

Mixing Dyadic and Deliberative Opinion Dynamics in an Agent-Based Model of Group Decision-Making

George Butler, Gabriella Pigozzi, Juliette Rouchier

▶ To cite this version:

George Butler, Gabriella Pigozzi, Juliette Rouchier. Mixing Dyadic and Deliberative Opinion Dynamics in an Agent-Based Model of Group Decision-Making. Complexity, 2019, 2019, 10.1155/2019/3758159. hal-02308172

HAL Id: hal-02308172 https://hal.science/hal-02308172

Submitted on 8 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Research Article **Mixing Dyadic and Deliberative Opinion Dynamics in an Agent-Based Model of Group Decision-Making**

George Butler, Gabriella Pigozzi, and Juliette Rouchier 🝺

Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243, LAMSADE, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France

Correspondence should be addressed to Juliette Rouchier; juliette.rouchier@dauphine.fr

Received 15 December 2018; Revised 6 May 2019; Accepted 10 June 2019; Published 7 August 2019

Academic Editor: José Manuel Galán

Copyright © 2019 George Butler et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this article, we propose an agent-based model of opinion diffusion and voting where influence among individuals and deliberation in a group are mixed. The model is inspired from social modeling, as it describes an iterative process of collective decision-making that repeats a series of interindividual influences and collective deliberation steps, and studies the evolution of opinions and decisions in a group. It also aims at founding a comprehensive model to describe collective decision-making as a combination of two different paradigms: argumentation theory and ABM-influence models, which are not obvious to combine as a formal link between them is required. In our model, we find that deliberation, through the exchange of arguments, reduces the variance of opinions and the proportion of extremists in a population as long as not too much deliberation takes place in the decision processes. Additionally, if we define the correct collective decisions in the system in terms of the arguments that should be accepted, allowing for more deliberation favors convergence towards the correct decisions.

1. Introduction

In a group, opinions are formed over affinities and conflicts among the individuals that compose it. Axelrod [1], a pioneer in opinion dynamics, cast light on two key factors required to model the processes of diffusion, namely, social influence (i.e., individuals become more similar when they interact) and homophily (i.e., individuals interact preferentially with similar others). He was the first to show that a radical differentiation of culture in a group could emerge from simple imitation through dyadic interactions. His results suggested that interactions through homophily and social influence could lead to collective states or collective opinions whose explanation, in many situations, went beyond the individual or micro level. Further, it was found that these collective states could be characterized by quantities like statistical distributions and averages, which explains why, in recent years, a growing body of research has endeavored to identify the conditions under which social influence at the micro (dvadic)-level translates into macropatterns of diffusion through repeated iterations [2]. In particular, several models have been developed to reproduce the emergent properties of opinion diffusion which may be classified in two groups: on the one hand, the discrete opinion models where opinions, or other ontological equivalents, take discrete values [3, 4]; on the other, the continuous opinion models where opinions are represented by real numbers [5–10].

In the context of continuous opinion dynamics introduced by Deffuant et al. [6], and later extended by other authors to include network effects [10, 11], trust [12], and many other social phenomena [7], individuals meet in random pair-wise encounters and then converge to a common opinion if and only if their respective opinions are sufficiently close to each other, in a kind of bounded confidence mechanism based on *confirmation bias*. After some transient evolution and social dynamics, this leads to final states in which either full consensus is reached or the population splits into a finite number of clusters such that all individuals in one cluster share the same opinion. So far, these models have been mostly applied to political issues such as societal cleavages and the emergence of extremism.

However, these representations of social interactions in opinion dynamics fail to take into account everyday communication settings that characterize democracies such as meetings, debate arenas, and the Media in which individuals exchange points of view and can influence one another in a collective manner. In effect, individual actions that translate into collective decisions, such as voting, are shaped by factors related to the structure and size of the channels of communication and deliberation. When a group engages in a collective discussion, group size, what arguments are advanced, how discussion is organized over time, and the acceptability criteria for proposals may lead to a transformation of preferences [13] and play a crucial role in consensus formation [14-16]. For instance, work in [14] defends that deliberation polarizes individuals in the direction of their initial opinions due to social pressure and to limited knowledge (biased or unbalanced argument pool) within the discussing group. In contrast, from several empirical studies of deliberation processes, other authors infer that deliberation can have a stabilizing or moderating effect on opinions [16, 17], which they interpret as opinions becoming more informed [18], balanced, and/or confused [13]. Authors in [15] express that deliberation may encourage moderate opinion consensuses if it is procedural, or be polarizing if it ego-involves the participants. All these phenomena may introduce some degree of discrepancy into the otherwise well-known steady states of opinions (clusters) obtained in classical opinion dynamics and should be modeled somehow.

Opinion diffusion has also been used to track convergence towards "correct" or accurate opinions in groups. Although authors in [19] study how interactions among agents diffuse true information, they focus their analyses on network effects and noisy signaling, and not on deliberation protocols. In [20], Rouchier and Tanimura explore the diffusion of information about an exogenous "true state" of the world (represented by bits) through social interactions and learning but assume the interactions to be only dyadic and not deliberative. In our context, a correct decision corresponds to one derived from a dialectical situation in which all the arguments for and against a proposed alternative are taken into account [15]. The existence of an ideal criteria of correctness of collective decisions can be used to evaluate the outcome of deliberative procedures and democratic decisionmaking. Since deliberation imposes regulatory conditions to decision-making processes, one may attribute a rational and democratic value to the collective decisions obtained from it. Deliberation is a way of getting closer to the ideal state in which a group judges propositions as if it had all arguments at its disposal. In [18], Barabas shows empirical evidence that deliberation increases knowledge and is correlated to correct responses to objective questions. Up until now, opinion diffusion models have not taken into account this dimension and models mixing deliberative and dyadic interactions may well do so. In the same manner, collective truth-seeking models can benefit from more insight on processes of peer-topeer information diffusion of the like of continuous opinion dynamics models.

From a social welfare perspective, a claim in favor of deliberation is that promoting dialogue leads to better decision-making, where "better" goes as improving social welfare. Preference structures that are rationally untenable or unjustifiable are eliminated from the pool of admissible preferences outright [21]. We argue that different ways of partaking deliberation, e.g. the structure and protocol of decision-making processes, may allow for more accurate collective decision-making.

In this direction, agent-based modeling proves to be an interesting method to study and observe the effect of deliberation on opinions. For one, it helps infer knowledge from models in which a multiplicity of different modes of communication among heterogeneous agents are, analytically, difficult to describe; second, given that empiricalbased conclusions on the topic may be costly and difficult to ascertain due to confronting ideologies and theories on the topic (see [15, 16]), it provides an alternative way of testing the effects of collective decisions and deliberation protocols on public opinion; and, not an exhaustive third, it furnishes an interesting modeling environment for collective choice analysis by giving the possibility to account for different levels of decision-making and a diversity of influence loops.

The aim of our model is to breach a gap between deliberation and opinion diffusion. We model dyadic or ego-involved dynamics using an opinion diffusion model based on social judgment theory [9, 22], whereas formal or deliberative discussion is modeled using abstract argumentation theory [23–25]. We build a process of collective decision-making where deliberation and voting are necessary conditions for collective choice as we are inspired from results in the deliberative democracy [16, 17, 26] and social psychology [14, 15] literature. We present the effects of deliberation on the opinions of a group and on its ability to correctly judge propositions. We call the latter the group's *judgment accuracy*. We also observe how deliberation impacts a group's ability to eventually vote in favor of proposals that are discussed and accepted during deliberation, in other words, on the group's coherence in decision-making. Furthermore, since a collective decision in the model is an outcome of a structured group decision-making process, a sequence of deliberative and dyadic interactions among agents, we seek to study the impact of its structure on the group's opinions, judgment accuracy, and coherence. The frequency of deliberative interactions within a decision-making process, the size of deliberation (number of agents that deliberate), and the majority voting rules used determine the collective acceptance of proposals are considered to this end. Ultimately, we strive to create a tool that helps policy-makers analyze deliberation protocols and make studied decisions about them.

Our model allows us to explore a new paradigm in opinion diffusion and answer the following research questions concerning decision-making in groups:

- (1) What effect can deliberation have on a group's opinion distribution when the opinion dynamics are described using bounded confidence models? In what way are opinion dynamics through deliberation alone interesting?
- (2) How do the structure and protocol of deliberative decision-making processes affect a group's distribution of opinions, coherence, and judgment accuracy?

(3) In what way can controlling for protocol relate to social situations in which collective decisions are coherent and accurate?

Simulations show that deliberation yields, on average, qualitative loose consensus and group polarization while reducing the number of different opinion clusters over the distribution of opinions. In effect, our model shows that deliberation has a significant overall impact on the distribution of opinions (on its variance) and on the shifts in individual opinion. In particular, when specifying opinion dynamics as only deliberative, the proportion of extremists and the variance of opinions are lower and the shifts of opinions greater than in a nondeliberative, dyadic opinion dynamics model. However, when considering a mix of deliberation and dyadic interactions, parameters that promote bigger or more frequent deliberation during decision-making processes increase the variance of opinions and the proportion of extremists and limit the effect of deliberation on opinions. The majority voting quota rule to accept a proposal plays a preponderant role as it determines if the consensus-driving power of deliberation outweighs the propensity to dissensus observed in bounded-confidence models with rejection.

The model sheds light on the fact that a group's coherence and judgment accuracy may depend significantly on how decision processes are structured. The number of debates allowed and the number of agents that may participate in them increase judgment accuracy in a marginally decreasing fashion but have little or no effect on the group's coherence in decision-making. Last, we point out that results are conditioned on how many arguments agents have at their disposal and on how they advance them during deliberation.

The remainder of this paper is organized as follows: in Section 2, we present the model and provide some necessary basics to understand its implementation; in the subsequent section, we introduce the metrics of interest and the calibration of the model. In Section 4, we report and discuss the results of the simulations; and in Sections 5 and 6, respectively, we survey related work and we conclude the article.

2. A Model for Collective Decision-Making with Deliberation

We propose a system of collective decision-making among agents made of three objects: arguments, agents, and tables for deliberation. Agents have agency, arguments represent pieces of information, and tables validate collective decisions while organizing all deliberative interactions among agents.

2.1. Model Overview. In this section, we propose an overview of the model. We quickly introduce the agents and the objects that make collective decisions possible in the multiagent system. We also provide a basic interpretation of the collective decision-making procedure, which we illustrate with an example.

2.1.1. Overview of Agents and Objects in the Model. Arguments are objects that relate to each other by a defeat relation.

They are characterized by their support of some value or principle. Agents are characterized by their sensitivity to deliberated ideas, by the arguments they possess, which reflect their opinions, and their knowledge about arguments. They communicate one-to-one or collectively by the dint of arguments in a public arena. Communication leads them to an eventual update of their opinions. Tables are entities that contain both: agents and arguments. They are controlled by a central authority (CA) [27] that fixes the rules in the deliberation process, the frequency of deliberative interactions, and the conditions for collective acceptability of deliberated information. The existence of a central authority, namely, a person or a machine that can reveal the correct epistemic status of any set of arguments can be equated to Habermas' claim that the "unforced force of the better argument" will triumph in an ideal speech situation¹. A central authority may also be associated with Rancière's notion of "police" that he defines as "an order of bodies that defines the allocation of ways of doing, ways of being, and ways of saying, and that sees that those bodies are assigned by name to a particular place and task" [28] (p.29).

2.1.2. Overview of the Decision-Making Process. Let N be a group of agents that is asked to deliberate and vote on a proposal \mathcal{P} . Given an argument I in favor of \mathcal{P} , they have to judge on whether \mathcal{P} is desirable or not. The argument I or proposal argument determines how much the proposal supports a principle \mathbb{P} or its opposite $\neg \mathbb{P}$. A proposal is a sentence that indicates a way of attaining a goal or solving a problem. A principle derives from the notion of value—values seen as fundamental social or personal goods that are desirable in themselves [29]. Environmentalism and patriotism are examples of values; to choose proposals that minimize environmental impact or maximize welfare are examples of two not mutually exclusive principles.

Agents are assumed to decide on the acceptance of the proposal on the basis of their *opinions* or their *adherence* to the said principle, whether they argue formally or informally about \mathcal{P} . When agents are not deliberating, they are subject to random pair-wise influence; when the deliberative exchange of arguments concludes, they are influenced by the acceptance status of the proposal that is being discussed. Only a fraction of all agents deliberate in each deliberative exchange; all agents are prone to pair-wise discussion. They vote for proposals according to their opinions. A proposal \mathcal{P} is accepted if and only if the argument *I* that is given with it is accepted after deliberation and the proposal is voted favorably by a majority of agents. To vote in favor of a proposal is comsidered to be equivalent to accepting the argument that comes along with it.

Example 1. \mathbb{P} = "Protect the environment" is a principle; a proposal may be \mathscr{P} = "Reduce carbon emissions by 2030 using electric cars." To justify the proposal, one may advance the argument I = "Electric cars will reduce society's dependency on fossil fuels and will result in a reduction of carbon emissions for 2030 while proctecting the environment" that expresses the degree to which the proposal is in line with environmental protection. An argument that tackles the



FIGURE 1: Argumentation framework $AF = (\mathcal{A}, \mathcal{R})$ with $\mathcal{A} = \{a, b, c\}$ and $\mathcal{R} = \{(a, b), (b, c)\}$. The labeling $\{\{b\}, \{c\}, \{a\}\}$ is conflict-free; $\{\{a, c\}, \{b\}, \emptyset\}$ is the only complete labeling.

proposal argument *I* and that opposes the principle \mathbb{P} could be a = "Electric cars may protect the environment by reducing effective carbon emissions but the batteries they depend on are very pollutant". The proposal \mathscr{P} may be accepted in the deliberation arena if some other argument that defeats *a*, say b = "Battery recycling businesses are hatching everywhere. By 2020, it is very likely that scientists find a viable solution to chemical pollution due to batteries" is advanced. It is not accepted otherwise. If *I* is accepted, then \mathscr{P} is collectively accepted if a majority of agents vote favorably for it.

2.2. Arguments as the Basic Units of Collective Discussion among Agents. In this subsection, we introduce the simplest object in our system: arguments. We present abstract argumentation theory and Caminada's labeling approach to argumentation [30] that allows us to track the epistemic status of arguments during deliberation.

2.2.1. Arguments. Arguments are objects that represent pieces of information that agents can understand. They are informational cues that enable agents to discuss with one another in a public, collective context and make decisions on the acceptability of other pieces of information. They are assumed to be nonfallacious. In our approach, each argument a is modeled by a real number $v_a \in [-1, 1]$ that stands for how much a respects or supports the principle \mathbb{P} . $v_a = 1$ means that argument *a* is totally coherent with the principle \mathbb{P} , whereas $v_a = -1$ reads that *a* is totally incoherent with the principle \mathbb{P} . Arguments relate to each other through an incompatibility relation that states that one cannot stand behind two conflicting arguments. Arguments are also characterized by their acceptability status that indicates whether they are accepted, refuted, or undecided in a given discursive context. Arguments have an epistemic reach (er) or a maximum number of arguments they can attack. This can be interpreted as the argument's level of generality or as the potential level of argumentative conflict within the argument pool.

2.2.2. Abstract Argumentation Theory. Deliberation, defined as an exchange of arguments, is modeled by confronting representations of different, eventually contending, arguments. In our model, we use abstract argumentation theory [23] to represent deliberation, where an argument is just a node in a graph, like arguments a, b, and c in Figure 1. Abstract argumentation theory models incompatibility between arguments and abstracts away from their internal structure. The intuition is that if an argument a attacks an argument b (represented by an arrow from node a to node b, as in Figure 1), a rational agent cannot accept both a and b.

Formally, let \mathscr{A} be a finite set of arguments and \mathscr{R} a subset of $\mathscr{A} \times \mathscr{A}$ called *attack relation*. The attack relation is intransitive and we note $(a, b) \in \mathscr{R}$ the fact that argument *a attacks* or is incompatible with argument *b*. One says that an argument *a defends* an argument *c*, noted $(a, c) \in \mathscr{D}$, if there exists an argument *b* such that $(a, b) \in \mathscr{R}$ and $(b, c) \in \mathscr{R}$ (see Figure 1). An argumentation framework $(AF = (\mathscr{A}, \mathscr{R}))$) is a digraph in which the nodes represent the arguments and the arcs represent the attacks among them. Given an argumentation framework AF, the classic problem in abstract argumentation resides in finding which of the arguments in AF can be accepted, rejected, or left undecided.

In the labeling approach [30], a label $\mathcal{L}ab(a) \in \Lambda = \{in, out, und\}$ of an argument $a \in \mathcal{A}$ denotes the epistemic status of *a*. Intuitively, an argument is labeled *in* if it is justifiable and *out* if it is not. If *a* is labeled *und*, it is considered to be in abeyance due to, for example, insufficient grounds for it to be labeled *in* or *out*. Furthermore, given an argumentation framework $AF = (\mathcal{A}, \mathcal{R})$, one calls *labeling* a complete function

$$\mathcal{L} : \mathcal{A} \longrightarrow \Lambda = \{in, out, und\}$$
$$a \longmapsto \mathcal{L}ab(a)$$
(1)

that assigns a label to each argument in *AF*. A labeling is written as a triplet $\bigotimes_{\lambda \in \Lambda} \lambda \mathscr{L}(\mathscr{A})$ where $\lambda \mathscr{L}(\mathscr{A}) = \{a \in \mathscr{A} \mid \mathscr{L}ab(a) = \lambda\}$. For example, *in* $\mathscr{L}(\mathscr{A})$ stands for the set of arguments in \mathscr{A} that are labeled *in* in the labeling $\mathscr{L}(\mathscr{A})$. A labeling is said to be *legal* if all argument it labels is legally labeled. An argument $a \in \mathscr{A}$ is said to be legally labeled

- (i) Lab(a) = in, if ∀b ∈ A s.t (b, a) ∈ R, Lab(b) = out:
 if a is not attacked or is only attacked by arguments that are themselves labeled out;
- (ii) Lab(a) = out, if ∃b ∈ A s.t (b, a) ∈ R, Lab(b) = in, that is if a is attacked by at least one argument that is labeled in;
- (iii) Lab(a) = und, if ¬(∀b ∈ A s.t (b, a) ∈ R, Lab(b) = out) ∧ (∀c ∈ A s.t (c, a) ∈ R, Lab(c) ≠ in); or equivalently, if there exists at least one argument labeled und that attacks a and there is no argument labeled in attacking a.

Roughly speaking, a *semantics* (denoted by σ) is a rationality criteria to decide which arguments to accept given an argumentation framework. The basic normative requirements for labeling-based semantics and argument acceptability in abstract argumentation repose on *conflict-free* labelings or labelings in which no two *in*-labeled arguments attack each other, and *admissible* labelings, which are conflict-free labelings that ask for arguments that are defended by the *in*-labeled arguments to be themselves *in*-labeled. The family of admissibility-based labelings goes from *complete* labelings to *preferred* and *grounded* labelings which are complete labelings that capture properties such as credulity and skepticism in argumentation. Formally, a labeling \mathcal{L} for the arguments in an argumentation framework (\mathcal{A}, \mathcal{R}) is said to be

- (i) conflict-free if $\nexists a, b \in \mathscr{A}$ s.t $(a, b) \in \mathscr{R} \land a, b \in in \mathscr{L}(\mathscr{A});$
- (ii) admissible if conflict-free and $\forall a \in \mathcal{A}, \mathcal{L}ab(a) \neq$ und $\Longrightarrow a$ is legally $\mathcal{L}ab(a)$;
- (iii) complete if admissible and $\forall a \in \mathcal{A}, \mathcal{L}ab(a) =$ und $\Longrightarrow a$ is legally labeled und;
- (iv) preferred if complete and if it maximizes *in* $\mathcal{L}(\mathcal{A})$;
- (v) grounded if complete and if it maximizes **und** $\mathcal{L}(\mathcal{A})$.

We choose admissibility-based semantics because, for one, they supply a comprehensive model of collective reasoning; they allow for a meaningful parameterization of credulous and skeptical collective reasoning, and, last, they are relatively easy to interpret since the differences between them are well documented in the literature (e.g. [18]). Additionally and in contrast to other rank-based or graded semantics in abstract argumentation, admissibility semantics assume that all arguments have the same weight.

2.2.3. Abstract Argumentation and Discursive Situations in Collective Decision-Making. Abstract argumentation is a convenient formalism for argument-based decision-making since it ignores difficulties relative to the nature, generation, and number of arguments and posits the possibility of using graph theoretic tools to model (collective) reasoning in a clear, coherent, and easy way [31]. Example 2 below provides a simple overview of a debate between two agents over the acceptability of the proposal "Tax the rich."

Example 2. Let $\mathscr{P} =$ "Tax the rich" and $\mathbb{P} =$ "*Liberalism.*" Figure 1 represents the abstract argumentation framework obtained from the following arguments:

- (i) I = c = "Only the rich should be taxed because they possess most of the capital in the country";
- (ii) b = "If you only tax the rich, the rich will leave and then you'll have no one to tax";
- (iii) a = "The rich will not leave because they have their livelihoods here and it would cost them more to leave than to pay the taxes."

Agent 1 advances the proposal argument *c*, agent 2 argues that, given the justification *c* for \mathcal{P} , \mathcal{P} is not tenable. Agent 1 defends his proposal by advancing *a* and defeating the argument *b*. The conclusion is that holding *c* and *a* is a tenable position, so the rich should be taxed. Notice that by simply accepting *c* and taking no position on *b* nor *a* is conflict-free and still corresponds to the position advocating that the rich should be taxed.

Abstract argumentation comes as an immediate application to our model in the construction of an ideal argumentation framework that loosely models Habermas' ideal speech situation. The ideal speech situation is important in our work because it corresponds to a normative state that is, in practice, difficult to reach and that allows us to observe how different deliberation protocols affect deliberative outcomes.

In the model, a label given to an argument, provided a semantics σ , is said to be *ideal* if it is obtained from a state of affairs in which all arguments are presented during deliberation. We call ideal or consensual argumentation framework the argumentation framework $(\mathcal{A}, \mathcal{R}) = AF_{\varepsilon}$ containing all arguments in the system and all consensual attacks among them. In a multiagent approach, a consensual attack between two arguments $a, b \in \mathcal{A}$ is a couple (a, b)such that $(a,b) \in \mathcal{R}$ if and only if a certain majority of agents recognizes that such attack exists. In our model, all agents agree on the attack relation over AF_{s} . It follows that if two agents advance two distinct arguments $a, b \in \mathcal{A}$ such that $(a, b) \in \mathcal{R}$, then all agents will recognize that such conflict exists. An immediate consequence of this assumption is that all deliberated results are consensual, even if an opinion *consensus*² on the principle is not necessarily reached.

2.3. Deliberative Social Agents. In this section we present the agents in our model. We define them on the basis of their opinions on the principle \mathbb{P} , the arguments they possess, their knowledge of the relationship among them, their behavior during deliberation, and their sensitivity to deliberated proposals.

2.3.1. Agents in Dyadic Social Interactions. Every agent i has an opinion, a relative position or degree of adherence $o_i \in$ [-1, 1] to a principle \mathbb{P} , and a couple $(U_i, T_i) \in [0, 2] \times [0, 2]$ $(U_i \leq T_i)$ of latitudes of acceptance and rejection, respectively, of informational cues. They live in [0, 2] since 0 and 2 are, respectively, the minimum and maximum distances of any two informational cues in the system. The idea behind the couple (U_i, T_i) is that there exist levels of relative tolerance from which informational cues have either an attractive or a repulsive effect on the individual [22, 32]. An o_i close to 1 implies that agent *i* fully supports the principle \mathbb{P} , and close to -1 that she rejects principle P or, equivalently, fully supports $\neg \mathbb{P}$. Moreover, if an agent *i*'s opinion is such that $|o_i| \ge 0.75$, *i* is considered to have an extreme opinion, a moderate one otherwise. U_i may be considered as i's uncertainty about her own opinion [33] or as the limit below which the object she judges may attract her. T_i may be seen as *i*'s bound of tolerance from which informational cues disgust her and confirm her initial position. Different combinations of (U_i, T_i) can be associated with agent *i*'s ego or personal involvement in discussion processes, as it is vividly explained in [32]. Finally, agents are assumed to be sincere and precise when communicating their positions to one another (no noise in the interactions).

2.3.2. Agents in Deliberative Social Interactions. Agents vote and participate in deliberation because they are aware of the potential changes an accepted proposal may induce in the opinion of the group. After all, proposals promote a principle that potentially leads to a shift in other agents' opinions. Agents' incentive to participate in deliberation is based on the idea that every single one of them wants to make her point across and, at the same time, reach a correct collective decision. Agents are endowed with a probability,



FIGURE 2: Representation of agents and their arguments on the opinion or adherence to the principles spectrum. The rectangles on the agents are their argument sacks. Arguments are depicted by dots and arrows from the arguments to the sacks depict both the arguments they hold in the opinion spectrum and the number of arguments they hold: $k_i = 4$ and $\phi_i^e = 0.75$, and $\phi_i^m = 0.5$ for all agent *i*.

 $x_i = f(d(o_i, v_I), p_a)$, of being attracted to a deliberated cue and a probability of being repulsed by it, $y_i = g(d(o_i, v_I), p_r)$. The two probabilities are assumed to be independent. They are a function of the distance between the opinion and the proposal argument and of the group's sensitivity³ to deliberation (p_a, p_r) . The former is decreasing on $d(o_i, v_I)$ and increasing on p_a , while the latter is increasing on both $d(o_i, v_I)$ and p_r . Agents are also assumed

- (ii) to trust⁴ one another when they utter informational cues.

2.3.3. Agents and Voting. At the end of the decision-making process, agents vote on whether they agree or not on the proposal argument that has been discussed during deliberation. Voting, for an agent, is the expression of her opinion in the final phase of the decision-making process. An agent *i* is said to vote favorably for a proposal \mathscr{P} of justification argument *I* if and only if $v_I \times o_i \ge 0$ or, equivalently, if the proposal argument does not adhere to the opposite of the principle *i* supports⁵.

2.3.4. Agent Knowledge of Arguments. Let \mathscr{A} be a finite set of arguments. Each agent *i* has a sack of arguments $\mathscr{A}_i \subset \mathscr{A}$ of size $|\mathscr{A}_i| = k_i$ whose content reflects her relative position, o_i , on \mathbb{P} (see Figure 2). For extreme-opinion agents, a proportion $\phi^e \ge 0.5$ of the *k* arguments in their sacks are of the same adherence to the principle than their opinions,

whereas for moderate agents a proportion $\phi^m \leq \phi^e$ of their *k* arguments are of the valence of their opinions. The hypothesis feeds from results found in [22] stating that "an individual places a verbal statement on an issue both in terms of the item's relative proximity to his own position and the latitude which is acceptable to him around that focal point of acceptance."

The arguments in an agent i's sack are those that she knows how to use and advance in a deliberative interaction. The size of the sack represents agent i's ability to communicate in a deliberative context. It follows that each agent *i* possesses partial knowledge $(K_i \subset \mathscr{A} \times \mathscr{A})$ of the attack relation between the arguments in \mathscr{A} and of AF_{ε} , which she derives from the arguments in her sack and what she observes in deliberation. Knowledge of arguments is assumed to be "attack-oriented"—an agent *i* knows an attack $(a,b) \in \mathscr{A}_i \times \mathscr{A}$ if she knows, upon observation, that neither she nor the group can rationally accept a alongside *b*. She may be aware of the existing defense relation among arguments when concatenating the information she accumulates on their attack relation during deliberation. She may use this information strategically in deliberative contexts.

Let $dist : \mathscr{A} \times \mathscr{A} \longrightarrow \mathbb{N}$ be the length of the shortest path between two elements in the argumentation framework induced by an agent *i*'s knowledge K_i . *i* sees an argument *a* as an *attacker* (*defender*) of an argument b if $dist(a, b) \equiv 1 \mod 2$ $(dist(a, b) \equiv 0 \mod 2)$. An argument a that is at an odd distance from another argument b attacks it since either it directly attacks the argument or it attacks a defender of the argument. Likewise, if a is at a pair distance of b, it attacks an attacker of b and thus defends it. If such path does not exist, then the distance is undefined and the agent does not see either argument as an attacker or defender of the other. Further, let $\mathscr{AF}^d = (\mathscr{A}^d, \mathscr{R}^d)$ denote the current state of affairs in deliberation. An agent *i*'s knowledge of \mathcal{R} is the set of attacks and defenses she can infer from the arguments she knows (Att_i, Def_i) and the attacks and defenses she can infer from deliberation (Att_i^d, Def_i^d) :

$$K_i = Att_i \cup Att_i^d \cup Def_i \cup Def_i^d \tag{2}$$

where,

- (i) $Att_i = \{(a, b) \in \mathcal{A}_i \times \mathcal{A}_i \mid dist(a, b) \equiv 1 \mod 2\};\$
- (ii) $Def_i = \{(a, b) \in \mathcal{A}_i \times \mathcal{A}_i \mid dist(a, b) \equiv 0 \mod 2\};\$
- (iii) $Att_i^d = \{(a,b) \in (\mathcal{A}_i \cup \mathcal{A}^d) \times \mathcal{A}^d \mid dist(a,b) \equiv 1 \mod 2\};$
- (iv) $Def_i^d = \{(a,b) \in (\mathcal{A}_i \cup \mathcal{A}^d) \times \mathcal{A}^d \mid dist(a,b) \equiv 0 \mod 2\}.$

Notice that agents have no restriction on the amount of information they can carry and use during deliberation. The model implies that agents have the ability to use and process all information on the attack relations if they could observe the attacks and that they all have equally high cognitive capacity. Agent knowledge resets at the end of each decisionmaking process, but argument sacks stay untouched. In other words, argument sacks are static in the model. 2.3.5. Agent Behavior during Deliberation. Agents may behave in two different ways in deliberation. They may behave *naively* or *focusedly*. Naive agents use deliberation to voice their opinions on the principle through arguments. Focused agents strategically argue in favor of proposal arguments that support the principle they favor. In both cases, agents advance arguments in terms of the arguments' relative proximity to their own positions [22].

Let $a_i \in \mathcal{A}_i$ denote the argument an agent advances in a debate d (of a deliberation process D_I) over a central argument I and v_a the valence of an argument $a \in \mathcal{A}$. Let

- (i) f_i(a) ≡ (o_i × v_a > 0 ∨ o_i = v_a = 0), the formula that is satisfied if the argument *a* is of the same valence as the opinion of agent *i*;
- (ii) $g_i(a) \equiv (a, b \in \mathcal{A}_i \setminus \mathcal{A}^d \land \forall b(|o_i v_b| > |o_i v_a|))$, the formula that is satisfied only if the argument *a* is in *i*'s sack (\mathcal{A}_i) and not in the debate and is the closest, in terms of adherence to the principle \mathbb{P} , to *i*'s opinion;
- (iii) $h_i(a) \equiv (f_i(a) \land g_i(a))$ the conjunction of the preceding formuli indicating that an argument *a* is in agent *i*'s sack and not in the debate, of the same sign as o_i , and closest to o_i in absolute value.

Naive agents choose $a_i = a$ such that $h_i(a)$ is true given any proposal argument *I*. Agents that behave strategically, or to say, focusedly, choose their moves in the debate as follows (*NIL* indicates no move or null argument):

(a) if agent *i*'s position with respect to \mathbb{P} is of the same sign as the proposal argument's adherence to $\mathbb{P}(f_i(I))$, she advances an argument *a* such that

(i)
$$h_i(a) \land ((a, I) \in Def_i^d))$$
 is true. If $a_i = NIL$,

(i1) $\neg h_i(a) \land g_i(a) \land ((a, I) \in Def_i^d)$ (chooses an argument not of the same sign and closest to her opinion and defending *I*) is true. If $a_i = NIL$,

(ii) $h_i(a) \land ((\exists c \in \mathscr{A}_i)((c, I) \in Att_i^d \land (a, c) \in Att_i))$ (anticipates an attack from an argument *c* on *I* and advances an argument that attacks the argument c)⁶ is true. If $a_i = NIL$,

(iii) $\neg h_i(a) \land g_i(a) \land ((\exists c \in \mathscr{A}_i)((c, I) \in Att_i^d \land (a, c) \in Att_i))$ is true. If $a_i = NIL$,

(iii) $h_i(a) \land ((a, I) \notin Att_i^d)$ (simply does not attack *I*) is true. If $a_i = NIL$,

(iiil) $\neg h_i(a) \land g_i(a) \land ((a, I) \notin Att_i^d)$ is true. If $a_i = NIL$, (iv) $a_i = NIL$ (the agent will advance no argument).

(b) if agent *i*'s position with respect to \mathbb{P} is of the opposite sign of the proposal argument's adherence to $\mathbb{P}(\neg f_i(I))$, she plays the argument *a* such that

(i) $h_i(a) \wedge ((a, I) \in Att_i^d)^7$. If $a_i = NIL$,

(i1) $\neg h_i(a) \land g_i(a) \land ((a, I) \in Att_i^d)$ is true. If $a_i = NIL$,

(ii) $h_i(a) \land ((\exists b \in \mathcal{A}_i)((b, I) \in Def_i^d \land (a, b) \in Att_i))$ (anticipates a defense from an argument *c* to *I* and advances an argument that attacks *c*) is true. If $a_i = NIL$,

(iii)
$$\neg h_i(a) \land g_i(a) \land ((\exists b \in \mathcal{A}_i))((b, I) \in Def_i^d \land (a, b) \in Att_i))$$
 is true. If $a_i = NIL$,

(iii) $h_i(a) \land ((\forall b \in \mathscr{A}^d)((b, I) \in Att_i^d \land (a, b) \notin Att_i^d))$ (avoids attacking any attacker of *I*). If $a_i = NIL$,

(iiil) $\neg h_i(a) \land g_i(a) \land ((\forall b \in \mathcal{A}^d)((b, I) \in Att_i^d \land (a, b) \notin Att_i^d))$. If $a_i = NIL$, (iv) $a_i = NIL$.

In English, a focused agent that is opposed to the proposal attempts to either attack it directly or attack an argument that is defending it, whereas an agent that agrees with the proposal attempts to either defend it or avoid attacking it altogether. This behavioral approach is similar to debate protocols in abstract argumentation but for the fact that agents may advance arguments that do not result in an "advancement" of the debate. In our case, the debates stop for other reasons (see Section 2.4).

Two important points deserve to be highlighted. Firstly, focused agents with small argument sacks will regularly find themselves applying rules (*iii*) or (*iv*). In effect, if they do not have enough arguments to infer attacks and defenses, their behavior in deliberation is likely to be similar to that of the naive agents. Secondly, the naive and focused behavioral assumptions presented here are very different in terms of computational complexity. The first assumes that the agent observes the debate but what she sees has no effect on her course of action; i.e., the argument she voices is uniquely determined by how she feels about the principle behind the proposal argument. In the second, an agent debates, essentially, to knock out any proposal argument she disagrees with⁸.

2.4. Tables for Deliberation. Tables are the physical or virtual places where the exchange of arguments occurs. Agents deliberate on these tables to determine whether the proposal argument is acceptable or not and are, thus, subdued by the deliberation procedure imposed by the table's central authority (CA) [27]. The CA decides on the structure and length of the collective decision-making process and the deliberation procedure. It controls the percentage of agents from the population that actively participates in the debate (n_D) and the labeling-based semantics used to extract acceptable arguments from the framework (σ). The percentage n_D denotes either the proportion of agents in the population that gets to advance arguments in a population-scale deliberation, or an independent sample of agents summoned to actively participate in the debate. σ stands for the procedure used to conclude on the epistemic status of the proposal argument and the arguments advanced during deliberation.

The CA has the ability to stop debates at will using a *stop rule* that inherently depends on the number of debates that ought to take place before a decision is deemed sufficiently discussed (\underline{m}), the maximum number of debates that can take place before abandoning deliberation (\overline{m}), and the label given to the proposal argument I ($sr(\underline{m}, \overline{m}, \mathcal{L}ab(I))$). *sr* is a Boolean function whose value "true" signals the call for a vote and/or the end of the decision-making process. \underline{m} is associated with a minimal dialectical or epistemic

requirement to consider a proposal for voting and to a lower bound of the length of the deliberation process; \overline{m} provides an upper bound. Moreover, the CA controls the size of the time interval between debates (t_D) , the collective decision rule (e.g. if there is voting on the deliberated proposals), and the majority quota rule for accepting proposal arguments (α) in the decision-making process. t_D may represent the frequency or density of pair-wise interactions in a decisionmaking process.

2.4.1. The Construction of a Decision-Making Process. A deliberation process or debate in our model is a tool to obtain labels for proposal arguments that are as close as possible to the ideal or consensual ones. To define a deliberation process formally, we introduce the notion of debate step as constituent of a deliberation process. Informally, a *debate step* is a time step in which a debate occurs. Formally and more event-oriented, a debate step (d, N', B', S') of a debate D_I on I is a quadruplet composed of a set of agents $N' \subset N$, a set of arguments $B' \subset \bigcup_{i \in N'} \mathscr{A}_i$, a set of attack relations $S' \subset \mathscr{R}$, and a mapping

$$d: (\mathscr{AF}; 2^{\mathscr{A}} \times 2^{\mathscr{R}}) \longrightarrow \mathscr{AF}$$
$$(AF; (B', S')) \longmapsto AF' = (B \cup B', S \sqcup S')$$
(3)

that adds the arguments in B' and some attacks in S' to a framework AF = (B, S) and where $S \sqcup S' = \{(a, b) \in S \cup S' \mid a, b \in B \cup B'\}$. In the same spirit, we define a deliberation process D_I on a proposal argument I as a sequence of debate steps $(d, N_s, B_s, S_s)_{s>0}$ such that $N_0 = \emptyset$ and

$$d\left(\left(\bigcup_{j=0}^{s-1} B_j, \bigsqcup_{j=0}^{s-1} S_j\right); \left(B_s, S_s\right)\right) \in (\mathcal{A}, \mathcal{R}), \quad (4)$$

with

- (i) $B_0 = I$ (the initial argument is the proposal argument);
- (ii) $S_0 = \emptyset$ (no reflexive attack from *I* to *I* is allowed);
- (iii) $(\forall s > 0)(\bigcup_{j \le s} B_j \cap B_s = \emptyset)$ (any newly added argument to the framework has yet to be added to it);
- (iv) $(\forall s > 0)(\bigsqcup_{j < s} S_j \cap S_s = \emptyset)$ (any newly considered attack among arguments cannot be declared among arguments that are not in the framework);
- (v) $\lim_{s\to\infty} (\bigcup_{s>0} B_s, \bigcup_{s>0} S_s) \subseteq (\mathcal{A}, \mathcal{R})$ (the system is stable; no arguments are created during deliberation).

Finally, let $\mathscr{L}ab_d(I)$ denote the label given to argument *I* at a debate step *d* during a deliberation process D_I . A *decisionmaking process*, noted $\mathscr{D}_I = (n_D, t_D, sr(\underline{m}, \overline{m}, \mathscr{L}ab_d(I)), \alpha, \sigma)$, over a proposal whose justification argument is *I* is a sequence of debate and *nondebate steps* $(z_t)_{t\in\mathbb{N}}, z_t \in \{d, \overline{d}\}$, such that

- (i) d are debate steps and \overline{d} nondebate steps that correspond to time steps at which there are no debates.
- (ii) If $t \pmod{t_D} \equiv 0$, then $z_t = d$, and d, otherwise;



FIGURE 3: A representation of a table of deliberation between two groups of 3 agents with 4 arguments each. Each agent has advanced an argument. Agents tacitly agree on the attacks between the arguments. The CA determines the labeling using the grounded labeling-based semantics.

\mathcal{D}_I :	d	\overline{d}	\overline{d}	\overline{d}	d	\overline{d}	\overline{d}	\overline{d}	d	\overline{d}	\overline{d}	\overline{d}	d
Time:	0	1	2	3	4	5	6	7	8	9	10	11	12

FIGURE 4: A decision-making process \mathcal{D}_I or collective decision step starting at t = 0 composed of debate (*d*) and nondebate (\overline{d}) substeps. Debate steps are shaded. The number of nondebate steps between debates is $t_D = 3$; the number of debates in the decision-making process is m = 4.

- (iii) The subsequence $(z_{l_t})_{l_t \in \mathbb{N}}$ of z_t with $l_t = t$ such that $t \pmod{t_D} \equiv 0$ is a deliberation process with $|N'| = n_D \times |N|$;
- (iv) $sr(\cdot, \cdot, \mathcal{S}ab_d(I))$ pursues the deliberation process as long as $\mathcal{S}ab_d(I) = und$ under the semantics σ ;
- (v) The length of the sequence |z_t|, or, equivalently, the duration of the process, is somewhere in the interval [[<u>mt_D</u>, <u>mt_D</u>]];
- (vi) The final labeling for the proposal argument I, $\mathscr{L}ab(I)$, is determined by a combination of its deliberated label $\mathscr{L}ab_d(I)$ and a majority vote of majority quota α .

A decision-making process ends when a final decision has been taken concerning the acceptability of the proposal \mathcal{P} ; e.g. *I* has been deemed unacceptable ($\mathscr{L}ab_d(I) = out$) or a majority of agents have voted against \mathcal{P} . For a representation of a deliberative interaction on a table, see Figure 3; for one of a decision-making process, see Figure 4

Please note that every collective decision can be seen as a "time step" in as much as it describes how agents update opinions and make collective decisions. In the model, the decision-making process and the parameters define the substeps or events that occur within the time step. Henceforth,



FIGURE 5: Process chart for a decision process \mathcal{D}_I on a proposal \mathcal{P} . *t* corresponds to time, and *m* (t_m) corresponds to the number of debate (non-debate) steps that have taken place in the deliberation process (in between two debate steps) over time for the mixed discussion model. We use $\mathcal{L}ab(I)$ and $\mathcal{L}(I)$ interchangeably in the figure for presentation.

comparing simulations on the basis of the different decisionmaking processes translates into comparing these *collective decision steps*.

2.4.2. Deliberation Protocol. To define the deliberation protocol held in the table, either as a consequence of the definition process or as a statement, we assume the following:

- (i) Agents may decide not to contribute to the debate.
- (ii) All agents have the same probability, conditional to their opinions, of being picked to participate in a debate or deliberation step.
- (iii) There is no restriction on the number of times each agent can participate in a deliberation process.
- (iv) Each agent may only place one argument per debate step.
- (v) Arguments that have already been advanced in the debate may attack newly placed arguments⁹.

The deliberation or debate protocol goes as follows:

- The CA randomly generates and makes public a central argument or proposal argument *I* and informs all agents about the rules of the decision-making process.
- (2) The CA randomly draws two sets of $(n_D \times |N|)/2$ agents with divergent views¹⁰ on \mathcal{P} and \mathbb{P} . It merges them to create the set of debaters N'.

- (3) Each agent *i* advances an argument from her sack *A_i*. The CA makes sure that there are no repeating arguments (agents already take argument repetition into account).
- (4) The CA establishes the debate step d's argumentation framework and computes its labeling, L^σ_d(A^d), using σ.
- (5) If the computed label for I is $\mathscr{L}ab_d(I) = und$ or the number of debates steps is inferior or equal to \underline{m} at time t, then the CA stops the debate and resumes it at the $(t + t_D + 1)$ 'th time step, by repeating (1), (2), (3), (4), and (5).
- (6) Let Lab(I) be the final label given to the proposal argument I. If voting is allowed, then if more than α × |N| agents agree with I, I is accepted (Lab(I) = in), it is rejected if strictly less than α × |N| agree with it (Lab(I) = out). When there is a tie, if Lab_d(I) = und, then Lab(I) = out; if Lab_d(I) = in, then Lab(I) = in. When voting is not part of the decision-making process (α = 0), Lab_d(I) = Lab(I).

One important point to notice about this protocol is that it induces a stop rule, *sr* (cf. steps (5) and (6) of protocol) for the decision process and, thus, it always ends. Either agents debate and agree on the proposal's acceptability through argumentation or, after \overline{m} debate steps, they directly vote on it so that a decision concerning the proposal is always reached. Refer to Figure 5 for a sweeping description of the deliberation protocol and the decision-making process.



FIGURE 6: A dyadic interaction between agents i and j as expressed in (5). Agent i's opinion falls within agent j's latitude of noncommitment and conversely. Neither agent affects the opinion of the other.

2.5. Opinion Dynamics for Social Interactions. In this subsection, we describe the opinion diffusion model based on pairwise interactions among agents and deliberation. Sherif's and Hovland's social judgment theory [22] motivates part of our approach. It describes how individuals' opinions change on the basis of their attitude structures. Attitude structures refer to the relative scope, width, or latitude of categories used by individuals when evaluating information, namely, the latitudes of acceptance, rejection, and noncommitment [32]. The idea behind this theory is that individuals change their positions only in accordance to how far or close the communicative cues they receive are from (to) their anchor positions. It holds that if communicative cues are far (close) from (to) an agent's position, say over her latitude of rejection (acceptance), then the agent shifts her position away from (towards) the position defended by the cues. In the case where the cues fall within the agent's latitude of noncommitment, her position does not change (see Figure 6 and Equation (5)).

2.5.1. Pair-Wise or Dyadic Opinion Dynamics. As agents may communicate and deliberate collectively, they may also engage in one-to-one conversations with other agents to ponder on their positions. We loosely associate this type of communication with dyadic nonargumentative exchange or discussion based on fallacious arguments and persuasion. In the light of social judgment theory and the description of agents in the system, we model pair-wise symmetric interactions following the opinion dynamics model in [9]. An agent *j*'s influence on an agent *i*'s opinion at time *t* is governed by the ensuing difference equation:

$$o_{i}^{t+1} = \begin{cases} o_{i}^{t} + \mu_{i} \left(o_{j}^{t} - o_{i}^{t} \right) & if \quad \left| o_{i}^{t} - o_{j}^{t} \right| < U_{i} \\ o_{i}^{t} - \mu_{i} \left(o_{j}^{t} - o_{i}^{t} \right) & if \quad \left| o_{i}^{t} - o_{i}^{t} \right| > T_{i} \\ o_{i}^{t} & otherwise \end{cases}$$
(5)

where the parameter $\mu_i \in [0, 0.5]$ controls for the strength of the attraction and repulsion in social influence and T_i and U_i are the latitudes of rejection and acceptance for agent *i*, respectively. The parameter μ_i may be thought of as the relative importance agent *i* gives to the opinions of her peers, or, analogously, the weight she gives to her own opinion when updating. $\mu < 0.5$ means that agents will never give more weight to the opinions of other agents than to their own (egocentric bias). When two agents i and j discuss, if, for instance, j advances an informational cue (argument, persuasion tactic) and it happens to be close enough to i's opinion anchor, then i shifts her opinion towards the direction of j's informational cue. The symmetric influence from agent i to agent j takes effect in the same way (see Figure 6).

We add to the dyadic dynamics a *rule of encounters*: at each step of pair-wise interactions, each agent meets exactly one other agent at random. This rule may be associated with random day-to-day encounters among agents. Steps of social influence correspond to nondebate steps (\overline{d}) in decision-making processes.

2.5.2. Deliberative Opinion Dynamics. We define an opinion update equation that links the proposal arguments that are advanced during deliberation and opinions. We combine the uncertain and probabilistic nature of the effect of deliberation, have it a moderating [15, 16] or polarizing [14] one, with a mechanism similar to the one in social judgment theory based on the distance between different informational cues and opinion anchors [32]. The probabilistic modeling of the opinion update can be related to deliberation encouraging open-mindedness¹¹ during collective discussion [18]. From this choice, it follows that deliberated cues can affect even the most extreme of agents, as opposed to some classic opinion diffusion models (e.g. [5, 8]) where once agents become extremists they may no longer become moderate.

Let \mathscr{P} be a proposal, v_i^t the proposal argument *I*'s level of support for a principle \mathbb{P} , and o_i^t an agent *i*'s opinion at time *t*. Then, given the difference $\delta_i^t = (v_i^t - o_i^t)/2$, we define *i*'s probability of being attracted to a decision's informational cue *I* by $x_i^t = p_a^{|\delta_i^t|}$, where p_a denotes a general probability parameter that characterizes how important deliberated results are for the group. The parameter may also be interpreted as the group's tendency to be swayed by a decisional majority. Similarly, we define *i*'s probability of being repulsed from a decision informational cue *I* by $y_i^t = p_r^{1/|\delta_i^t|}$, where p_r denotes a general probability parameter that characterizes the group's dislike of deliberated results. p_r can also be thought of as the group's distrust of the system symbolized by deliberation and democracy.

Let $\mathcal{L}ab(I)$ denote the epistemic status of the proposal argument *I* at the end of a decision process over \mathcal{P} ; every agent *i* updates her opinion as follows:

(i) If
$$\mathscr{L}ab(I) = in$$
:

$$o_{i}^{t+1} = \begin{cases} o_{i}^{t} + 2\gamma_{i}\delta_{i}^{t} & with \ probability \ x_{i}^{t} \\ o_{i}^{t} - 2\gamma_{i}\delta_{i}^{t} & with \ probability \ y_{i}^{t} \\ o_{i}^{t} & with \ probability \ 1 - x_{i}^{t} - y_{i}^{t} \end{cases}$$
(ii) If $\mathscr{L}ab(I) = out: o_{i}^{t+1} = o_{i}^{t}$

where $\gamma_i \in [0, 0.5]$ is the strength of repulsion and attraction in the dynamic. The meaning of γ_i is analogous to the one of μ_i in the dyadic interactions model but for the observation that *i* weights her opinion to an argument rather than to an opinion. The interpretation of this dynamics is straightforward. If the deliberated proposal argument is close to an agent's opinion, then it is very likely that the agent shifts its opinion towards it. Please note that the probability x_i^t (y_i^t) that an agent is attracted (repelled) to (from) an accepted deliberated proposal argument is bounded below (above) by p_a (p_r) (equivalently, $x_i^t \in [p_a, 1]$; $y_i^t \in [0, p_r]$). Steps of deliberative opinion dynamics are associated with debate steps in which a concluding collective decision is made.

2.5.3. Mixed Opinion Dynamics. Opinion dynamics in the context of decision-making processes can be best described as a combination of the two preceding dynamics. In this opinion dynamics model, agents are engaged in a democratic system in which they deliberate, vote for proposals, and occasionally discuss with one another on the principle supported by the proposals. The opinion dynamics for an agent *i* can be written as a sequence of dyadic and deliberative opinion updates:

$$o_{i}^{t+1} = \begin{cases} Equation (5) & if \ t \neq 0 \ (modt_{D}) \land sr(\cdot) = false \\ Equation (6) & if \ t \equiv 0 \ (modt_{D}) \land sr(\cdot) = true \\ o_{i}^{t} & otherwise \end{cases}$$
(7)

where $sr(\cdot)$ is the decision process's stop rule, t_D the number of time steps between each deliberation step, Equation (5) describes the dynamics for the dyadic interactions, and Equation (6) posits the changes in opinions due to the deliberated cues. Equation (5) applies when agents are not deliberating and when each agent encounters another agent to exchange information that may lead to local opinion updates. When there is deliberation, a handful of agents deliberate and, if they reach the stop condition imposed by the table, all agents vote for the proposal. They update (see Equation (6)) their opinions on the basis of the proposal argument's support for the principle and the result of the scrutiny. Otherwise, there is no change in the opinions of the agents. A vote indicates the end of the decision-making process.

3. Experiments and Calibration

In this section, we introduce the metrics that enable us to observe the simulations and characterize the calibration of the model.

3.1. Observations and Initialization. In this section we describe the simulations and the protocol used to test and observe the results of our model. We introduce the metrics of interest, explain the calibration of the model, describe some results obtained from the simulation data used for calibration, and conclude with a brief discussion on the expected outcomes of the mixed opinion dynamics model.

3.1.1. Metrics or Statistics of Interest. Let S denote the end of a simulation. We are interested in the effect of deliberation and of the model's procedural parameters on the following metrics or statistics:

(i) *Variance of opinions* (*Var*(*o*)): the variance of opinions at time *S*. Since opinions live in the union of the positive and negative unit intervals, $Var(o) \in [0, 1]$. The higher the variance of the distribution is, the more "diverse" the opinions are.

(ii) Proportion of extremists in the population $(prop_{ex})$: the proportion of agents in the population with opinions such that $|o_i| \ge 0.75$. A high proportion of extremists makes "healthy" consensus¹² difficult to reach and deliberation more or less informative.

(iii) *Shifts of opinions (Sh) [33]*: statistic that measures the aggregated change in individual opinion at time *S* with respect to aggregated individual opinion at the beginning of the simulation,

$$Sh = \frac{2\sum_{i \in N} |o_{0,i} - o_{S,i}|}{\max_{i \in N} o_{0,i} - \min_{i \in N} o_{0,i}}.$$
(8)

Sh is positive and bounded above by 2|N| since $o_i \in [-1, 1]$. A low shift statistic implies that the process has a small impact on opinions.

(iv) Judgment or consensual inaccuracy (ec): consensual accuracy of a group consists of an ad hoc statistic measuring a group's ability to infer correct labels for proposal arguments from a decision-making process, given the ideal consensual labeling based on full-information, $\mathscr{L}_{\varepsilon}^{\sigma}(\mathscr{A})$. We use a Hamming-based distance on labelings [34] to define the statistic,

$$ec = \frac{1}{|\mathcal{F}|} \sum_{I \in \mathcal{J}} a_I \left| \mathscr{L}ab_{\varepsilon}\left(I\right) \neq \mathscr{L}ab\left(I\right) \right|, \tag{9}$$

where \mathcal{I} is the set of all discussed proposal arguments and $a_I = 1/2$ if $\mathcal{L}ab_{\varepsilon}(I) = und$ or if $\mathcal{L}ab(I) = und$, $a_I = 1$ otherwise.

ec lives in the interval [0, 1]. An inaccuracy statistic close to 1 indicates that agents, subdued to a particular decision-making process, make many mistakes in judging the labels of the proposal arguments. Note that, when there is voting, all proposal arguments are labeled either *in* or *out* after deliberation.

(v) Coherence (ir): let $\mathcal{L}ab_D(I)$ be the label obtained for I from the deliberation process without voting. The coherence statistic measures how well voting results adjust to results obtained during deliberation only. We use the proportion of arguments that have been labeled *in* in the debate and that agents have voted favorably for,

$$ir = \frac{\left|\left\{I \in \mathcal{F} \mid \mathcal{L}ab_{D}\left(I\right) = in \land \mathcal{L}ab\left(I\right) = in\right\}\right|}{\left|\left\{I \in \mathcal{F} \mid \mathcal{L}ab_{D}\left(I\right) = in\right\}\right|}.$$
 (10)

The coherence statistic's domain is [0, 1]. If after M debates no deliberated central argument has been labeled *in* or voting is not part of the procedure of collective decision-making, then the statistic equals

1 or, said differently, agents are perfectly coherent. This comes for the fact that if the central argument is labeled *out*, then it is not even eligible for voting; if labeled *und*, then the debate will always be coherent with the preferences of the agents since the result will be a simple aggregation of their votes. If the statistic equals 0, then none of the proposal arguments labeled *in* are voted in favorably by the agents. It follows that a high coherence statistic implies that when agents vote for acceptable proposal arguments, the results of the scrutiny reflect the consensual rationality expressed in the deliberation process.

3.1.2. Parameters of Interest. We recall the parameters of interest in our study that are linked to the structure of the decision-making processes:

- (i) <u>m</u>: the minimum number of time steps in which a debate or a scrutiny occurs in a decision-making process before a final deliberated decision is submitted to a vote;
- (ii) n_D: the number of agents that deliberate, as a proportion of the population;
- (iii) t_D : the number of time steps between debates in which pair-wise interactions among agents may occur;
- (iv) α : the proportional majority requirement for the acceptance of a proposal. When $\alpha = 0$, it stands for no voting: "any proposal argument that is labeled *in* during deliberation is accepted."

Please recall that, in terms of the definition of a decisionmaking process, each parameter controls either the length or the content of the sequence $\mathcal{D}_I(t_D, \underline{m})$ or the rules that are applied during the debate steps (α, n_D) (refer to Figure 4).

3.1.3. Initialization. All |N| = 400 agents start off with an opinion o_i drawn from a uniform distribution $\mathcal{U}(-1,1)^{13}$. Given o_i , every agent *i* randomly draws a set \mathcal{A}_i ($|\mathcal{A}_i| = k$) of arguments from a balanced¹⁴ argument pool \mathcal{A} of M = 600 nonneutral ($v_a \neq 0$) arguments on the basis of o_i : if $|o_i| < 0.25$ (the opinion is moderate), then agent *i* randomly fills half of her argument sack with arguments such that $v_a < 0$ and the other half with arguments such that $v_a > 0$ ($\phi_i^m = 0.5$). Otherwise, she randomly fills $\phi_i^e \geq 0.5$ of her sack with arguments such that v_a and o_i are of the same sign and the remaining $(1 - \phi_i^e)$ with arguments of the opposite sign¹⁵.

Like for the opinion of agents, every argument *a*'s adherence to the principle \mathbb{P} , v_a , is drawn from a uniform distribution $\mathcal{U}(-1, 1)$. The attack relation that gives birth to the ideal argumentation framework (AF_{ε}) is established on the basis of the v_a 's and the arguments' epistemic reach, *er*.

Let $\theta(a, b, r) = (|v_a + v_b|/2)^r$ be an auxiliary positive real-valued function that takes two arguments (a, b) and a positive real number (r > 0) as input. The probability that any argument $a \in \mathcal{A}$ creates a link to any other argument $b \in \mathcal{A}$ is given by the equation $p = p_e^{\theta(a,b,2)}$, and, thus, arguments supporting opposing views of the principle always attack each other. We fix p_e , the epistemic correlation parameter, to 0.15. If p_e was greater than 0.15, then arguments of the same sign would attack each other too often (and focused agents would be rarely incited to advance favorable arguments during deliberation). If p_e was lower, then the arguments of the same type would induce an almost empty graph, which is unrealistic. The number of arguments that any argument a can attack is bounded by a's epistemic reach that we fix to 15. A higher value of epistemic reach makes the argument lattice too conflicting and nearly bipartite. A lower epistemic reach makes the lattice not sufficiently conflicting in the sense that too many arguments can attack the proposal argument relatively to the few that can defend it. If any argument attacks the proposal argument, the chances that another argument attacks the attacker are low. Hence, the proposal argument is almost never accepted and deliberative opinion updates become rare.

The arguments in the resulting argumentation framework, AF_{ε} , are given a permanent labeling, $\mathscr{L}_{\varepsilon}^{\sigma}(\mathscr{A})$ (in short $\mathscr{L}_{\varepsilon}$), using σ = grounded semantics. We choose the grounded labeling-based semantics because it provides a unique admissible labeling, respects minimal rationality constraints while simplifying the model, and models a skeptic approach to accepting arguments (see [23, 25, 31]), as it maximizes the cardinality of the set **und** $\mathscr{L}_{\varepsilon}$. To some extent, if we consider proposals as committing because they may guide collective action, choosing a skeptic semantics seems reasonable insofar as it labels an argument *in* or *out* only if it has no reason to label it **und**. For instance, if a proposal is a public policy that requires large amounts of resources and engages to future course of action, then it may also determine policy cycles and heavily burden a group and its future decisions. In important situations like these, it seems reasonable to lengthen deliberative cues and ask for more demanding and "grounded" criteria for policy argument acceptability.

On the proposal's side, we create an argument I whose adherence to \mathbb{P} is also drawn from a $\mathcal{U}(-1, 1)$, and we label it $\mathcal{L}ab(I) = und$. I is interpreted as being the main argument justifying the discussed proposal \mathscr{P} and v_I^t as its support for \mathbb{P} at time t. I cannot attack other arguments but other arguments can attack it. For each argument $a \in \mathcal{A}$, a directed arc or attack from a to I is activated with probability $p_p =$ $0.03^{\theta(a,I,0.2)}$. When p_p 's lower bound is too high, the proposal argument is almost always defeated and, thus, deliberative opinion updates do not happen. When p_p 's lower bound is lower than 0.03, then the opposite occurs—the proposal argument is always accepted due to an absence of attacks towards it. Finally, we set the maximum number of debates to $\overline{m} = 7$ for the decision-making processes stop rules¹⁶.

3.2. Calibration and Simulation Protocol. In this subsection, we discuss the calibration of the models, the termination conditions for runs in the deliberative and mixed model, and the expected outcomes in terms of the observations.

3.2.1. Calibrating Dyadic Opinion Dynamics. Dyadic opinion dynamics correspond to a space of parameters in which argumentation and deliberation spaces are not taken into

TABLE 1: Parameters and parameter types and their domains used to compare dyadic, collective, and mixed opinion dynamics.

Le	ength of D	ØI	Decisional	l in D_I	5	Social influ	ience	Delib	eration	n influence	Cognitive	
t_D	<u>m</u>	\overline{m}	n_D	α	μ	T	U	P_a	p_r	γ	k	Behavior
{1, 3,	6} {1, 3, 6}	{7}	$\{0.01, 0.02, 0.05\}$	$\{0, 0.5, 0.66\}$	{0.1}	$\{1.4, 1.8\}$	$\{0.2, 0.6\}$	{0.3, 0.5}	{0.1}	$\{0.05, 0.1, 0.2\}$	$\{4, 8, 16\}$	{Naive, Focused}

account. Deliberated arguments and informational cues have no effect on agent opinion, but agents still vote for the proposal argument¹⁷. We use and calibrate this model for comparability in terms of all our metrics since they are not explicitly observed in [9], a reference opinion dynamics model. Furthermore, for simplicity and comparability again, we suppose that agents are homogeneous in terms of their attitude structures and opinion weightings. We fix $\mu_i =$ $\mu, T_i = T$, and $U_i = U$ for all $i \in N$ and for some triplet $(\mu, U, T) \in [0, 0.5] \times [0, 2] \times [U, 2]$.

Experiments suggest that the strength of the dyadic interactions is an explanatory factor of the time of convergence ($\mu < 0.5$) and of the dyamics' steady states. Bigger values of μ speed up convergence towards a stable set of opinion clusters and affect the size and the relative position of these clusters in the opinion distribution. μ , however, does not play a major role in the number of opinion clusters observed at the steady state of the dynamics.

We set μ to the value given to it in Jagger et al.'s opinion dynamics model ($\mu = 0.1$) [9]. For one, it is the smallest real number for which most results found in [9] hold and the dynamics are smooth; secondly, it seems to be a reasonable "strength" of influence considering that we would not want pair-wise social interactions to completely shadow or mask an eventual effect of deliberation in the opinion dynamics.

For different values of attitude structures, T and U, we get exactly the same results as in [9] in terms of the number and density of the opinion clusters. The different couples of values of U and T model the size of the latitude of noncommitment (T - U) and, thus, the agents' propensity to update opinions. Whenever $T - U \leq 0.2$, regularities in convergence and opinion clustering appear. If U > 1, there is always central convergence; if T < 1, there is always bipolar convergence with clusters of similar sizes. For T – U > 0.2, relatively low values of T (T < 1) and relatively high values of U always yield extreme bipolar convergence with a relatively small cluster of moderate agents (bipolarcentral convergence). For T sufficiently high $(T \ge 1.2)$ and U sufficiently low ($U \leq 0.8$), in other words not very egoinvolved agents, there is a bigger spectrum of steady states. For this reason, we studied the metrics of interest for these value differences and kept the couples of (U, T) depicted in Table 1. For these couples, central, bipolar, bipolar-central convergence (3 clusters), and multicluster convergence (4 or 5 clusters) are possible and result from meaningfully different attitude structures. Judgment accuracy varies significantly across the four different attitude structures we retain and so it happens with the variance of opinions. These scenarios provides us with reference results from which to build upon.

3.2.2. Calibrating Deliberative Social Interactions. Deliberative social interactions correspond to social situations in which individuals are not influenced by pair-wise discussion with peers but are sensitive to collectively deliberated informational cues. On these terms, parameters such as the size of agents' argument sacks (k_i) , the distribution of arguments in them (ϕ_i) , the minimum number of debates (\underline{m}) , and the proportion of debaters from the population (n_D) are no longer mute. We assume, for simplicity, that the formula $\exists \phi^e \exists k \forall i (k_i = k \land \phi_i^e = \phi^e \land \phi_i^m = 0.5)$ is true with $i \in N$, for some value $\phi^e \in [0.5, 1]$, and for $k \in \mathbb{N}$.

The chosen domains for these parameters (see Table 1) are justified by the size of the population and the size of the argument pool (M = 600) which we set from the start. If k is too big ($k \ge 32$) and/or too many agents are allowed in the debate ($n_D \ge 0.2$), then individuals always disprove the central argument and deliberation is never or rarely taken into account in the update of opinions. Similar effects occur when \underline{m} is high, yet we only calibrate it with respect to the mean running time of a simulation and the values of t_D . Different distributions of ϕ_i^e seem to significantly or directly explain none of the statistics we analyze, so we conveniently set $\phi_i^e = 0.75$. We chose the parameter k to be always divisible by 4, for it is convenient given the initial conditions regarding the argument distribution in the agents' argument sacks.

We fix the value space for the parameters γ , p_a , p_r as in Table 1 to account for different intensities of the effect of deliberative voting retroaction on the system. For example, the scenarios where the three parameters are at their lowest values correspond to the scenarios with the smallest fixed prior effect of deliberation in the model. Increasing any of these parameters should, mechanically, be associated with a world in which agents are more sensitive to collective decisions. The values for p_a are taken from [13], where the author states that, by means of deliberation, 7% to 28% of individuals changed their opinion from agreeing to disagreeing or vice versa on a referendum question for Denmark's participation in the Euro. We choose two different values for p_a that account for high (0.3) and very high (0.5) sensitivity to deliberation in a group. p_r , on the other hand, is taken to be small (equal to 0.1), since we believe that there exist agents that will always go against the reached consensus. Their numbers are meager.

For simplicity, we assume that, for all $i \in N$, $\gamma_i = \gamma$ for some $\gamma \in [0, 0.5]$. In effect, we posit that the heterogeneity of agents in the deliberative context only derives from their arguments and their opinions. We decide to calibrate γ in allusion to the strength of the social dynamics in pair-wise interactions. In doing so, we limit ourselves to consider three different scenarios in which deliberation has, respectively, half, equal, and twice as much opinion-shifting power than one-to-one social influence. Finally, we include the acceptability voting quota rule α whose domain is inspired from classical majority rules [35] observed empirically (Table 1).

TABLE 2: Parameters and their domains used to study the effect of the procedural parameters on the metrics.

Procedural parameters							Other parameters					
t_D	<u>m</u>	n_D	α	\overline{m}	μ	T	U	Рa	p_r	γ	k	Behavior
{1, 2, 3, 4, 5, 6	{1, 2, 3, 4, 5}	$\{0.01, 0.02, 0.03, 0.04, 0.05\}$	$\{0, 0.5, 0.66\}$	{7}	$\{0.1\}$	{1.6}	{0.2}	{0.3}	{0.05}	{0.2}	{12}	$\{Naive, Focused\}$

3.2.3. Calibrating the Mixed Social Interactions Model. The mixed interaction model ascribes to a parameter space in which the effect of collective choices on our metrics is nontrivial and where deliberation and voting on proposals determine their acceptability. Since the mixed model is equivalent to periodic iterations of pair-wise and collective social interactions, the calibration of the parameters in both preceding models holds (see Table 1 for the calibration). The first reason for this is comparability; the second is that the previous calibration takes into account the fact that both dynamics are going to be combined. We include the frequency of debates by adding the parameter t_D which controls for the amount of pair-wise interaction among agents between two distinct deliberation steps.

3.2.4. Termination Conditions for Runs. Simulations stop once 100 proposals are deliberated on and/or voted for, in other words, after 100 collective decision steps have occured. The number of proposals discussed seems arbitrary, yet it is high enough to observe the effects of deliberation on opinion distributions and on other metrics related to coherence and judgment accuracy. We choose the number of runs as a function of our research questions, which give relevance to the procedural parameters in the mixed model rather than to the parameters set to describe the population and its behavior.

3.2.5. Expected Outcomes of the Simulations. We expect more deliberation, in terms of higher m and n_D , to increase judgment accuracy and, at the same time, to reduce the variance of opinions. In turn, a smaller variance of opinions implies that the argument pool for deliberation is smaller and, therefore, judgment accuracy should be lower. Moreover, variance increasing dyadic interactions (or rejection) should also increase judgment accuracy and coherence since bipolarization and dissensus foster argument diversity in deliberation. The coherence statistic should be stronger in simulations in which central convergence appears quickly and agents are naive. Attitude structures, sensitivity to informational cues, and weights given to deliberated cues that makes paths to bipolar convergence smaller or to central convergence longer should be associated with higher judgment accuracy. Shifts of opinion should be more visible in scenarios in which extreme agents are pulled away from the extremes. Hence, the sensitivity to deliberated cues (p_a, p_r) and γ should explain the shifts. When crossed with high values of U and T, the shifting power of deliberation should be at its highest.

In the end, we do not know how these mechanisms will play out. The results of the subsequent experiments give us insight on the interplay between the aforementioned effects.

4. Results

Before pointing at any result obtained from the mixed interactions model, we describe and compare the dyadic and deliberative interactions models in the parameter space obtained from the calibration in Section 3 (see Table 1). Primarily, we require our results to describe two orthogonal types of opinion dynamics: the pair-wise dyadic and the deliberative. The latter comprises scenarios in which individuals do not influence each other by means of pair-wise discussion and only update their opinions from deliberation; the former scenarios where only pair-wise interactions among agents determine the dynamics. We describe both on the parameter space given in Table 1 by performing at least 20 simulations per scenario. Quantitative results from the pair-wise interaction model alone are described during the calibration and are not treated here-deliberative parameters have no effect on it. Moreover, we show how these different model specifications yield qualitatively different opinion distributions and compare them on the basis of the metrics of interest. To this end, we perform independent two-sample Student's t-tests¹⁸ or Welch t-tests¹⁹ and compute confidence intervals at a 0.95 level of confidence. Last but not least, we comment ordinary least squares (OLS) regression estimates to account for the direction and magnitude²⁰ of the effects of the parameters on the metrics. Estimates are declared significant at a 5% level of risk.

We perform the same analysis on the mixed interactions model, as we compare it to the pair-wise dyadic and deliberative interactions models and discuss the marginal effects of the parameters on the metrics. More precisely, we obtain two different types of results for the mixed interactions model: the first compares the mixed scenarios to their monolithic counterparts with respect to each metric and with respect to the regression estimates; the second gives a clear idea of the marginal effect of each governance parameter (or parameter of interest) on our observations. In the first, we allow control and procedural parameters to vary on a parameter space similar to those of the dyadic and deliberative opinion dynamics (see Table 1). In the second, we fix all of our control parameters, execute 36,000 balanced²¹ runs, and focus only on the one-way, pooled effects of our procedural and behavioral (all agents are focused or naive) parameters. We generate comprehensive graphs and tables to account for the obtained results.

Please notice the slight change in the size of the parameter space for the procedural parameters and in the values of the nonprocedural ones for the first and second type of results (refer to Tables 1 and 2, respectively). Since we want to have a finer idea of the marginal effects of each of the procedural parameters on how well a group decides on proposals, we make the value jumps for each parameter (1)



FIGURE 7: From left to right, common opinion trajectories and distributions for dyadic, deliberative, and mixed opinion dynamics within the structure of different decision-making processes.

small enough to detect significant differences in the estimates and (2) meaningful for interpretation. For the fixed, onevalued parameters in Table 2, we use Table 1 to set them to their minimum, mean, or median values.

4.1. Comments on Dyadic and Deliberative Opinions Dynamics. In this subsection, we present our first results regarding the deliberative opinion dynamics described in Equation (6) and its differences with the dyadic dynamics (Equation (5)) in terms of our observations.

4.1.1. Qualitative Analysis of Pair-Wise Opinion Dynamics. Pair-wise opinion dynamics alone produces multicluster stable convergence with variable cluster sizes (see Figures 7(a) and 7(d)). Insight on the complexity of these dynamics was discussed in Section 3.2.1.

In Figures 7(a) and 7(d), agents discuss randomly with one another in pairs. Clusters in the extremes form quickly since either agents that are close to the extremes attract one another or are convinced to stay close to the extremes by interacting with other agents they disagree with. Other agents with near-extreme positions may be simply attracted to one of the several moderate foci of agents in the opinion distribution. On the other hand, moderate agents are either attracted to the closest opinion focus or pushed towards the extremes of the distribution. Over time, agents in the central focus either attract other agents into it or ignore the opinions of the extreme agents that interact with them. The result of these dynamics yields a multiclustered convergence where the opinion foci are at least at distance U from one another and only the extreme foci (clusters of agents with $|o_i| \ge 0.75$) are at a distance greater than or equal to T from one another. The foci at the bounds of the opinion distribution are very stable and, in analogy to the (1/2d)-rule in assimilation bounded-confidence models [33], the number of clusters that form is roughly equal to U^{-1} .

4.1.2. Qualitative Analysis of Deliberative Opinion Dynamics. Qualitatively, deliberative opinion dynamics yield a loose unipolar convergence of opinions near the center (central convergence, see Figure 7(b)), near the center of either the left or right portions of the distribution (group polarization, see Figure 7(e)) or to a sparse bipolar opinion distribution with clusters at the center and at the extremes (see Figure 7(c)) if mixed with pair-wise interactions. This is probably a reason to expect the variance of opinions, and the proportion of extremists (almost perfectly correlated: $\rho = 0.89$, p < 0.001), to be low for these scenarios. The side towards which the opinion cluster skews depends, essentially, on the valence of the first argument that is collectively accepted.

In Figures 7(e) and 7(b), agents that update their opinions do so at the same time and only when deliberation is successful. In these scenarios, every time a decision is collectively accepted, agents update their opinions towards

		Dependent v	ariable:	
	Var(o)	Sh	ес	ir
	(1)	(2)	(3)	(4)
Constant	0.610^{***}	4.429***	0.472***	0.000
Constant	(0.002)	(0.206)	(0.003)	(0.000)
II = 0.6 (II = 0.2)	-0.012***	2.981***	0.061***	0.000
0 = 0.0 (0 = 0.2)	(0.002)	(0.260)	(0.004)	(0.000)
T = 1.9 (rof T = 1.4)	-0.264^{***}	-3.261***	0.004	0.000
I = 1.6 (Ie1. $I = 1.4$)	(0.002)	(0.260)	(0.004)	(0.000)
$\alpha = 0.66 (rof \alpha = 0.5)$	0.002	-0.059	-0.100^{***}	0.000
a = 0.00 (ref. $a = 0.5$)	(0.002)	(0.184)	(0.003)	(0.000)
(IIT) = (0.6, 1.8) (ref (IIT) = (0.2, 1.4))	-0.324^{***}	-3.728***	-0.033***	0.000
(U, 1) = (0.0, 1.0) (101, (U, 1) = (0.2, 1.4))	(0.003)	(0.368)	(0.006)	(0.000)
Observations	1,440	1,440	1,440	1,440
\mathbb{R}^2	0.983	0.389	0.529	
Adjusted R ²	0.983	0.388	0.527	
Residual Std. Error (df = 1435)	0.032	3.492	0.053	0.000
F Statistic (df = 4; 1435)	21,272.740***	228.686***	402.326***	

TABLE 3: Ordinary least squares estimates for pair-wise interactions scenarios for most of the observed metrics.

Note: *p<0.1; **p<0.05; ***p<0.01

the argument's defended position. Hence, if there is voting and if agents happen to obtain *in* proposal arguments, they inexorably cluster towards one or the other side of the opinion distribution, resulting in opinion convergence remeniscent of group polarization (see Figure 7(e)). Furthermore, as agents cluster in the same side of the opinion spectrum, the arguments that may be advanced in deliberation are fewer and more skewed towards that one side of the opinion spectrum. Agents become more easily persuaded in deliberation and only by arguments supporting one side of the spectrum. Convergence of opinions towards one loose opinion focus becomes faster and more certain. If there is no vote (see Figure 7(b)), the dynamics are similar but for the fact that it is the uniform randomness of the proposal argument that guarantees a central convergence of opinions.

4.1.3. Quantitative Analysis of Pair-Wise Opinion Dynamics. The three parameters that allow for quantitative analysis are U and T, the agents' attitude structure, and the majority quota voting rule α . Table 3 supports the claim that going from $\alpha = 0.5$ to $\alpha = 0.66$ does not affect the variance nor the shifts of opinions. However, judgment accuracy is significantly higher when accepting proposals becomes more difficult. This is rather surprising but also intuitive since decisions that are taken using more restrictive methods of scrutiny should be more accurate.

Concerning the attitude structures, the interaction term between U and T shows that U has a small effect on the variance of opinions given that T is low. This shows that even when the assimilation threshold is high, a sufficiently small rejection threshold can make its one-way effect vanish. However, high levels of both U and T reduce the variance of opinions dramatically. In contrast, when it comes to overall changes in opinion, the magnitude of U's one-way marginal effect increases the shifts of opinions, whereas T lowers it. This can be interpreted as having the rejection dynamics push moderate-to-extreme agents to positions that are close to the extremes and very moderate agents to the center of the distribution. Not too many quantitative changes (on average) can therefore be observed. The fact that higher U is associated with faster convergence of opinions seems to imply lower judgment accuracy. At some point, voting for the same type of proposals for too long may result in many collective mistakes. This is an interesting conclusion to consider in the analysis of the mixed model. Finally, the interaction term shows that T has no strong impact on the metric.

4.1.4. Quantitative Analysis of Deliberative Opinion Dynamics. Seven parameters can be considered for quantitative analysis in this dynamics: the parameters of interest recalled in Section 3.1.2, the size of the argument sacks (k), agent behavior during deliberation, the strength of the deliberation dynamics (γ) , and the measure sensitivity to deliberated results (p_a) . The most "unexpected" result obtained from the deliberation dynamics is that it shows how deliberation contributes to the proportion of extremists and the variance of opinions. It happens that having more agents discuss proposals and having more of these discussions increases the variance of opinions and the proportion of extremists while lowering judgment inaccuracy and coherence. For the former, the fact that far too many arguments are advanced during deliberation plummets the chances that the central argument is accepted; hence, deliberation does not move opinions too often. For the latter, we observe that only "big" differences (going from the lowest value of a parameter to the highest) in protocol actually prove to have a preponderant impact on the metrics. These observations hint at a possible trade-off between judgment accuracy and the variance of

Complexity

		Depender	ıt variable:	
	Var(o)	Sh	ес	ir
	(1)	(2)	(3)	(4)
Constant	0.032***	13.092***	0.370***	1.007***
Constant	(0.001)	(0.406)	(0.001)	(0.001)
$\frac{1}{100} = 3 (\operatorname{rof} \frac{1}{100} = 1)$	0.012***	-2.062***	-0.012***	-0.001
m = 5 (iei. $m = 1$)	(0.001)	(0.276)	(0.001)	(0.001)
$\overline{m} = 6 (rof \overline{m} = 1)$	0.023***	-3.285***	-0.018^{***}	-0.004^{***}
m = 0 (ref. $m = 1$)	(0.001)	(0.276)	(0.001)	(0.001)
n = 0.02 (ref $n = 0.01$)	0.007***	-0.965***	-0.006***	-0.002**
$n_{\rm D} = 0.02$ (IeI. $n_{\rm D} = 0.01$)	(0.001)	(0.276)	(0.001)	(0.001)
n = 0.05 (ref $n = 0.01$)	0.020***	-3.316***	-0.015***	-0.004^{***}
$n_{\rm D} = 0.03$ (IeI. $n_{\rm D} = 0.01$)	(0.001)	(0.276)	(0.001)	(0.001)
$\alpha = 0.5 (ref \alpha = 0)$	0.146***	68.932***	0.068***	-0.507^{***}
u = 0.5 (101. $u = 0$)	(0.001)	(0.276)	(0.001)	(0.001)
$\alpha = 0.66 (rof \alpha = 0)$	0.291***	-13.842***	0.0004	-0.000
u = 0.00 (ici. $u = 0$)	(0.001)	(0.276)	(0.001)	(0.001)
k = 8 (ref k = 4)	0.004^{***}	-0.815^{***}	-0.002^{***}	-0.001
k = 0 (101. $k = 4$)	(0.001)	(0.276)	(0.001)	(0.001)
k = 16 (ref k = 4)	0.008^{***}	-1.118^{***}	-0.004^{***}	-0.003^{***}
k = 10 (101. $k = 4$)	(0.001)	(0.276)	(0.001)	(0.001)
Rehavior - "Focused" (ref "Naive")	0.015***	-1.980^{***}	-0.012^{***}	-0.004^{***}
Denuvior – Toeuseu (Tei. Tvaive)	(0.0005)	(0.225)	(0.0004)	(0.001)
v = 0.1 (ref $v = 0.05$)	-0.020^{***}	5.762***	-0.0003	-0.001
γ = 0.1 (101. γ = 0.03)	(0.001)	(0.276)	(0.001)	(0.001)
y = 0.2 (ref $y = 0.05$)	-0.025^{***}	8.648***	-0.0003	-0.001
$\gamma = 0.2$ (101. $\gamma = 0.03$)	(0.001)	(0.276)	(0.001)	(0.001)
p = 0.5 (ref p = 0.3)	-0.013^{***}	1.581^{***}	0.0004	0.0003
$p_a = 0.5$ (ref. $p_a = 0.5$)	(0.0005)	(0.225)	(0.0004)	(0.001)
Observations	19,440	19,440	19,440	19,440
R ²	0.931	0.844	0.562	0.958
Adjusted R ²	0.931	0.844	0.562	0.958
Residual Std. Error (df = 19427)	0.033	15.701	0.030	0.050
F Statistic (df = 12; 19427)	21,844.790***	8,733.563***	2,079.585***	36,761.250***

TABLE 4: Ordinary least squares estimates for the deliberative interactions scenarios for most of the observed metrics.

Note: *p<0.1; **p<0.05; ***p<0.01

opinions that may be of interest in the mixed discussion model.

Group coherence, shifts of opinions, and the variance of opinions are highly affected by the voting threshold: the coherence metric is reduced by half when allowing for classic majority voting, shifts explode, and the variance of opinions increases dramatically. This is not surprising since the majority quota rule works as a buffer to the success of decision-making processes and, in consequence, it hampers the deliberative opinion updates. For the shifts of opinions, it is clear that the two-third majority voting rule results in little opinion change and that the simple majority rule induces group polarization. As expected, there are less extremists in scenarios where there is no voting and the shifts of opinion are more likely to happen when agents are more sensitive to deliberation. When we include voting in the system, the group's judgment accuracy decreases due to the correct deliberated decisions that are vetoed by a salient opinionagent majority. Strangely, the effect only holds for the simple majority decision rule. Lastly, the proportion of individuals that participate in the debate and the minimal amount of debates have similar marginal effects on all the metrics (see Table 4).

4.1.5. Comparing Deliberative and Pair-Wise Opinion Dynamics in a Nuthshell. Qualitatively, both models yield similar outcomes but for the fact that deliberation never produces opinion polarization with both types of extremists (see Figure 7). Furthermore, clusters in the deliberative interactions model are less homogeneous in size and multicluster convergence occurs only when deliberation is rarely successful, which happens sporadically given the calibration of

N 11			Metric		
Model	Var(o)	<i>prop</i> _{ex}	Sh	ес	ir
Dyadic vs. Deliberative	[0.191, 0.216]	[0.170, 0.197]	[-29.48, -28.28]	[0.074, 0.082]	[0.165, 0.172]
Mixed vs. Deliberative	[0.142, 0.147]	[0.150, 0.155]	[-20.85, -19.76]	[0.003, 0.002]	[-0.038, -0.030]
Mixed vs. Dyadic	[-0.072, -0.046]	[-0.045, -0.018]	[8.332, 8.859]	[-0.079, -0.071]	[-0.205, -0.201]

TABLE 5: Mean difference 0.95 confidence intervals for metrics by model comparison.

the model. On the other hand, pair-wise opinion dynamics produces multicluster convergence most of the time and central convergence only in the infrequent cases where the latitude of acceptance is large.

The comparison of these two scenarios in terms of the metrics is straightforward. Whilst dyadic interactions allow for more variability in opinions through pair-wise interactions and more mistakes in judging proposals, deliberative updates simply act as an antagonist force. On average, the deliberative interaction model produces opinion distributions with smaller variance and yields steady states with higher judgment accuracy, a higher shift statistic, and, naturally, lower group coherence since only proposals that are deliberated *in* are eligible for voting (see Table 5).

4.2. Articulating Dyadic and Deliberative Social Interactions in Opinion Dynamics. So far we have considered scenarios where argumentation had no place and where individuals made decisions according to opinions that were purely constructed from pair-wise discussions. We also observed cases in which agents formed their opinions only by integrating deliberated proposals into their opinion updates. The scenarios we present in this section combine the two preceding scenarios to account for both, the effects of pair-wise discussion and the effects of deliberation on opinions, judgment accuracy, and group coherence. From the preceding observations, we conclude the following: deliberative opinion dynamics reduces the variance and the proportion of extremists because it shifts opinions greatly and polarizes groups by eliminating at least one type of extremist, the one whose opinions are initially opposed to the deliberated results; or it brings opinions close to neutrality (see Figure 7(b)). This may not be desired in a group since the diversity and stability of opinions are necessary for deliberation to make sense and lead to the collective acceptance of controversial proposals. Dyadic opinion dynamics, on the other hand, tend to increase the variance of opinions and the number of extremists while having a lesser aggregated effect on the shifts of opinion. This is an immediate result of the pair-wise interaction dynamics inasmuch as it polarizes individuals very quickly.

To sum up, in the mixed interactions model, deliberation hinders the effect of voting on opinions by making some proposals that would normally be submitted for voting not eligible for voting. It may also undo opinion changes due to pair-wise interactions. The mixed model can be seen as the deliberation opinion dynamics model in which pair-wise interactions occur in between collective discussion and may affect which arguments are advanced during deliberation. Equivalently, it can be seen as the dyadic opinion dynamics model in which deliberation accounts for endogenous "shocks" on the opinion distribution.

4.2.1. Qualitative Analysis of the Mixed Dynamics. In Figures 7(c) and 7(f), agents update their opinions through deliberation and pair-wise interactions. Opinion trajectories visuals show a combination of both previously described opinion trajectories. Over time, as agents discuss with one another, opinion clusters form as in the pair-wise or dyadic opinion dynamics. Deliberation may disrupt the formation of these clusters and three different situations may occur:

- (i) Deliberation disrupts the formation of the opinion clusters but not sufficiently to jeopardize pair-wise opinion cluster formation (e.g. for high values t_D). Two phenomena may be responsible for it: (1) the variability of the proposal argument's support for the principle \mathbb{P} and (2) too few (successive) deliberation steps that result in opinion updates. Once the opinion clusters are formed, successful deliberation foreshadows the merging of the previously formed clusters, may bring extremists closer to moderate agents, and, thanks to the assimilation component of the pair-wise dynamics, allow bigger and more moderate opinion clusters to form (refer to Figure 7(c)).
- (ii) Deliberation significantly disrupts the pair-wise dynamic formation of clusters. This may happen when the requirements for accepting proposals are weak in the decision-making process (e.g. low α). In this case, one has either central convergence, since all agents follow the deliberated results in the same manner, or 3-cluster convergence with two extreme opinion foci. In the latter situation, the extreme groups form quickly (like in the pair-wise model) and deliberation cannot pull extreme agents away from the extreme foci they are in. The main reason for this is that immediately after an extreme agent attempts to leave her foci through deliberation updates, she gets pulled or pushed back into it by means of a dyadic interaction with another extreme agent.
- (iii) Deliberation does not disrupt the formation of the opinions cluster. This is the rare case scenario for very demanding decision-making processes in terms of inclusion (e.g. n_D) and argument acceptability (e.g. α). Opinions converge as in the pair-wise opinion dynamics model (refer to Figures 7(a) and 7(d)).

Last, when there is voting ($\alpha \neq 0$), the position of the moderate opinion cluster is determined by the first

Complexity



FIGURE 8: Significant correlation estimates for the metrics of interest in the mixed opinion dyamics scenarios.

accepted proposal argument and/or the opinion majority that is formed from previous dyadic interactions (see Figure 7(f)).

4.2.2. Relating the Metrics of Interest in Mixed Opinion Dynamics. The first thing to notice is that for all of these scenarios, all the metrics are significantly correlated (see Figure 8). The variance of opinions and the proportion of extremists are almost perfectly correlated. To some extent, this means that a high variance of opinions, known that the mean of opinions is statistically null, equates to having many extremists in the population, be towards one or the other side of the opinion distribution (e.g. extreme bipolarization). Similarly, we find a negative and strong correlation between the judgment inaccuracy measure and the coherence coefficient (see Figure 8)—when judgment inaccuracy is high (low), coherence to deliberation is low (high). Cases like these are telling since they may be related to sequences of collective decisions in which agents do not make too many mistakes when judging proposal arguments and still make deliberated decisions that are, on average, in line with their principles.

The shift statistic, interestingly, is negatively correlated with the variance of opinions, with judgment accuracy, and with the coherence coefficient. In a way, this means that having more moderate agents conduces, on average, to more mistakes in judging arguments and to less coherent collective decisions. An explanation for this phenomenon could be that as agents reach consensus in opinion, important (possibly extreme) arguments for correctly judging argument proposals are not advanced in the deliberation arena, which results in more mistakes and in possibly accepting proposal arguments that, after voting, are not meant to be accepted. Indeed, the 4.2.3. Comparing Mixed Dynamics with Other Monolithic Dynamics of Interactions. Qualitatively, opinion trajectories in the mixed model are more chaotic than in the pairwise and deliberative interaction models (see Figure 7). As analyzed previously, deliberation dynamics disrupt the pairwise dynamics and create situations in which central and multiclustered convergence (with two clusters at the extremes of the distribution) are possible, at the expense of more (pair-wise dynamics) or less (deliberative dynamics) clusters. It tolerates the survival of very small isolated group of agents that the previous models rarely tolerate in a hundred collective decision-making steps (see Figure 7(c)).

One of the most interesting observations one can make of this family of models regards how groups change (in terms of our metrics) when dyadic social interactions are added to deliberative interactions and vice-versa. When comparing the mixed interactions with only deliberative ones, shifts, agents' aggregated change in opinions, are, on average, less pronounced. The variance of opinions and the proportion of extremists are significantly higher for the former than for the latter; group coherence, in contrast, is lower with respect to the scenarios where opinions are formed only through deliberation. Judgment accuracy follows that same trend. A reason for this might be that mixed dynamics are faster to yield semistable clusters and/or opinion convergence. In consequence, the potential argument pool for deliberative exchange shrinks faster and deliberation becomes less effective over time, which results in more collective mistakes. The outcome on coherence is predictable inasmuch as deliberative interactions produce group polarization. If a group agrees in opinion, debates are sterile because the members of the group will most likely refrain from attacking the proposal arguments they all support and will work together to disparage the proposal arguments they do not support (see Table 5).

In the same manner, we report that the proportion of extremists and the variance of opinions are barely lower in the pair-wise interaction model when including deliberation in the decision-making process. Shifts of opinions are also barely lower when there is no deliberation. This can indicate that deliberation does not undo the changes in opinion due to pair-wise interactions nor does it amplifies them. In contrast, when we consider the labeling-based metrics, the mixed interaction model ensures lower judgment inaccuracy because of deliberation and lower group coherence (see Table 5).

In sum, the mixed discussion model is a compromise between the pair-wise interaction and deliberative opinion dynamic models in that it offers some guarantees in terms of judgment accuracy and coherence, while allowing for "reasonable" variance in the opinion pool. It establishes some kind of trade-off between labeling-based metrics and the variance of opinions that may be of interest for decisionmaking process design.

Metrics	Parameters										
	t_D	<u>m</u>	n_D	α	T	U	Pa	γ	k	Behavior	
Var(o)	∕XX	$\nearrow X$	//X	110	$\searrow \searrow X$	$\searrow \searrow X$	$\searrow \searrow X$	$\searrow \searrow X$	/ / X	//X	
prop _{ex}	$\nearrow XX$	$\nearrow X$	//X	110	$\searrow \searrow X$	$\searrow \searrow X$	$\searrow \searrow X$	$\searrow \searrow X$	//X	//X	
Sh	$\nearrow XX$	$\searrow \searrow X$	$\searrow \searrow X$	110	$\nearrow \nearrow X$	$\searrow \searrow X$	//X	177	$\searrow \searrow X$	$\searrow \searrow X$	
ес	∕XX	$\searrow \searrow X$	$\searrow \searrow X$	225	∕OX	\nearrow	OOX	$O \bigcirc X$	$\searrow \supset X$	$\searrow \searrow X$	
ir	$\searrow XX$	$\searrow \supset X$	$\searrow \searrow X$	\searrow	$\searrow OX$	$\searrow O X$	OOX	OOX	$\searrow \supset X$	$\searrow \searrow X$	

TABLE 6: Summary of sign or relative marginal effects for all varying parameters on dyadic, deliberative, and mixed interactions models in and for all experiments.

Nota: gray symbols correspond to dyadic dynamics, light gray to deliberative dynamics, and bold black to mixed dynamics. Given a scenario, the symbol X stands for constant parameter and O for insignificant marginal effect, and as parameters grow: $2 \lor 1$ for positive (negative) marginal effect for the parameter. From left to right at the parameter level, the leftmost arrow corresponds to the scenario with the strongest (in absolute value) marginal effect of the parameter (with respect to to its factors). Example: Let *p* be a parameter, and $2 \lor 0$ stands for "*p* has a significant positive effect in dyadic discussion scenario, negative in the deliberative one, and insignificant in mixed discussion. The positive effect in dyadic discussion is greater than the absolute value of the negative one in the purely deliberative scenario."

4.2.4. Comments on the Differences between the Three Models for OLS Regression Estimates. The significance of most of the regression parameters estimated for the mixed interactions scenarios are similar to those estimated for the pure dyadic and deliberative scenarios. Nevertheless, the magnitude of the effects is different insomuch that the dynamics themselves are different. So, instead of commenting the linear regression estimates of the mixed model, we compare, when possible, the standardized significant coefficients of the regressors of the mixed model with those of the pair-wise and deliberative interaction models (see Table 6 for a summary of the results).

In mixed interactions, for instance, the influence of *m* on the opinion distribution is much more preponderant than in the deliberative dynamics scenarios. Going from m = 1 to m = 3 increases the standard variation units of the variance of opinions by twice as much as the one in the deliberative scenarios. This, in particular, is due to the design of the decisionmaking process. If m is higher, then the number of social influence steps (t_D) in the decision process is also higher (compared to $t_D = 0$ in the deliberative case). If we decide to consider these steps as contributing to the variance of opinions, then the effect of *m* on the variance of opinions has to be stronger in the mixed interactions model. In contrast, the number of agents participating in the debate has a weaker effect on the variance of opinions in the mixed interactions model than in the deliberative dynamics, specially for n_D = 0.05 (more than twice as much). The difference may come from the fact that, in mixed discussions, the part of the argument pool used in deliberation is more restrictive than in the deliberative case. We can argue that since pair-wise dynamics may get individuals closer to one another after some deliberation, the same happens to the arguments that individuals will most likely use in deliberation. Taking this into account means that deliberation will, at many instances, be unfruitful and have no effect on opinions, thus failing to moderate extreme opinions and/or polarize the group. In the same direction, a high number of participants in deliberation can make the deliberation dynamics stiff. In this case, the distribution of opinion is close to the uniform distribution, which has a relatively high-to-moderate variance with respect to the reference results.

The strength of the deliberative dynamics has a positive stronger effect on the variance of opinions in the mixed model than in the deliberative model when $\gamma = 0.2$, and a negative weaker one when $\gamma = 0.1$. A sound interpretation of this result is that when deliberation is successful and its effect on opinions does not get agents to move sufficiently away from their current positions, agents are pulled or pushed back to their previously held opinions: the variance of opinions does not change much. On the other hand, if the effect of successful deliberation is strong, opinions can change sufficiently to make agents much more prone to assimilation into moderate opinion foci: the variance of opinions falls.

The deliberative opinion dynamics model is much more sensitive to α than the mixed discussion model. Differences in marginal effects can double. The probability of attraction to deliberated results (p_a), for instance, has a strong moderating effect (negative) on agents' opinions in the deliberative opinion dynamics but a meager one (two-fold) in the mixed and dyadic dynamic models. This is because opinion updates in the deliberative model depend only on that parameter, whereas in the mixed model more numerous types of interactions make opinion updates possible.

In terms of judgment accuracy, all parameters except the size of the argument sacks (k) have a significant, slightly stronger effect on the deliberative dynamics than in the mixed model. In effect, the difference may be attributed to the upshot of pair-wise interactions on the distribution of opinions and on the arguments that would be advanced during deliberation. The result respecting parameter k, however, is counterintuitive. Since argument sacks are static, scenarios where shifts are higher should make k more important for deliberation to succeed and it is precisely in the scenarios of pure deliberative dynamics that we find the biggest opinion shifts. This may point to the hypothesis that in deliberative interactions agents change opinions substantially but do not often change their adherence to the principle (go from a negative opinion to a positive one or vice versa) when there is no group polarization.

4.3. Sensitivity Analysis for the Procedural Parameters of *Interest*. We perform a sensitivity analysis of the observations on the parameters of interest in the mixed interactions model.

We extend the parameter domains of the pure dyadic and deliberative scenarios, as we make it explicit in Table 2.

4.3.1. Minimum Number of Debates (\underline{m}). We observe that the minimum number of debates has a significant, well-observed effect on all of our metrics excluding group coherence. For the variance of opinions, we can observe that the more debates there are, the bigger the value of the metric is; although the higher n_D and t_D are, the weaker the overall effect of \underline{m} becomes (Figures 9(d) and 9(a)). Furthermore, the marginal effect of increasing minimal debates is decreasing: an additional deliberation step requirement increases the variance of opinions but less and less as it grows. \underline{m} , coupled with n_D , has a shy S-shaped effect (see Figure 9(d)) on the variance of opinions, meaning that it has a stronger effect when n_D is low, which suggests a possible trade-off between the number of arguments accepted in the debate and the number of debates required before accepting a proposal.

In contrast, the shifts in opinion are less and less likely as m grows and this is independent of the variations of the other parameters. Again, the effect is marginally decreasing and is only truly significant when $\alpha = 0.5$. An explanation to these effects may be linked to the design of the system. First, the variance of opinions is higher when deliberation is asked for because the more deliberation steps there are, the higher the chances are that the central proposal argument is deemed unacceptable, specially when agents are focused. Furthermore, mechanically speaking, increasing the minimal amount of debates implies that, whenever a decision is to be taken, at least $\underline{m} \times t_D$ time steps have to take place and if, at any moment, a proposal argument is considered undecided, t_D time steps are added to the process. So, unless the debate does not yield **und** labels for proposal arguments (highly unlikely considering that σ = grounded semantics), the more nondeliberation steps there are in the decision process, the higher the variance of opinions is. Concerning the shifts, when $\alpha \neq 0.5$, either the system is too stiff to accept any proposal argument, and opinions do not change much, or the effects of pair-wise discussion and deliberation cancel out in such a way that individual shifts are minimal (Figure 9(b)).

On the side of labeling-based metrics, the more debates are asked for, the more accurate a group is in its judgment, the effect being smaller as \underline{m} grows. When agents are naive, the effect is quasilinear, while when they are focused, the strongest effects of acquiring more deliberation are found when levels of deliberation are already low (Figure 9(f)). This can be explained by the fact that the more debates there are in the decision process, the closer one gets to the ideal argumentation framework.

4.3.2. Proportion of the Population in Deliberation Steps (n_D) . Like for \underline{m} , n_D has a significant effect on the proportion of extremists and on the shifts and variance of opinions (Figure 9(e)). This may result from the fact that being able to put more arguments in play at the same debate step may increase the odds of revealing cycles around the proposal argument in AF_{ε} . Given that we use grounded semantics, the arguments in the cycles are labeled **und** and, in consequence, debates are more often postponed when n_D is higher. Postponing debates, in turn, increases the number of nondebates steps in the decision process, which increases the variance of opinions and limits, over time, the moderating effect of deliberation. Moreover, the effect of this parameter is very dependent on the value of α (Figure 9(e)). For shifts, for instance, $\alpha = 0.5$ makes the effect of n_D negative, while $\alpha = 0$ makes it positive but to a lesser extent. Furthermore, for higher requirements of deliberation (\underline{m}) , adding more individuals to the deliberation process has weaker effects on the variance of opinions and on the other metrics. It is also quite interesting to notice that it has the same effect on agents independently of agent behavior during deliberation. This is a surprising result as one would have expected more focused agents in a deliberation arena to cause a raise in the proportion of extremists as they play to knock out opposing proposal arguments, and thereof reducing the effect of deliberation (lowers variance) on the whole population.

Finally, similar to \underline{m} , adding more people into the deliberation process increases judgment accuracy (see Figure 9(c)) and has no clear effect, if any, on group coherence (see Figure 9(i)).

4.3.3. Steps between Deliberation Steps (t_D) . In all configurations, t_D increases the proportion of extremists (variance) and decreases the shifts of opinions among agents. The shifts and the effect on the variance of opinions are present for $\alpha = 0.5$, insignificant otherwise (Figure 9(b)). t_D is highly linked to \underline{m} by construction. When $\underline{m} = 1$, the curve linking the variance of opinions and t_D is convex. As m increases, the curve becomes more and more concave, which means that t_D has a more important effect on the opinion distribution as collective decisions take longer to be achieved. This seems counterintuitive but in reality it reflects the multiplicative relationship between deliberation and pairwise interactions and the semantics in the model. If \underline{m} is low and t_D is high, the effective number of pair-wise interactions is, on average, smaller in the deliberation process. Hence, the structure of the decision-making process hinders any increase in the variance of opinions. If *m* is high, the opposite effect is observed. Additionally, since the grounded semantics yields few in arguments with respect to other admissibilitybased semantics, getting closer to the ideal argumentation framework may make it difficult to obtain the expected effects of deliberative discussion on agents. Lower *m* constrains the variance of opinions by making deliberation more influential on opinions.

In terms of judgment accuracy and coherence, we observe that t_D does not explain coherence, yet it tends to decrease judgment inaccuracy. This may be due to a more varied argument pool in the debate which, in turn, may be a consequence of a higher variance of opinions or vice versa.

4.3.4. Acceptability Voting Quota (α). By far, it is the most influential parameter in our study. It changes the direction and the intensity of the effect of all the other procedural parameters and, by construction, heavily constrains the road to accepting a proposal (see Figure 9).



FIGURE 9: Curves of mean observations for different parameters and metrics for 36,000 runs. Bars are 0.95 confidence intervals. Plots with the same observables (ordinate), except for *ir*, are on the same scale.

The higher the requirement for accepting the proposal is, the higher the proportion of extremists and the variance of opinions are and the lower the shift statistic is. In few words, α constrains the world to dyadic discussion or throws it into a process in which deliberation is much more important than dyadic interactions. This is why, for any $\alpha \neq 0.5$, one either gives too much weight to deliberated results, which shadows pair-wise interactions, or too little weight in the sense that no deliberated result is ever accepted and thus never integrated into the opinions of the agents that have voted for it. The reason for the latter is that the latitude acceptance is too low and, thereafter, pair-wise interactions are not enough to unevenly polarize the population in such a way that deliberated proposal arguments are voted favorably by a non-negligeable majority. Shifts for $\alpha = 0$ are stable for different levels of t_D , which implies that deliberation keeps in check all the shifts of opinions related to an increase in the number of pair-wise interactions. For $\alpha = 0.5$, deliberation is less likely to be successful and shifts tend towards the shift levels for $\alpha = 0$ (see Figure 9(b)).

Concerning labeling-based metrics, α entirely determines the coherence statistic. For $\alpha \neq 0.5$, there is no difference in coherence because of how coherence is defined: it is maximal when $\alpha = 0$ and, strangely, maximal when $\alpha = 0.66$ (see Figure 9(h)). Surprisingly, α has no effect on judgment accuracy. One would have expected a higher α to increase judgment accuracy since agents would hardly collectively accept a proposal argument discussed during deliberation. Proposal arguments have, given how the argument lattice is generated, a higher probability of being rejected than of being accepted in the ideal argumentation framework.

4.3.5. A Word on Focused and Naive Agents. Focused and naive specifications for agent behavior are an important parameter in the decision process since they model how agents choose arguments in the deliberation arena. In the scenarios where agents are focused, agents show equal or higher variance of opinions and are therefore more extreme (see Figure 9(g)), independently of the voting requirements. Shifts of opinion are less likely as well. We can see this happening because focused agents knock out proposal arguments more often than naive agents, provided that n_D is low.

One could have expected naive agents to yield more coherent decisions since they argue sincerely. However, this seems not to be the case (see Figure 9(i)). Agents are unable to deliberate in a way that makes deliberation reflect their voting intentions either because they do not have the necessary arguments to do so or because such arguments do not exist.

In terms of judgment accuracy, focused agents do better than naive agents in finding the ideal labels for the proposal arguments. This is likely because, when reconstructing the framework, focused agents take into account the deliberated proposal argument and choose the most pertinent arguments in their sacks that relate to the state of affairs at the deliberation arena. *In fine*, debates result in a better approximation of $\mathscr{L}ab_{\varepsilon}(I)$ even though one could have believed the opposite; that is, focused agents are ready to sacrifice a correct collective label of a proposal argument in the goal of defending their positions. 4.4. Points of Discussion. In this subsection, we discuss the results obtained from the simulations. We attempt to make sense of them from a social sciences perspective.

4.4.1. Opinion Consensus and Dissensus. The deliberative opinion dynamics corresponds to the simulated scenarios in which the variance of opinions is at its lowest and judgment accuracy at its highest. Dyadic opinion dynamics are associated with the scenarios in which the variance of opinions is at its highest, judgment accuracy at its lowest, and coherence at its maximum possible level. The mixed interactions dynamics is a combination of these aspects for all the metrics except for coherence at which it is the worse-deliberation leads to some kind of loose opinion consensus that is reinforced by pair-wise interactions. Likewise, the mixed discussion model confirms some results exposed in [16] expressing that opinions and voting intentions in a deliberative context "often change." That being said, and more particularly in deliberative opinion dynamics, outputs align well with two interesting results on consensus formation given by an interpretation of Moscovici and Doise's [15] work, Sunstein's [14] account on deliberation and group polarization, and social impact theory [4]. The first states that mild consensus is reached in deliberation as procedure disengages agents; the second is that, within discussion groups, majorities tend to become larger and opinions more extreme-not on the basis of competing arguments but of the initial balance of opinions. After several decision processes, we find opinion distributions that reflect these observations even though we are not able to precisely coin the set of parameters that yields such distributions (see Figure 7).

Although differences between mixed and dyadic opinion dynamics in terms of the variance of opinions and shifts are slim, the former does much better in terms of judgment accuracy. In this sense, deliberation does moderate opinions, but not as much when non-deliberation steps are relatively frequent throughout the decision-making process. In terms of Moscovici and Doise theory of consensus [15], this makes sense since pair-wise discussion engages or egoinvolve agents in the discussion and pushes to convergence of opinion at the extremes of the opinion distribution. However, what remains unclear is the reason why procedural parameters that constrain deliberation do not disengage agents from discussion and thus result in a mild consensus, as also expressed by Moscovici and Doise [15]. A possible explanation of this phenomenon may stem from the fact that deliberation in the model is always accompanied by a series of ego-involving discussions that shadow or counter the moderating effect of deliberation. Not only this, deliberation could also be responsible for crystallizing the opinions of extremists for it favors one or the other part of the distribution of opinions, as explained in [14] and to some extent in [13]. Furthermore, the mixed discussion model gives an example of polarization that results from a minority of dissenters on one side of an issue acceding to views of the majority's side [4], as opposed to the strong polarization observed in dyadic interactions. Situations in which the status quo or the undecidability of the proposal argument lingers cannot possibly yield situations of moderation. In cognitive science, one can be inclined to believe that the longer a deliberation process is, the more likely individuals will take an immutable stance on the situation. In view of the assumption that long decision-making processes can be more costly for a group or society as a whole, agents may feel forced to take a stance just for the sake of ending the process. The stance that is adopted would therefore depend on pair-wise interactions among individuals and not on deliberation, since deliberation takes too long to be conclusive.

4.4.2. For Hypothetical Recommendations in Decision-Making Process Elaboration. On another tune, results from the model suggest that procedural parameters have marginally decreasing positive or negative effects on the metrics of interest. In a hypothetical policy analysis perspective, this may indicate that, after choosing a correct number of individuals and imposing the right amount of debates, one can attain reasonable levels of judgment accuracy, coherence, and variance of opinions. As accounted in [18], deliberation increases knowledge (collective knowledge as well) and, therefore, it heightens the chances of making correct decisions. If one believes that social welfare is linked to deliberation and eventually to the metrics studied previously, then procedural parameters can both be used as an instrument to attain desired levels of social welfare and reinforce the notion of legitimacy of collective decisions. It can also provide a justification for choosing a deliberation regime rather than another on the basis of how important (or urgent) a topic of discussion is. Indeed, if we adopt the claim that deliberation is a forerunner of welfare, then any state in which deliberative cues are at their maximum (highest number of participants, many deliberation steps before making a decision, and as frequent as possible) has to be mapped to the highest attainable social welfare. In this case, is a situation where judgment accuracy is at its highest yet the variance of opinions or the proportion of extremists at its lowest an ideal situation? We cannot say for sure since strong opinion consensus and long decision-making processes, in many situations, may result in a loss of welfare rather than in a gain of it.

Another interesting result in procedure for deliberation is the very similar estimates we find for the effects of the minimum number (m) of debates necessary to make a decision on a proposal and the proportion (n_D) of individuals participating in deliberation on judgment accuracy. In the scope of social welfare, this raises the question of whether advocating for longer decision processes by demanding more deliberation steps (higher \underline{m}) rather than bigger debates (higher n_D) may increase welfare. Depending on the objectives of the decider of the deliberation procedure, he or she may choose to concentrate on one or the other. A recommendation we can assert from the mixed social interactions model is that if one prefers low to high variance of opinions, and/or high to low coherence, the decider is better off if she focuses on increasing the size of the debate rather than organizing debates very frequently. One of the reasons for this might be that when making bigger debates, all agents are truly considered as equals in the decision-making process. In organizing many debates, individuals that are chosen to participate in many of these will be more influential and will

somehow bias the set of arguments in the deliberation arena. The majority rule modeled through α also determines the size of the shifts and the distribution of opinions. Because a bigger α is associated with higher levels of extremism and opinion variance, deliberated cues are harder to accept and, therefore, it becomes more difficult for agents to reach concensus and/or moderate their opinions.

On the other hand and further into the idea of legitimacy, the goal of deliberative processes is the agreement among all participants. This agreement stands for the right or best decision from a formal or procedural perspective and it is also the best from a substantive point of view. In other words, the legitimacy of decisions not only derives from well-established procedural requirements (respect of the protocol) but also begs for the fulfillment of two essential conditions: (1) satisfying the procedural requirements for a correct procedure (formal legitimacy) and (2) the rational acceptability of the results of this procedure (substantive acceptability). Legitimacy can be then obtained from the coherence statistic scrutinized in the model insofar as it measures how well agents accept the results of a correct procedure (e.g. deliberation). So, for a proposal to be legitimate, one has to choose parameters that would maximize group coherence in decision-making processes.

As deliberation may give rise to consensus in the deliberation arena, it may also sow the seeds of dissensus in the group. Our model clearly illustrates this phenomenon through the procedural parameters of deliberation and the metrics observed. From another perspective in deliberative democracy, dissensus in collective decision-making may be desirable. According to Landermore and Page's accounts [36], lack of consensus can be perceived as having agents with alternative ideas that, combined with other ideas, can provide a better approximation of the ideal status of an argument. The authors in [36] call this dissensus "positive dissensus" and it makes normative sense when solving complex collective decision-making problems. Our model captures the idea of having normative requirements to define a good and legitimate collective decision: requirements of high consensual deliberative accuracy (judgment accuracy), of legitimacy (coherence), and of diversity (variance of opinions) are all taken into account, and, for such, we can conclude that our model is successful in describing a sound process of collective decision-making with interpretable outcomes.

Of course, the present simulation results and analysis do not prove that all opinion dynamics can be accounted for by such simple processes like the one presented here. They do, however, shed light on the expressiveness of combining different paradigms to account for more interpretable models and pave the way to creating original models of the sort. The simulations suggest the desirability of discovering the consequences of relatively simple laws of communication at different levels (micro, meso, and macro) to determine what still needs to be explained.

5. Related Work

We see our model as a contribution to the influence and opinion dynamics field in ABM and a pragmatic application of abstract argumentation theory. To our knowledge, we are unaware of existing literature on agent-based modeling that explicitly relates collective choice and the notions of deliberation and opinion diffusion through abstract argumentation as we have done it.

Most models in the literature of opinion diffusion are interested in opinions because these have an influence in collective decision and in questions of social order. For instance, in [3] the authors are interested in consensus and in how a group collectively chooses among two alternatives. In other models, authors are interested in the emergence of extremism [8] and on the distribution of opinions when extremists are introduced in the population [5, 33, 37], while other authors coin the notion of opinion polarization as an emergent property of the system [8, 12]. They show, using models of "bounded confidence" and opinion diffusion with trust, that three different kinds of steady states (unipolar, bipolar, and central) are possible depending on whether agents are sufficiently uncertain about their opinions and sufficiently connected and/or a certain proportion of individuals are already extreme. Recent articles on information and opinion dynamics stress the importance of governance and government intervention in the spread of emotions (opinions), in the frangibility of social consensus system. In [10], the authors show that government intervention in the spread of negative emotions can lead to an even faster spread of negative emotions (single-peeked convergence) and in a faster collapse of the social consensus system. Opinion as a function of trust is studied in [12] and relates to how individuals form their opinions on the basis of how much they trust agents in their networks.

Another stream of discrete opinion dynamics models that explain the emergence of opinion can be found in the literature. Computer multiagent models of attitude change based on Latane's social impact theory are presented in [4]. In the theory, an agent changes her opinion on the basis of the informational impact she is subject to which depends on the persuasiveness (strength), supportiveness, and immediacy (group structure) of the environment and information she receives. The simulations predict two emergent groups of phenomena: the shifting of attitudes towards incompletely polarized equilibria and the formation of coherent clustering of subgroups with deviant attitudes. Similarly, but in a continuous, more argumentative fashion, Mäs and Flache [11] present a model of opinion diffusion using arguments. Arguments are considered to be for or against a proposal and are given an agent-dependant relevance in dyadic persuasive interactions. Arguments also determine the agents' opinions and their dissimilarities, which posits the rules of pair-wise encounters in their model as agents are assumed to form their opinions according to the arguments they own. They show that in an argumentative opinion diffusion model, the only observable steady states are bipolarization and consensus and that bipolarization can emerge from interactions among similar agents. In other words, bipolarization of opinions is possible with homophily and without negative influence if there is argumentation. They subsequently compare their results to experimental data. In [38], a similar approach to argumentation and opinion formation is taken but for the fact

that they introduce different types of arguments, explicitly apply social judgment theory [22], and use survey data to establish model-to-real-world comparisons.

Closer to opinion formation and abstract argumentation, Gabbriellini and Torroni [39] are the first to have originally and soundly merged opinion diffusion and abstract argumentation. They define an agent's opinion as a function of the arguments she holds and the attack relation among them. They devised a focused peer-to-peer dialogue system of persuasion (NetArg), inspired from Mercier and Sperber's argumentative theory of reasoning [40], which used only abstract argumentation to study opinion polarization and opinion dynamics. Moreover, they considered networks and the notions of trust and epistemic vigilance to define a dynamics for knowledge revision and trust itself, which we clearly lack in our model. They showed that if a conservative belief operator in argumentation was applied by agents when they reasoned, then their dialogue protocol did not increase polarization among agents. That said, they used the model to study the effect of Groenvetter's weak link theory in the spreading of arguments and were not particularly interested in notions like judgment accuracy or coherence, nor in tackling questions related to collective decision-making procedures. Their system is very expressive and helps position our work in the litterature: our work is at the frontier of pure argumentative opinion diffusion models, opinion diffusion with arguments, and continuous boundedconfidence models of opinion diffusion à la Deffuant.

On another note, work on collective cognitive convergence [41] and opinion sharing [19] shows that consensus towards a certain "correct" opinion or cognitive state is always possible yet dependent on noise, variability, and awareness of agents. In [20], the authors show that learning about an exogenous correct state of the world (represented by bits) under confidence was possible but only if the agents were not too confident. In a homogeneous population, they show that the higher the confidence, the worse the learning—the very confident agents do not learn the properties of the true state of the world and disrupt the learning process of the less confident ones. Collective cognitive convergence can be seen as a result of deliberation in truth-seeking models.

When it comes to abstract argumentation theory, we take an approach that wires two types of dialogues that are wellstudied in the literature: persuasion dialogues [42] and deliberation dialogues [43]. Another interesting line of work is the one on mechanism design [44], or the problem of devising an argumentation protocol where strategic argumentation is not a liability for success in debates. We tackle mechanism design in a different way, though. Instead of considering strategy-proofness, we are interested in how differences in protocol can result in epistemologically "better" collective choices and guarantee that opinion distributions are favorable for deliberation. For a survey on persuasion dialogue, see [42].

Work on agent-based argumentation usually assumes that the semantic relationship between arguments is fixed. In other words, if two individuals were to put two arguments in the public arena such that one logically attacked the other, then everyone would agree that such attack exists [45, 46]. Other models which do not make this restrictive assumption can also be found in the literature [27, 47] and derive from the class or family of opponent models [27, 48] in which two opposing sides attempt to win a dialogue. Our model is in the intersection of these two, but the frame combining opinion diffusion of the kind and argumentation remains new.

The idea of mixing interpersonal influence and vertical communication, however, is not original. It is described and implemented in innovation diffusion models such as [5, 49]. In both, vertical communication is modeled as exogenous transparent information-agents are aware of the existence of an innovation thus triggering several processes of choice and stabilization of opinions. Also, in [50] an Eulerian model is implemented to show the effect of media and exogenous information on opinion distributions. With respect to this point, the originality of our work is in that the information emitted as vertical communication is endogenous. It is issued from a deliberation model that agents shape on the basis of their opinions, arguments, and behavior. In the spirit of [51], where the authors control for the design of vertical communication (to whom the vertical communication is addressed, its timing, how it influences agents opinions), we control for the process generating the information for our own set of variables: length of debates, majority quotas, and the frequency at which discussions take place.

6. Conclusion and Perspectives

The main objective of this article was to build a bridge where decision-making, argumentation, and opinion diffusion could come together. We proposed a model that combined abstract argumentation theory and a bounded confidence opinion diffusion model and showed to what extent it could explain variance of opinions, extremism, coherence in collective decisions, and judgment accuracy on arguments, given an ideal state of full information. The second objective was to show in what way governance, providing agents with arenas of discussion and deliberation, was important in addressing the success of collective decisionmaking processes and the quality of their outcomes.

The model revealed that allowing for more deliberation time, though in lower frequency, and allowing for wider participation in deliberation increased the variance of opinions and the proportion of extremists in a group. These observations are consistent with the combination of results found in [9] and in [13, 14], which stress that deliberation may polarize groups and may have a meager effect on the shifts of opinions, and inconsistent with [16] where it is argued that opinions tend to moderate after deliberation. It happens that deliberation alone did moderate opinion, yet when integrated into a more complex system in which individuals were allowed to interact with one another, and not all that was deliberated was accepted, its influence was shadowed by other more individualistic dynamics.

Undeniably, the grounded semantics played an important role in the weak effect of deliberation since it models a skeptical way of reasoning over arguments. For agents, accepting arguments, thus updating their opinions accordingly, happened not too often. Nevertheless, asking for more deliberation did increase judgment accuracy, as observed in [18], yet in a marginally decreasing fashion. We showed that voting within the deliberation protocol not only increased the proportion of extremists and the variance of opinions in a group but also determined how coherent deliberation and voting procedures were with each other. Lastly, we showed that agents that are focused judged arguments better, had more stable opinions, and constituted a group with higher proportion of extremists than their naive counterparts.

In terms of governance, the model says that there is perhaps no trade-off between extremism and judgment accuracy. Instead, it asserts that the higher the variance of opinions, the closer the group gets to the correct decisions. This situation may be interpreted as Landermore and Page's notion of "positive dissensus" (in complex collective decision-making tasks) [36] as opposed to consensus as a normative requirement for correct collective decision-making. If a decider has to organize deliberation to legitimize a public policy, depending on what kind of world he or she wants, different set of parameters may be chosen to account for different levels of legitimacy and correctness. For instance, one may ask for a decision to be legitimate to respect a certain level of coherence or of judgment accuracy and to be discussed by a sundry of different-opinionated peoples. If so, deliberation has to be made more often, more participants have to be included in debates or in deliberation instances, and not too many time steps should be introduced in between two deliberation instances. So, the model may capture normative ideas to define correct and legitimate collective decisions and to make recommendations on those grounds. Requirements of high consensual deliberative accuracy (judgment accuracy), legitimacy (coherence), and diversity (variance of opinions) are all taken into account, and for such a reason the model is rather successful in describing a sound process of collective decision-making with interpretable outcomes.

6.1. Extensions of the Model and Other Ideas. Our model can be extended in many ways. Value-based abstract argumentation frameworks (VAFs), for example, as an extension of Dung's argumentation framework, provide a formal description of the process of decision-making in which arguments are given values and audiences (set of agents) ignore attacks between arguments on the basis of their preferences over values. Social abstract argumentation frameworks (SAFs) also provide an interesting and seemingly convenient formalism for our model. In SAFs, agents vote on the acceptability of arguments, and the resulting labeling is a combination of the arguments that are voted for and labeling-based semantics. In our model, we apply the same rule yet only for one argument and given precise agent-based voting mechanisms and voting rules, which are eventually parametrized.

To consider our model as a specification of a mixed VAF and SAF model may be of great interest for future work. This formalism can provide solid foundations to describe, explain, interpret, and eventually extend the results found here from an argument-based perspective. On those grounds, future work will focus on creating a hybrid argumentation framework—with its semantics—that takes heed of both: values or principles and voting.

Concerning other possible extensions of the model, a natural reflex would be to argue on the validity and the robustness of the mix between the bounded confidence opinion diffusion models and deliberation. Consequentially, adding deliberation to different bounded confidence models of opinion diffusion such as the Hegelsmann-Krause model [37] or to models of social impact theory like in [4] can be of great interest for future work. We acknowledge that a finer understanding of argumentation frameworks, random lattices, and their respective applications for this model are necessary. Testing for different labeling-based semantics, or even trading labeling-based semantics for scoring-based semantics (semantics that defines acceptability on scores given to arguments) such as debate semantics and rankingbased semantics (semantics that yield a preorder), may be interesting to consider in future work.

More into modeling deliberation, it is not inadequate to believe that agents may also advance proposals during deliberation and replace, to some extent, the proposal for which the deliberation is taking place. Endogenizing the process that produces the type of argument to be deliberated on may be interesting to explore either by having agents strategically replace proposals at certain moments during the debate or having the central authority attempt to choose proposals such that the probability of having them accepted is high. Agents may also be thought to be autoorganizing (independent of a central authority) and concede on the size of the instances of deliberation. To consider differences in semantics across agents or even endogenizing the semantics on how urgent or important the proposal that is being discussed is can also be an interesting direction to take.

Last but not least, many nontrivial modifications of our model are possible and most should include better-thought deliberation protocols. It may be interesting to design and observe protocols in which deliberation only affects agents that are actually debating, for example. Trust, network effects, multidimensionality of opinions, new processes of argument exchange, or learning among agents are notions to further develop in order to make the model more realistic and relax some unreliable assumptions. In sum, the model is to be refined and extended with the objective of either studying concrete cases of deliberative polling and opinion dynamics or implementing more intuitive thought experiments.

Appendix

We present below some of the pseudocode we used to implement the model.

Algorithms:

- (a) Algorithm 1 implements the dyadic influence in the model. It describes how two agents interact and update their opinions.
- (b) Algorithm 2 implements the deliberative influence in the model, which describes how agents update their opinions in sight of the accepted deliberated proposal arguments.

- (c) Algorithm 3 is used to obtain the grounded labeling of an argumentation framework.
- (d) Algorithm 4 provides an overview of a run in the simulations. *M* is the number of proposal arguments to discuss, which we set to 100 in the model.

Each object in the system is indexed by a *who* number. Hence, a set of agents is annotated by a set of natural numbers. We use this fact in the implementation of the algorithm. *Objecti* denotes the object of *who* number *i*. Agents' opinions (*o*) are calibrated so that $o \in [-1, 1]$.

Data Availability

The program and data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

A preliminary version of this work was presented and discussed at the conference *Bridging the Gap between Formal Argumentation and Actual Human Reasoning* on October 4-5, 2018, at Bochum, Germany. The authors thank the audience for their valuable comments and suggestions.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

Acknowledgments

The contribution of Gabriella Pigozzi was supported by the Deutsche Forschungsgemeinschaft (DFG) and the Czech Science Foundation (GACR) as part of the joint project From Shared Evidence to Group Attitudes (RO 4548/6-1).

Endnotes

- 1. In other words, the superiority of the "right answer", or labeling in our case, will appear as such to all.
- 2. All agents hold the same opinion.
- 3. By sensitivity, we mean how much an agent is ready to change her opinion on the sole fact that an informational cue has been debated on.
- 4. The assumption is strong, but this trust among agents may be rooted from a genuine individual motivation to reach a democratic consensus.
- 5. We adopted a rule of thumbs that says that if an agent is neutral with respect to the principle, then she will vote for any proposal that is presented to her during deliberation.
- 6. If the argument is not unique, she chooses one at random.
- 7. If there are more than one candidate, the agent will favor the argument at minimal distance from *I*.

Require: Vector of values for parameters $(V = (\mu, U, T))$ Ensure: Successful interactions among agents $k \longleftarrow 0, N_e \longleftarrow \emptyset$ while $k \leq |N|/2$ do \triangleright Fill N_a with |N|/2 agents. $k \leftarrow k+1, N_{e} \leftarrow Random(|N|+1)$ \triangleright Pseudo-random generator. Generates integer between 0 and |N|. $N_r \longleftarrow \emptyset, \delta_{i,j} \longleftarrow 0$ ▷Variable to store differences in opinion. for all $i \in N_e$ do Randomly pick a number $j \in N \setminus (N_e \cup N_r)$ $N_r \leftarrow N_r \cup \{j\}, \delta_{i,j} \leftarrow |o_i - o_j|$ procedure DYADIC-UPDATE(*i*, *j*) ▷Two agents *i* and *j* discuss and update opinions. $o' \leftarrow o_i, (\mu_i, U_i, T_i) \leftarrow (\mu_i, U_i, T_i) \leftarrow (\mu, U, T)$ >Save agent i's opinion before updating and homogenize agents. if $\delta_{i,j} < U_i$ then $o_i \leftarrow \mu_i o_i + (1 - \mu_i) o_i$ if $\delta_{i,j} > T_i$ then $o_i \leftarrow (1 + \mu_i)o_i - \mu_i o_i$ if $\delta_{i,i} < U_i$ then $o_j \leftarrow \mu_j o' + (1 - \mu_j) o_j$ if $\delta_{i,j} > T_j$ then $o_i \leftarrow (1 + \mu_i)o_i - \mu_i o'$

ALGORITHM 1: Dyadic social influence.

```
Require: Vector of values for parameters (V = (\gamma, p_a, p_r))
Ensure: Successful interaction between proposal arguments and agents
     for all i \in N do
         procedure PROBABILITY-UPDATE(v_a, p_a, p_r) \trianglerightUpdate individual probabilities of change due to deliberated cues
             x_i \longleftarrow p_a^{|v_a - o_i|/2}, y_i \longleftarrow p_r^{2/|v_a - o_i|}
        q \leftarrow 0
         if \mathcal{L}ab(I) = in then
            procedure DELIBERATIVE-UPDATE(i, a)
                                                                                                                                      ⊳Agent i, argument a
              \gamma_i \leftarrow \gamma, q \leftarrow u \sim \mathcal{U}(0, 1)
                                                                                                  \triangleright u is a realization of a uniform random variable
              if q < x_i then
                o_i \leftarrow \gamma_i o_i + (1 - \gamma_i) v_a
                break
              if q < y_i/(1-x_i) then
                 o_i \leftarrow (1 + \gamma_i)o_i - \gamma_i v_a
```



- 8. In terms of computational complexity, naive agents make their moves in $O(|\mathscr{A}_i| \log(|\mathscr{A}_i|))$, while focused agents do at least as bad.
- 9. Tables have memory. Therefore, debates are not necessarily trees.
- 10. If there are less than $(n_D \times |N|)/2$ agents for any of both groups, the CA summons all of them to the deliberation table.
- 11. Or careful consideration of the viewpoints of others which implies that citizens keep an open mind and do not reject arguments outright.
- 12. A "healthy" consensus refers to the case in which many consensual decisions are taken, but, opinion-wise, agents do not agree with each other. Quotes are used on the word healthy because some political science theorists

(e.g [36]) believe that, in certain situations, agreement in opinions for deliberation is epistemologically damaging.

- 13. Simulations on groups of |N| = 800 and |N| = 400 agents shows that scaling the population has an effect on all the metrics, but it is confounding depending on which couples of attitudes structures are considered (U, T). We stick to |N| = 400 as in [9] for comparability and simulation running time.
- 14. By balanced we mean with as many arguments with $v_a < 0$ as with $v_a > 0$.
- 15. Experiments on the system suggest that a bigger (smaller) argument pool M = 1000 (M = 200) is associated with a higher (lower) variance of opinions (and extremism) and higher (lower) judgment accuracy. For the rest of the observations, M seem to have no effect. We choose a middle value for M. Similarly, the argument pool chosen to be balanced or uniformly

Require: Argumentation framework $AF = (\mathscr{A}, \mathscr{R})$ **Ensure:** Assignment of labels to arguments from the grounded labeling of $AF: \mathscr{L}(\mathscr{A})$ $\mathscr{L}_0(\mathscr{A}) \leftarrow (\emptyset, \emptyset, \emptyset)$ $k \leftarrow 0$ **do** *in* $\mathscr{L}_{k+1}(\mathscr{A}) \leftarrow$ *in* $\mathscr{L}_k(\mathscr{A}) \cup \{a \mid a \text{ is not labeled in } \mathscr{L}_k(\mathscr{A}), and \forall b \in \mathscr{A} s.t b \,\mathscr{R} a \text{ then } b \in out \, \mathscr{L}_k(\mathscr{A})\}$ *out* $\mathscr{L}_{k+1}(\mathscr{A}) \leftarrow out \, \mathscr{L}_k(\mathscr{A}) \cup \{a \mid a \text{ is not labeled in } \mathscr{L}_k(\mathscr{A}), and \exists b \in \mathscr{A} s.t b \,\mathscr{R} a \text{ and } b \in$ *in* $\, \mathscr{L}_{k+1}(\mathscr{A})\}$ while $\mathscr{L}_{k+1}(\mathscr{A}) = \mathscr{L}_k(\mathscr{A}) = 0$



```
Require: Vector of values for parameters (V)
Ensure: Vector of statistics (U)
      procedure Set-parameters(V)
                                                                                                                                ▷Initialize parameters for scenario
     procedure INIT-OBJS(n, m, k,...)
                                                                                                                           \trianglerightCreate m arguments, n agents, and AF<sub>e</sub>
     procedure Solve-Argumentation-Framework(AF)
                                                                                                                           \trianglerightAssigns the epistemic labeling \mathscr{L}_{\varepsilon} to \mathscr{A}
     i \leftarrow 1
     t \leftarrow 1
                                                                                                                \triangleright M the number of arguments to discuss
     while i \leq M do
        I_i \leftarrow Gen\_arg(1)
                                                                                                                             \trianglerightGenerates proposal argument I
        A_{i,t} \leftarrow (I_i, \emptyset)
        \mathscr{L}ab(I_i) \leftarrow und
                                                                                                                                             \trianglerightInitialize label for I
                                                                                                                                   \trianglerightInitialize debate counter j
        j \leftarrow 1
        while ((\mathscr{L}ab(I) = und \lor \underline{m} \le j) \land j \le \overline{m}) do
           if t \equiv 0 \pmod{t_D} then
               procedure BUILD-ARGUMENTATION-FRAMEWORK(n_{D}, behavior)
                                                                                                             ▷Build the argument graph, agents deliberate
                  (B_{i,t},S_{i,t}) \longleftarrow (X,Y) \subset (\bigcup_{l \in N_{n_D}} \mathcal{A}_{l,i,t},\bigcup_{l \in N_{n_D}} \mathcal{R}_{l,i,t})
                                                                                                                \rhd \ N_{n_{\mathrm{D}}} \subset N are agents sampled by the CA
                  A_{i,t} \longleftarrow d(A_{i,t}, (B_{i,t}, S_{i,t}))
                                                                                                                                                                \trianglerightFind \mathscr{L}(\mathscr{A}^d)
               procedure Solve-Argumentation-Framework(\mathscr{A}^d)
                  \mathscr{L}ab(I_i) \longleftarrow \mathscr{L}ab_D(I_i)
               j \leftarrow j + 1
           else
               k \leftarrow 1
               while k \leq t_D do
                 procedure DYADIC-SOCIAL-INFLUENCE(\mu, T, U)
                                                                                                                                           ⊳Agents discuss one-to-one
                  k \longleftarrow k + 1
                  t \longleftarrow t + 1
        if \alpha > 0 \land \mathscr{L}ab(I_i) \neq out then
                                                                                                                         ⊳Agents vote for the proposal
           if (|\{l \in N | o_l \times v_{I_i} \ge 0\}| \ge \alpha \times n) then
                \mathcal{L}ab(I_i) \leftarrow in
           else
                \mathcal{L}ab(I_i) \leftarrow out
        if \mathcal{L}ab(I_i) = in then
           procedure Deliberation-influence(\gamma, p_a, p_r)
                                                                                               >Agents are influenced by the result of the decision process
        i \leftarrow i + 1
      t \longleftarrow t+1
      U \leftarrow Compute\_Statistics(V)
                                                                                                                                                 ⊳Get the statistics
      return U
```



generated does not seem to play a significant explanatory role on any of the metrics.

16. Experiments show that the maximum number of debates is not an important parameter even if it is an intuitive one when considering decision-making processes. In practice, debates end after at most $\underline{m} + 1$ deliberation steps, and even then, it happens rarely. However, to avoid unpleasant surprises like infinite debates on a complete argument graph containing the proposal argument, we added the constraint \overline{m} . Its value is precisely 7 because the biggest minimum number of debates we study is 6.

- 17. To ensure comparability with the deliberative and mixed opinion dynamics models on judgment accuracy and coherence, agents vote on proposal arguments every $t_D \times (\underline{m}+1)$. t_D and \underline{m} does not explain variability in judgment accuracy nor in coherence.
- 18. Student t-tests for independent samples are robust to the violation of two hypothesis on the distribution of the samples: normality and homogeneity of variances as long as the sample size is big enough (central limit theorem) and if the difference in size of the compared samples is small, respectively.
- Student's test version for when variances are not assumed equal and the difference in size of the compared samples is big.
- 20. We do not correct for homoskedasticity nor normality of errors in the estimates.
- 21. Balanced as the same number of runs per scenario or such that no significant correlations are found between parameters.

References

- R. Axelrod, "The dissemination of culture: a model with local convergence and global polarization," *Journal of Conflict Resolution*, vol. 41, no. 2, pp. 203–226, 1997.
- [2] D. M. Rousseau, Reinforcing the Micro/Macro Bridge: Organizational Thinking and Pluralistic Vehicles, SAGE Publications Sage, Los Angeles, Calif, USA, 2011.
- [3] S. Galam and S. Moscovici, "Towards a theory of collective phenomena: consensus and attitude changes in groups," *European Journal of Social Psychology*, vol. 21, no. 1, pp. 49–74, 1991.
- [4] A. Nowak, J. Szamrej, and B. Latané, "From private attitude to public opinion: a dynamic theory of social impact," *Psychological Review*, vol. 97, no. 3, pp. 362–376, 1990.
- [5] G. Deffuant, S. Huet, and F. Amblard, "An individual-based model of innovation diffusion mixing social value and individual benefit," *American Journal of Sociology*, vol. 110, no. 4, pp. 1041–1069, 2005.
- [6] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems* (ACS), vol. 3, no. 01n04, pp. 87–98, 2000.
- [7] J. Rouchier, P. Tubaro, and C. Emery, "Opinion transmission in organizations: an agent-based modeling approach," *Computational and Mathematical Organization Theory*, vol. 20, no. 3, pp. 252–277, 2014.
- [8] M. Meadows and D. Cliff, "The relative disagreement model of opinion dynamics: where do extremists come from?" in *Proceedings of the International Workshop on Self-Organizing Systems*, Lecture Notes in Computer Science, pp. 66–77, Springer, 2013.
- [9] W. Jager and F. Amblard, "Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agentbased simulation model of attitude change," *Computational & Mathematical Organization Theory*, vol. 10, no. 4, pp. 295–303, 2005.
- [10] Y.-F. Zhang, H.-Y. Duan, and Z.-L. Geng, "Evolutionary mechanism of frangibility in social consensus system based on negative emotions spread," *Complexity*, vol. 2017, Article ID 4037049, 8 pages, 2017.

- [11] A. Flache and M. W. Macy, "Small worlds and cultural polarization," *The Journal of Mathematical Sociology*, vol. 35, no. 1-3, pp. 146–176, 2011.
- [12] A. Tsang and K. Larson, "Opinion dynamics of skeptical agents," in *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, pp. 277–284, International Foundation for Autonomous Agents and Multiagent Systems, France, May 2014.
- [13] K. M. Hansen, Deliberative democracy and opinion formation, University Press of Southern Denmark Odense, 2004.
- [14] C. R. Sunstein, "The law of group polarization," *Journal of Political Philosophy*, vol. 10, no. 2, pp. 175–195, 2002, Wiley Online Library.
- [15] S. Moscovici and W. Doise, Dissensions et Consensus: une théorie générale des décisions collectives, Presses Universitaires de France-PUF, 1992.
- [16] J. S. Fishkin and R. C. Luskin, "Experimenting with a democratic ideal: deliberative polling and public opinion," *Acta Politica*, vol. 40, no. 3, pp. 284–298, 2005.
- [17] R. C. Luskin, J. S. Fishkin, and S. Iyengar, Considered Opinions on US Foreign Policy: Face-To-Face versus Online Deliberative Polling, International Communication Association, New Orleans, La, USA, 2004.
- [18] J. Barabas, "How deliberation affects policy opinions," *American Political Science Review*, vol. 98, no. 4, pp. 687–701, 2004.
- [19] O. Pryymak, A. Rogers, and N. R. Jennings, "Efficient opinion sharing in large decentralised teams," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, vol. 1, pp. 543–550, Spain, June 2012.
- [20] J. Rouchier and E. Tanimura, "When overconfident agents slow down collective learning," *Simulation*, vol. 88, no. 1, pp. 33–49, 2012.
- [21] D. Miller, "Deliberative democracy and social choice," *Political Studies*, vol. 40, S1, pp. 54–67, 1992.
- [22] M. Sherif and C. I. Hovland, Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change, Yale University Press, 1961.
- [23] P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games," *Artificial Intelligence*, vol. 77, no. 2, pp. 321–357, 1995.
- [24] M. Caminada and G. Pigozzi, "On judgment aggregation in abstract argumentation," *Autonomous Agents and Multi-Agent Systems*, vol. 22, no. 1, pp. 64–102, 2011.
- [25] P. Baroni, M. Caminada, and M. Giacomin, "An introduction to argumentation semantics," *The Knowledge Engineering Review*, vol. 26, no. 4, pp. 365–410, 2011.
- [26] S. Chambers, "Deliberative democratic theory," *Annual Review* of *Political Science*, vol. 6, no. 1, pp. 307–326, 2003.
- [27] E. Bonzon and N. Maudet, "On the outcomes of multiparty persuasion," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems 2011, AAMAS 2011*, vol. 1, pp. 47–54, International Foundation for Autonomous Agents and Multiagent Systems, Taiwan, May 2011.
- [28] J. Rancière, *Disagreement, Politics and Philosophy*, University of Minnesota Press, 1999.
- [29] T. Bench-Capon and K. Atkinson, "Abstract argumentation and values," in *Argumentation in Artificial Intelligence*, pp. 45–64, Springer, 2009.
- [30] M. Caminada, "On the issue of reinstatement in argumentation," in Proceedings of the European Workshop on Logics in

Artificial Intelligence, vol. 4160 of Lecture Notes in Computer Science, pp. 111–123, Springer Berlin Heidelberg.

- [31] G. A. Vreeswijk, "An algorithm to compute minimally grounded and admissible defence sets in argument systems," in *Proceedings of the 2006 conference on Computational Models of Argument: Proceedings of COMMA 2006*, vol. 144, pp. 109–120, IOS Press Amsterdam, The Netherlands, 2006.
- [32] C. W. Sherif, M. Kelly, H. L. Rodgers et al., "Personal involvement, social judgment, and action," *Journal of Personality and Social Psychology*, vol. 27, no. 3, pp. 311–328, 1973.
- [33] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure, "How can extremism prevail? A study based on the relative agreement interaction model," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 4, p. 1, 2002.
- [34] E. Awad, M. W. Caminada, G. Pigozzi, M. Podlaszewski, and I. Rahwan, "Pareto optimality and strategy-proofness in group argument evaluation," *Journal of Logic and Computation*, vol. 27, no. 8, pp. 2581–2609, 2017.
- [35] D. Grossi and G. Pigozzi, "Judgment aggregation: a primer," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 8, no. 2, pp. 1–151, 2014.
- [36] H. Landemore and S. E. Page, "Deliberation and disagreement: problem solving, prediction, and positive dissensus," *Politics, Philosophy and Economics*, vol. 14, no. 3, pp. 229–254, 2015.
- [37] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence: models, analysis and simulation," *Journal* of Artificial Societies and Social Simulation, vol. 5, no. 3, 2002.
- [38] A. Stefanelli and R. Seidl, "Moderate and polarized opinions. Using empirical data for an agent-based simulation," in *Proceedings of the Social Simulation Conference*, 2014.
- [39] S. Gabbriellini and P. Torroni, "A new framework for ABMs based on argumentative reasoning," in *Advances in Social Simulation*, vol. 229, pp. 25–36, Springer, Berlin, Germany, 2014.
- [40] H. Mercier and D. Sperber, "Why do humans reason? Arguments for an argumentative theory," *Behavioral and Brain Sciences*, vol. 34, no. 2, pp. 57–74, 2011.
- [41] H. V. Parunak, T. C. Belding, R. Hilscher, and S. Brueckner, "Modeling and managing collective cognitive convergence," in *In Proceedings of the 7th iNternational Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 3, pp. 1505– 1508, International Foundation for Autonomous Agents and Multiagent Systems, 2008.
- [42] H. Prakken, "Formal systems for persuasion dialogue," *The Knowledge Engineering Review*, vol. 21, no. 2, pp. 163–188, 2006.
- [43] K. Atkinson, T. Bench-Capon, and P. Mcburney, "A dialogue game protocol for multi-agent argument over proposals for action," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 2, pp. 153–171, 2005.
- [44] M. Thimm and A. J. García, "Classification and strategical issues of argumentation games on structured argumentation frameworks," in *Proceedings of the 9th International Conference* on Autonomous Agents and Multiagent Systems, vol. 1, pp. 1247– 1254, International Foundation for Autonomous Agents and Multiagent Systems, Canada, 2010.
- [45] I. Rahwan and K. Larson, "Mechanism design for abstract argumentation," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, *AAMAS 2008*, pp. 1031–1038, International Foundation for Autonomous Agents and Multiagent Systems, Portugal, May 2008.

- [46] H. Prakken, R. Riveret, A. Rotolo, and G. Sartor, "Heuristics in Argumentation: A Game-Theoretical Investigation," *European University Institute*, 2008.
- [47] N. Hadidi, Y. Dimopoulos, and P. Moraitis, "Tactics and concessions for argumentation-based negotiation," in *Proceedings* of the COMMA, vol. 245, pp. 285–296, 2012.
- [48] N. Oren and T. J. Norman, "Arguing using opponent models," in Proceedings of the International Workshop on Argumentation in Multi-Agent Systems, pp. 160–174, Springer, 2009.
- [49] S. Delre, W. Jager, T. Bijmolt, and M. Janssen, "Targeting and timing promotional activities: An agent-based model for the takeoff of new products," *Journal of Business Research*, vol. 60, no. 8, pp. 826–835, 2007.
- [50] A. Mirtabatabaei, P. Jia, and F. Bullo, "Eulerian opinion dynamics with bounded confidence and exogenous inputs," *SIAM Journal on Applied Dynamical Systems*, vol. 13, no. 1, pp. 425– 446, 2014.
- [51] V. Barbet, N. Guiraud, V. Laperrière, and J. Rouchier, *Haggling* on Values: Towards Consensus or Trouble, 2019.



Operations Research

International Journal of Mathematics and Mathematical Sciences







Applied Mathematics

Hindawi

Submit your manuscripts at www.hindawi.com



The Scientific World Journal



Journal of Probability and Statistics







International Journal of Engineering Mathematics

Journal of Complex Analysis

International Journal of Stochastic Analysis



Advances in Numerical Analysis



Mathematics



Mathematical Problems in Engineering



Journal of **Function Spaces**



International Journal of **Differential Equations**



Abstract and Applied Analysis



Discrete Dynamics in Nature and Society



Advances in Mathematical Physics