \mathbf{d}_i^p$ form a coalition around p. If $\mathbf{D}$ contains a coalition that supports r, the same argument applies with $\ell - 1$ coalitions. □

Lemma 4. *Let $\mathbf{D}$ be a terminal coalition structure in the d-hypercube with respect to ℓ-compromises for some $d \geq 2$, $\ell \geq 2$. Then for each point p in the d-hypercube it holds that $\mathbf{D}$ contains at most one deliberative coalition that contains agents of type p.*

Proof. Suppose $\mathbf{D}$ contains two deliberative coalitions $\mathbf{d}_1$, $\mathbf{d}_2$ that contain agents of type p. We can assume without loss of generality that $|\mathbf{d}_1| \geq |\mathbf{d}_2|$. But then the agents of type p approve the proposal supported by $\mathbf{d}_1$, so the agents of type p in $\mathbf{d}_2$ have an incentive move to $\mathbf{d}_1$, a contradiction with $\mathbf{D}$ being terminal. □

Proof of Proposition 7. The claim is clearly true for $d = 2$, so we can assume $d \geq 3$. Consider a deliberative coalition structure $\mathbf{D}$ that does not admit any ℓ-transitions for $\ell = 2^{d-1} + d$. We say that a coalition $\mathbf{d} = (C, p)$ in $\mathbf{D}$ is *pure* if all agents in C have the same type. We claim that $\mathbf{D}$ can contain at most $d + 1$ pure coalitions.

Indeed, suppose that $\mathbf{D}$ contains $d + 2$ pure coalitions. By the pigeonhole principle, there are two pure coalitions $\mathbf{d} = (C, p)$ and $\mathbf{d}' = (C', p')$ in $\mathbf{D}$ such that all agents in $\mathbf{d}$ are of type $i_1 i_2 \ldots i_d$, all agents in $\mathbf{d}'$ are of type $j_1 j_2 \ldots j_d$, and there exists an $s \in [d]$ such that $i_s = j_s = 1$. Let e_s be a point in the hypercube that has 1 in position s and 0 in all other positions. Then there is a transition in which all agents in C, all agents in C' and all agents currently supporting s form a coalition that supports s; this is a contradiction, as this transition only involves 3 coalitions, and $2^{d-1} + \frac{d+1}{2} > 3$ for $d \geq 3$.

Thus, all but $d+1$ coalitions in $\mathbf{D}$ are not pure, and hence contain agents of at most two different types. Further, by Lemma 4, for each $p \in \{0, 1\}^d$ it cannot be the case that two coalitions in $\mathbf{D}$ contain agents of type p. Hence, if $\mathbf{D}$ contains t pure coalitions, it contains at most $(2^d - t)/2$ non-pure coalitions, i.e., at most $2^{d-1} + t/2$ coalitions altogether; as $t \leq d + 1$, our bound on $\ell^*(d)$ now follows from Lemma 3. □

Proof of Proposition 8. We partition the proof into four claims.

Claim 1. *$\ell^*(3) \geq 3$.*

Proof. This follows immediately from Proposition 6. □

Claim 2. *$\ell^*(3) \leq 3$.*

Proof. Let $\mathbf{D}$ be a terminal coalition structure, and suppose that $\mathbf{D}$ is not successful.

Let $\mathcal{C} = \{C_{110}, C_{101}, C_{011}, C_{111}\}$. Note that all agents in C_{110} are of type 110 or 111, and similarly for C_{101} and C_{011}. On the other hand, C_{111} may contain agents of type 111, 110, 101 or 011.

Suppose first that at least two of the coalitions in $\mathcal{C}$ are not empty. Then these two coalitions have a merge transition (which is also a 3-compromise transition) in which they form a coalition around 111, a contradiction with $\mathbf{D}$ being a terminal coalition structure. On the other hand, if all coalitions in $\mathcal{C}$ are empty, then $|\mathbf{D}| = 3$, and we obtain a contradiction by Lemma 3.

Thus, we can focus on the case where exactly one coalition in $\mathcal{C}$—say, C—is non-empty. Lemma 3 then implies that each of the coalitions C_{001}, C_{010}, and C_{100} is non-empty. As argued above, if there is an agent in C of type (i, j, k) then $i + j + k \geq 2$. Now, if C contains no agent of type 110, then all agents in C approve 001 and hence there is a follow transition where C joins C_{001}, a contradiction with $\mathbf{D}$ being a terminal coalition. Similarly, if C contains no agent of type 101, then C can join C_{010}, and if C contains no agent of type 011, then C can join C_{100}. We conclude that C contains agents of types 110, 101, and 011; hence, it has to be the case that C supports 111, i.e., $C = C_{111}$.

Now, suppose that $|C_{111}| \leq |C_{ijk}|$ for some i, j, k with $i + j + k = 1$; assume without loss of generality that $|C_{111}| \leq |C_{001}|$. We have already argued that C_{111} contains some agents of type 011; note that these agents approve 001. Hence there is a subsume transition in which all agents in C_{111} who approve 001 move to 001, a contradiction with $\mathbf{D}$ being a terminal coalition structure.

The remaining possibility is that $|C_{111}| > \max\{|C_{001}|, |C_{010}|, |C_{100}|\}$. In this case, if for some i, j, k with $i + j + k = 1$ the coalition C_{ijk} contained some agents who approved 111, then there would be a subsume transition in which all these agents would move to 111. Thus C_{111} already contains all agents who approve 111. As we assume that **D** is not successful, we have $|C_{111}| < m^*$ and hence $111 \notin M^*$. It follows that 110, 101, and 011 are not in M^* either, as every agent who approves one of these proposals also approves 111. Moreover, as C_{111} contains all agents that approve 111, each coalition C_{ijk} with $i + j + k = 1$ consists exclusively of agents of type (i, j, k). Thus, if, say, 100 is in M^*, then all agents who approve 100 are either in C_{100} or in C_{111} and hence there is a follow transition in which all these agents join forces around 100, forming a coalition of size m^*, a contradiction with **D** being terminal. □

Claim 3. *$\ell^*(4) \geq 5$.*

Proof. Consider a coalition structure $\mathbf{D} = \{\mathbf{d}_{1000}, \mathbf{d}_{0100}, \mathbf{d}_{0010}, \mathbf{d}_{0001}, \mathbf{d}_{1110}\}$, where

- C_{1000} contains 21 agents of type 1000 and 10 agents of type 1001;
- C_{0100} contains 21 agents of type 0100 and 10 agents of type 0101;
- C_{0010} contains 21 agents of type 0010 and 10 agents of type 0011;
- C_{1000} contains one agent of type 0001; and
- C_{1110} contains 30 agents of type 1101, 30 agents of type 1011, 30 agents of type 0111, 10 agents of type 1100, 10 agents of type 1010, and 10 agents of type 0110.

One can verify that each agent approves the proposal supported by her deliberative coalition.

There are 121 agents who approve 0001: this includes the one agent in C_{0001}, 10 agents from each of the coalitions C_{1000}, C_{0100} and C_{0001}, and 90 agents from C_{1110}. For these agents to gather at 0001, all five coalitions need to be involved. In particular, it cannot be the case that only the agents of types 1001, 0101 and 0011 move to 0001, as this would result in a coalition of size 31, which is equal to the size of their current coalitions.

It remains to argue that **D** is stable with respect to transitions that involve at most four coalitions. The argument in the previous paragraph already shows that every such transition should involve agents in C_{1110}.

For each $s = 1, 2, 3, 4$ we will argue that there is no transition in which agents move to a point of weight s.

- $s = 4$: It is immediate that there is no transition to the unique point of weight 4, i.e., 1111, as all agents who approve 1111 are currently in C_{1110}.
- $s = 3$: No agent not currently in C_{1110} approves 1110, so there is no transition in which some agents move to 1110. Now, consider a point of weight 3 that differs from 1110; for concreteness, take 0111. There are 100 agents in C_{1110} that approve 0111, as well as 10 agents in C_{0100} and 10 agents in C_{0011}. Thus, the coalition at 0111 would contain 120 agents, which is equal to the size of C_{1110}. Therefore, agents in C_{1110} would not benefit from this transition.
- $s = 2$: For every proposal of weight 2, there are at most 60 agents in C_{1110} who support this proposal and at most 10 other agents who support it; as $|C_{1110}| > 70$, there is no transition to a point of weight 2 that is attractive to agents in C_{1110}.

- $s = 1$: There are 80 agents in C_{1110} who approve 1000; these agents could move to 1000, but this would result in a coalition of size 111, which is smaller than their current coalition. Hence, a transition in which agents who approve 1000 move to 1000 is not possible; for a similar reason, there is no transition where agents who approve 0100 or 0010 move to the respective points.

□

Claim 4. $\ell^*(4) \leq 5$.

Proof. Let $\mathbf{D}$ be a terminal coalition structure with respect to 5-compromise transitions, and suppose for the sake of contradiction that $\mathbf{D}$ is not successful. By Lemma 3, $\mathbf{D}$ contains at least six coalitions.

Consider an agent who approves 1111. Then her ideal point has weight 3 or 4. Consequently, this agent approves all proposals of weight 3 or 4: indeed, her distance to 0000 is 3, while her distance to any proposal of weight at least 3 is at most 2. It follows that if $\mathbf{D}$ contains a coalition $\mathbf{d}_{1111} = (C_{1111}, 1111)$ then it does not contain any coalitions that support a proposal of weight 3: indeed, the ideal point of each agent in C_{1111} has weight 3 or 4, so if there is another coalition C' that supports a proposal p of weight 3, then there is a follow transition in which C_{1111} and C' merge around p.

Furthermore, $\mathbf{D}$ can contain at most two coalitions that support proposals of weight at least 3. Indeed, assume for the sake of contradiction that there are three coalitions in $\mathbf{D}$ that support a proposal of weight at least 3. As argued in the previous paragraph, it cannot be the case that one of them supports 1111, so we can assume without loss of generality that $\mathbf{D}$ contains coalitions $\mathbf{d}_{1110} = (C_{1110}, 1110)$, $\mathbf{d}_{1101} = (C_{1101}, 1101)$, and $\mathbf{d}_{1011} = (C_{1011}, 1011)$, with $|C_{1110}| \geq |C_{1101}| \geq |C_{1011}|$. Then C_{1011} can only contain agents of type 0011: all other agents who approve 1011 also approve either 1110 or 1101, so they would have an incentive to move to the weakly larger coalitions that support these proposals. Similarly, C_{1101} can only contain agents of type 1001 or 0101, as all other agents who approve 1101 also approve 1110, so they can move to the weakly larger coalition C_{1110}. But then all agents in C_{1101} and C_{1011} approve 0001, so there is a 3-compromise in which all agents in C_{1101}, C_{1011} and C_{0001} merge around 0001, a contradiction.

Thus, we have argued that $\mathbf{D}$ contains at most two coalitions that support proposals of weight 3 or higher. Now, consider coalitions in $\mathbf{D}$ that support proposals of weight at most 2.

Suppose $\mathbf{D}$ contains a coalition $\mathbf{d}_{1100} = (C_{1100}, 1100)$. Note that all agents in C_{1100} approve 1000 and 0100. Thus, if $\mathbf{D}$ contains a coalition that supports 1000 or 0100, then there would be a follow transition in which all agents in C_{1100} adopt the proposal of this coalition, a contradiction with $\mathbf{D}$ being a terminal coalition structure. Similarly, if $\mathbf{D}$ contains a coalition $\mathbf{d}_{1010}$ or $\mathbf{d}_{1001}$, there would be a merge transition in which C_{1100} merges with this coalition around 1000, and if $\mathbf{D}$ contains a coalition $\mathbf{d}_{0110}$ or $\mathbf{d}_{0101}$, there would be a merge transition in which C_{1100} merges with this coalition around 0100.

We conclude that if $\mathbf{D}$ contains $\mathbf{d}_{1100}$ then it contains at most two other coalitions that support proposals of weight at most 2: namely, it may contain $\mathbf{d}_{0011}$ (in which case it contains no coalitions that support a proposal of weight 1), or it may contain one or both of the coalitions $\mathbf{d}_{0010}$ and $\mathbf{d}_{0001}$. By symmetry, it follows that if $\mathbf{D}$ contains a coalition that supports a proposal of weight 2, then it contains at most three coalitions that support proposals of weight at most 2. As we have argued that $\mathbf{D}$ contains at most two proposals of weight at least 3, we obtain a contradiction with $|\mathbf{D}| \geq 6$. Hence, we may assume that $\mathbf{D}$ contains no coalitions that support a proposal of weight exactly 2.

Thus, if $|\mathbf{D}| \geq 6$, it must be the case that $\mathbf{D}$ contains four coalitions supporting proposals of weight 1, as well as two coalitions supporting proposals of weight at least 3. We have argued that among these two coalitions, one (a weakly smaller one) only contains agents of two types of weight 2 each; moreover, our analysis shows that all agents in this coalition approve some proposal p of weight 1. But then there is a follow transition in which this coalition joins the coalition that currently supports p, a contradiction again. Thus, we can conclude that $\mathbf{D}$ is successful. □

Proposition 8 now follows by combining Claims 1–4. □

Edith Elkind
University of Oxford
Oxford, United Kingdom
Email: eelkind@gmail.com

Davide Grossi
University of Groningen and University of Amsterdam
Groningen and Amsterdam, the Netherlands
Email: d.grossi@rug.nl

Ehud Shapiro
Weizmann Institute of Science
Rehovot, Israel
Email: ehud.shaprio@weizmann.ac.il

Nimrod Talmon
Ben-Gurion University
Be'er Sheva, Israel
Email: talmonn@bgu.ac.il