An opinion diffusion model with deliberation

George Butler, Gabriella Pigozzi, Juliette Rouchier

An opinion diffusion model with deliberation

George Butler, Gabriella Pigozzi, and Juliette Rouchier

Univeristé Paris-Dauphine, PSL Research University, CNRS, UMR 7245, Paris 75016, France
{name,surname}@lamsade.dauphine.fr

Abstract. In this article, we propose an agent-based model of opinion diffusion and voting where agents influence each other through deliberation. The model is inspired from social modeling as it describes a process of collective decision-making that iterates on a series of dyadic inter-individual influence steps and collective deliberation procedures. We study the evolution of opinions and the correctness of decisions taken within a group. We also aim at founding a comprehensive model to describe collective decision-making as a combination of two different paradigms: argumentation theory and agent-based influence models, which are not obvious to link since a formal translation and interpretation of their relationship is required. From a sequence of controlled simulations, we find that deliberation, modeled as an exchange of arguments, reduces the variance of opinions and the number of extremists, as long as not too much deliberation takes place during the decision-making process. Insofar as we define "correct" decisions as those whose supporting arguments survive deliberation, promoting deliberative discussion favors convergence towards correct decisions.

Keywords: Opinion diffusion · abstract argumentation · agent-based modeling · deliberation

1 Introduction

In a group, opinions are formed over affinities and conflicts among the individuals that compose it. Axelrod [3], a pioneer in opinion dynamics, shed light on two key factors required to model the processes of opinion diffusion, namely, social influence (i.e., individuals become more similar when they interact) and homophily (i.e., individuals interact preferentially with similar others). He showed that interactions through those factors lead to emergent collective opinions of which the individual had poor control. Since, a growing body of research has endeavored to identify the conditions under which social influence, at the micro (dyadic) level, translates into macro patterns of diffusion through repeated iterations [26]. Two types of models appear in the literature: on the one hand, the Ising-type models where opinions take discrete values [3,15]; on the other, the continuous opinion models where opinions are represented by real numbers [10,25,18,30,19].

The question of group deliberation, defined as an exchange of arguments, is not explicitly taken into account in opinion diffusion. Opinion dynamics seem to miss the intuition that individual behavior may be determined by factors related to non-dyadic channels of interaction, such as deliberation arenas, and to the structure and size of the channels of communication themselves. When a group engages in a 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 [17] and play a crucial role in consensus formation [28,20,13]. Moscovici and Doise [20] explain that there are two types of "discussions" in deliberation, informal or *warm* and formal or *cold*, that potentially lead to consensus. They show that when a group is asked to reach an agreement through informal

or non-procedural deliberation, the obtained consensus is more likely to be extreme compared to the average of the pre-consensus individual opinions. When deliberation is procedural, the obtained consensus tends to be milder and opinions less polarized. Opinion diffusion has also been used to track convergence towards "correct" opinions. For example, authors in [22,16] study network effects and signaling, but not deliberative protocols. A correct decision corresponds to one derived from a state of the world in which all arguments for and against the decision are taken into account [8]. Deliberation reveals such arguments. Hence, it may help a group converge towards correct decisions. For this reason, decision-making processes with deliberation are interesting to explore.

The aim of our model is to breach the gap between deliberation and opinion diffusion. Drawing from [20], we model warm discussion using an opinion diffusion model based on social judgment theory [27,18], and cold discussion using abstract argumentation theory [12,7]. We engineer a decision-making process with voting that terminates according to deliberated decisions, as we draw inspiration from the literature in deliberative democracy [13,8,28] and opinion diffusion [10,25,30]. We describe the effects of deliberation on opinions and on the correctness, a group's ability to correctly judge propositions, and coherence, a group's ability to accept deliberated proposals, of group decisions by modeling decision-making processes as a sequence of deliberative and dyadic interactions among agents. In particular, we study the impact of the frequency of deliberative interactions and their construction (number of agents, voting rules, etc...) on opinions.

Our model shows that deliberation has a significant overall impact on the distribution of opinions (variance) and on the overall shifts of opinion. We provide evidence of Moscovici and Doise's [20] results on consensus: when specifying opinion dynamics as only deliberative, the proportion of extremists and the variance of opinions are lower than in a non-deliberative specification of the dynamics. However, as observed in [28], if deliberation is mandatory in decision-making processes, more deliberation translates to an increase in the variance of opinions and of the proportion of extremists. The model also explains that the frequency of deliberative interactions as well as the number of agents that participate in deliberation increase judgment accuracy in a marginally decreasing fashion, but have little to no effect on coherence in the decision-making process. Last, we point out that results are strongly conditioned to the voting majority quota and to how agents advance arguments during deliberation.

The remainder of this paper goes as follows: in section 2, we present the model, provide the necessary basics to understand its implementation, and we introduce the metrics of interest. In section 3, we report and discuss our results; sections 4 and 5 are dedicated to related works and to the conclusion of the article.

2 A model for collective decision-making with deliberation

Let N be a group composed of $|N| = n$ agents. The group faces the question of whether to accept or reject a proposal $\mathcal{P}$ justified by an argument I. I, or *proposal argument*, is judged on how well it supports a principle $\mathbb{P}$ or its opposite $\neg\,\mathbb{P}$. Agents discuss the proposal on the basis of their adherence to the principle $\mathbb{P}$. When agents discuss informally, they are subject to random pair-wise influence; when they argue formally, they are impelled by the results obtained in the decision-making process. A *decision-making process* $\mathcal{D}(\mathcal{P}, I)$ on a proposal $\mathcal{P}$ is a sequence of formal and informal discussions that leads to a decision on the acceptance of $\mathcal{P}$ (Fig. 1). A proposal $\mathcal{P}$ is accepted if the argument I that justifies it is accepted in deliberation and/or voted favorably by a majority of agents.

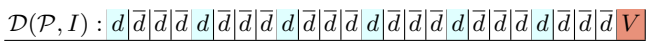


Fig. 1: A decision-making process $\mathcal{D}(\mathcal{P}, I)$. $\overline{d}$ stand for informal discussion steps, d for deliberative interaction, and V to a vote over I. Agents update their opinions according to the obtained result.

2.1 Deliberative agents and opinion dynamics with deliberation

Every agent i has an opinion, a relative position or degree of adherence $o_i \in [-1, 1]$ to the principle $\mathbb{P}$ and a couple $(T_i, U_i) \in [0, 2] \times [0, 2]$ $(U_i \leq T_i)$ of latitudes of rejection and acceptance, respectively, of informational cues. The idea is that there exist levels of relative tolerance from which informational cues have either an attractive or a repulsive effect on the individual [27]. An o_i close to 1 implies that agent i fully supports the principle $\mathbb{P}$, close to -1 that she rejects principle $\mathbb{P}$ or, equivalently, fully supports $\neg\,\mathbb{P}$.

Let $\mathcal{A}$ be a finite set of arguments, seen through the principle $\mathbb{P}$, that agents may hold in a debate over a proposal $\mathcal{P}$. Each agent i has a sack of arguments $\mathcal{A}_i \subset \mathcal{A}$ whose content reflects her relative position, o_i, on $\mathbb{P}$. Thus, agents possess partial knowledge on the relationships between the arguments in $\mathcal{A}$. If $a \in \mathcal{A}_i$, then agent i knows which arguments are in conflict with a. Each argument $a \in \mathcal{A}$ is given 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 "argument a is totally incoherent with the principle $\mathbb{P}$".

Agents have an incentive to deliberate because they know that deliberation is an opportunity to either support or undermine a proposal that opposes their position on $\mathbb{P}$. They may present two types of behavior, *naive* and *focused*. Naive agents will only use deliberation to voice their opinions on the principle. Focused agents strategically argue in favor of proposal arguments that support the principle they favor, thus using all the information they have on the relationship between arguments. All agents (1) are able to assess the degree of support for $\mathbb{P}$ of all arguments, (2) agree on the existence of a conflict between any two arguments if such is announced during deliberation, and (3) are sincere when communicating their positions to each other. At time t, an agent i votes favorably for a proposal $\mathcal{P}$ of justification argument I if and only if $v_I(t) \times o_i(t) \geq 0$.

A dynamics for informal discussion At each informal discussion time step t, every agent i randomly meets one other agent j and updates her opinion according to the following dynamic equation:

$$o_i(t+1) = \begin{cases} o_i(t) + \mu(o_j(t) - o_i(t)) & if \quad |o_i(t) - o_j(t)| < U_i \\ o_i(t) + \mu(o_i(t) - o_j(t)) & if \quad |o_i(t) - o_j(t)| > T_i \\ o_i(t) & otherwise \end{cases} \tag{1}$$

where the parameter $\mu \in [0, \frac{1}{2}]$ controls for the strength of attraction and repulsion in social influence and (T_i, U_i) is the couple of latitudes of rejection and acceptance for informational cues of agent i.

The meeting and updating of opinions in this situation are loosely associated to Moscovici and Doise's warm discussion [20] and will be denominated the *warm discussion* model.

Fig. 2: Argumentation framework $AF = (\mathcal{A}, \mathcal{R})$ with $\mathcal{A} = \{a, b, c\}$ and $\mathcal{R} = \{(a, b), (b, c), (c, I)\}$. The labeling $\{\{c\}, \{I\}, \{a, b\}\}$ is conflict-free, a and b are undecided, c is accepted and I is rejected. $\{\{a, c\}, \{b, I\}, \emptyset\}$ is the only complete labeling obtained from the framework.

2.2 Abstract argumentation and deliberative models for collective decision-making

Deliberation, defined as an exchange of arguments, may be modeled by confronting, eventually contending, arguments. Following Dung's abstract argumentation theory [12], let $\mathcal{A}$ be a finite set of arguments and $\mathcal{R}$ a subset of $\mathcal{A} \times \mathcal{A}$ called *attack relation*. $(a, b) \in \mathcal{R}$ stands for the fact that argument $\mathcal{A}$ *attacks* argument b, meaning that argument a is in conflict with argument b. One says that an argument c *defends* an argument a if there exists b such that $(c, b) \in \mathcal{R}$ and $(b, a) \in \mathcal{R}$. One names *argumentation framework* (AF) the couple $(\mathcal{A}, \mathcal{R})$ composed of a set of arguments and their attack relation, which can be seen as a digraph in which the nodes are the arguments and the arcs are the attacks. A *label* $\mathcal{L}ab(a) \in \{\mathbf{IN}, \mathbf{OUT}, \mathbf{UND}\}$ of an argument $a \in \mathcal{A}$ denotes the acceptability status of a in a deliberation process. Intuitively, an argument is labeled **IN** if it is acceptable, **OUT** if it is not and **UND**, if nor **IN** nor **OUT** labels are applicable. Moreover, one defines a *labeling* on an argumentation framework $AF = (\mathcal{A}, \mathcal{R})$ as a complete function $\mathcal{L} : \mathcal{A} \to \Lambda = \{\mathbf{IN}, \mathbf{OUT}, \mathbf{UND}\}$, $a \mapsto \mathcal{L}ab(a)$ that assigns a label to each argument in AF. A labeling-based *semantics* is a set of criteria that yields acceptable labelings. For example, if an argument a attacks an argument b, then an acceptable labeling should not assign the label **IN** to both arguments. Basic semantics demand labelings to be *conflict-free*, meaning that no two arguments that attack each other are labeled **IN**, or *admissible*, implying that the labeling is conflict-free and that for any **IN** labeled argument a, there exists another **IN** labeled argument c such that c defends (or reinstates) a.

The family of admissibility-based labelings goes from *complete* labellings, which are admissible labelings for which all labels (including the undecided) are justified [5], to *preferred* and *grounded* labellings which are complete labellings obtained from, respectively, maximizing and minimizing the number of arguments that are labeled **IN**. They capture properties such as credulity and skepticism in argumentation. For a more extensive account of semantics and labellings, refer to [5].

The reason why we incur to abstract argumentation is technical in nature. The theory provides a comprehensive formalism that bypasses difficulties related to the nature and construction of arguments. The formalism also lends itself well to graph theory and to model (collective) reasoning in a clear, coherent and easy way [29]. Given an argumentation framework, Dung's extension-based [12] approach is only interested in the set of acceptable arguments (according to a certain semantics). The labelling approach [7] assigns a label to each argument in the framework. Hence, this approach is more expressive since it distinguishes arguments that are not accepted from those that are undecided. Such distinction is crucial since the existence of undecided arguments is one of the reasons why deliberation takes place and carries on over time. This is the primary justification for using labeling-based semantics in our model. Figure 2 provides an example of an argumentation framework that models one "step" of deliberation over a proposal justified by an argument I.

Deliberative collective decision-making protocol. Debates take place on a table in which a *central authority* (CA) [6] fixes the deliberation procedure. The CA chooses the percentage of agents

(n_D) from the population that may actively participate in the deliberation, the labeling-based semantics (σ) used to assess the label of the proposal argument, and the number ($\underline{m}$) of debates that ought to take place before a decision is deemed sufficiently discussed. Additionally, it also controls the maximum number ($\overline{m}$) of debates that can take place before abandoning deliberation, and the number (t_D) of informal discussion steps between debates. The CA also decides which collective decision rules to apply during the process (e.g. whether there is voting on proposals) and the proportion (α) of favorable votes in the population necessary to accept a proposal. Given a proposal $\mathcal{P}$, the deliberation or debate protocol goes as follows:

1. The CA generates and makes public a central argument or proposal argument $I \notin \mathcal{A}$;
2. The CA randomly draws two sets of $\frac{n_D}{2} \times n$ agents with divergent views on $\mathcal{P}$;
3. Each agent advances an argument from her sack $\mathcal{A}_i$. The CA makes sure that there are no repeated arguments with respect to previous debates on the same proposal (tables have memory);
4. The CA builds the debate's argumentation framework on the previously held debates over the proposal. It computes a labeling for the arguments using the semantics σ;
5. If the obtained label for I is undecided ($\mathcal{L}ab_d(I) = \mathbf{UND}$) or the number of debates steps held in the decision process 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 $\mathcal{L}ab(I)$ be the final label given to the proposal argument I. If voting is not part of the process ($\alpha = 0$), $\mathcal{L}ab_d(I) = \mathcal{L}ab(I)$; otherwise if more than $\alpha \times n$ agents agree with I, I is accepted ($\mathcal{L}ab(I) = \mathbf{IN}$), refused if strictly less than $\alpha \times n$ agents agree with it ($\mathcal{L}ab(I) = \mathbf{OUT}$). If there is a tie $\mathcal{L}ab_d(I) = \mathbf{UND} \Rightarrow \mathcal{L}ab(I) = \mathbf{OUT}$ and $\mathcal{L}ab_d(I) = \mathbf{IN} \Rightarrow \mathcal{L}ab(I) = \mathbf{IN}$.

Notice that deliberation always ends: either agents debate and agree on the proposal's acceptability through procedural argumentation or, after $\overline{m}$ debate steps, they directly vote on it. Also, observe that voting for a proposal is the same as voting for the argument that justifies it.

2.3 Linking deliberation and informal discussion through opinions

Let $\mathcal{P}$ 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 distance $\delta_i(t) = \frac{1}{2}|v_I(t) - o_i(t)|$ and the acceptability status $\mathcal{L}ab(I)$ of I at the end of a decision process over $\mathcal{P}$, agent i updates her opinion as follows:

$$o_i(t+1) = \begin{cases} o_i(t) + \gamma(v_I(t) - o_i(t)) & if\ \ \mathcal{L}ab(I) = \mathbf{IN},\ \textit{with probability}\ p_a^{\delta_i(t)} \\ o_i(t) + \gamma(o_i(t) - v_I(t)) & if\ \ \mathcal{L}ab(I) = \mathbf{IN},\ \textit{with probability}\ p_r^{\frac{1}{\delta_i(t)}} \\ o_i(t) & if\ \ \mathcal{L}ab(I) \neq \mathbf{IN},\ \textit{with probability}\ 1 - p_r^{\frac{1}{\delta_i(t)}} - p_a^{\delta_i(t)} \end{cases}, \quad (2)$$

where $\gamma \in [0, \frac{1}{2}]$ is the strength of repulsion and attraction in the dynamics. p_a and p_r are probability parameters that control for the possibility that an agent is attracted to and repulsed from agreements reached during debates. The equation combines the probabilistic nature of the effect of deliberation based on a principle similar to the one in social judgment theory [27], be it a moderating [20,13] or polarizing [28] one. It follows that deliberated informational cues may potentially influence any agent in the group. We call the model in which agents only update their opinions by Equation 2 the *cold discussion* model, as we associate it to Moscovici and Doise's [20] cold discussion. We call the *mixed discussion* model the model defined by Equations 1, 2 and the decision-making protocol.

2.4 Simulations

A time step in the model corresponds to either a debate, a step of dyadic social influence or a vote that makes agents update their opinions[1]. Simulations stop once 100 decision-making processes over 100 randomly generated proposals terminate. We observe how deliberation affects opinion distributions, coherence between majority voting and deliberative results, and judgment accuracy taking as reference the warm and cold discussion models.

Observations. At the end of each simulation ($t = S$), we observe the following metrics:

- **Variance of opinions** ($Var(o)$): the variance of opinions at time S. The higher the variance of the distribution, the more "diverse" opinions are in the opinion pool;
- **Shift in opinions** (Sh) [9]: statistic that measures the aggregated change in individual opinion at time S with respect to time 0, $Sh = \frac{2\sum_{i\in N}|o_i(0)-o_i(S)|}{max_{i\in N}o_i(0)-min_{i\in N}o_i(0)}$;
- **Proportion of extremists in the population** ($prop_{ex}$): percentage (%) or proportion of agents in the population with non-moderate opinions (i.e. $|o_i(S)| \geq 0.75$);
- **Judgment or consensual inaccuracy** (ec): it consists of an ad hoc statistic measuring a group's ability to infer correct labels for proposal arguments. Correct labels are obtained from the argumentation framework AF_ε that contains all arguments and their attacks. Let $\mathcal{I}$ be the set of all discussed proposal arguments up to S and $\mathcal{L}ab_\varepsilon(I)$ the label given to I in AF_ε. We use a Hamming-based distance on labellings as introduced in [2] to explicitly define the statistic: $ec = \frac{1}{|\mathcal{I}|}\sum_{I\in\mathcal{I}} a_I|\,\mathcal{L}ab_\varepsilon(I) \neq \mathcal{L}ab(I)|$, where $a_I = \frac{1}{2}$ if $\mathcal{L}ab_\varepsilon(I) = \textbf{UND}$ or $\mathcal{L}ab(I) = \textbf{UND}$ and $a_I = 1$, otherwise;
- **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: $ir = \frac{|\{I\in\mathcal{I}\,|\,\mathcal{L}ab_d(I)=\textbf{IN},\mathcal{L}ab(I)=\textbf{IN}\}|}{|\{I\in\mathcal{I}\,|\,\mathcal{L}ab_d(I)=\textbf{IN}\}|}$.

Initialization. All agents start off with an opinion o_i drawn from a uniform distribution $\mathcal{U}(-1,1)$. For all agent i, we set $(U_i, T_i) = (U, T)$ for some $(U, T) \in]0, 2[\times]0, 2[$, μ to 0.1 and p_r to 0.05. Given o_i, agents randomly draw a set $\mathcal{A}_i$ ($|\mathcal{A}_i| = k$) of arguments from a balanced[2] argument pool $\mathcal{A}$ of $m = 600$ non-neutral arguments on the basis of o_i. Each argument $a \in \mathcal{A}$ is given a level of support for the principle $\mathbb{P}$, v_a, obtained from a uniform distribution $\mathcal{U}(-1,1)$. The attack relation $\mathcal{R}$ that gives birth to the consensual argumentation framework AF_ε is established according to the v_as and is given a permanent labeling $\mathcal{L}^\sigma_\varepsilon$ computed using σ = grounded semantics. On the proposal side, we create an argument $I \notin \mathcal{A}$ whose support for $\mathbb{P}$ is also drawn from a uniform distribution $\mathcal{U}(-1,1)$, and is given the label $\mathcal{L}ab(I) = \textbf{UND}$. We allow I to attack no argument, yet allow other arguments to randomly attack it. Finally, we set the maximum number of debates to $\overline{m} = 7$ and following [18], we set the number of agents in the model to 400.

3 Simulation results

We obtain two kinds of results. The first is global and answers the question on how deliberation affects opinion formation. It consists of the comparison between the warm (Eq. 1), cold (Eq. 2),

[1] In warm discussion, agents vote for the proposal arguments, but do not update their opinions.
[2] By balanced we mean with as many arguments with $v_a < 0$ as with $v_b > 0$.

Length of $\mathcal{D}_I$		**Decisional in D_I**		**Social influence**		**Deliberation influence**		**Cognitive**	
t_D	$\underline{m}$	n_D	α	T	U	p_a	γ	k	$focused$
{1,3,6}	{1,3,6}	{0.01,0.02,0.05}	$\{0, \frac{1}{2}, \frac{2}{3}\}$	{1.4,1.8}	{0.2,0.6}	{0.1,0.3,0.5}	{0.05,0.1,0.2}	{4,8,16}	$\{True, False\}$

Table 1: Multimodal parameter domains used to compare warm, mixed, and cold discussion.

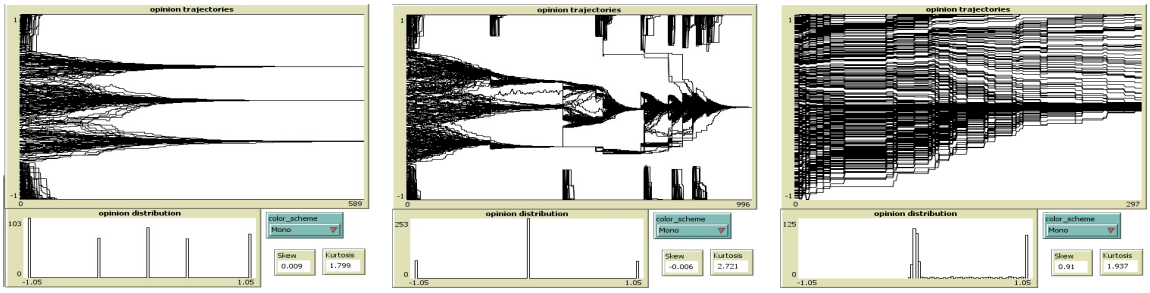


Fig. 3: From left to right, opinion trajectories and distributions for warm, mixed, and cold discussion.

Model \ Metric	$Var(o)$	$prop_{ex}$	Sh	ec	ir
Warm vs. Cold discussion	[0.191,0.216]	[0.170,0.197]	[-29.48,-28.28]	[0.074,0.082]	[0.165,0.172]
Mixed vs. Cold discussion	[0.142,0.147]	[0.150,0.155]	[-20.85,-19.76]	[0.003, 0.002]	[-0.038,-0.030]
Mixed vs. Warm discussion	[-0.072,-0.046]	[-0.045,-0.018]	[8.332, 8.859]	[-0.079,-0.071]	[-0.205,-0.201]

Table 2: Mean difference 0.95 confidence intervals for metrics by model comparison.

and mixed (Eq. 1, Eq. 2 w. deliberation protocol) discussion models. We simulate from 10 to 30 runs for each model and scenarii on the parameter space induced by the initialization and Table 1.

The second kind of results consists of a sensitivity analysis. It addresses the questions regarding the importance of procedural deliberation parameters, namely n_D, t_D, α, and $\underline{m}$, and agent behavior on our metrics in the mixed discussion model. We span their domain as described in Table 3, and generate 36,000 observations. Simulations and analyses are performed in Netlogo 6.0.4. and R 3.2.3.

Comparing the different models. From the simulations, we observe that the variance of opinions and the proportion of extremists are strongly correlated ($\rho \approx 0.95$, $p < 0.001$). We infer that cold discussion favors judgment accuracy, reduces the variance of opinions and the proportion of extremists. Although there is opinion polarization, only one group of extremists forms, probably the one in favor of the first deliberated results (see Fig. 3). Otherwise, a moderate consensus around neutrality forms. Warm discussion, on the other hand, is responsible for an increase in the variance of opinions and in the number of extremists. Coherence is maximal since agents only vote for proposals. Interestingly, we see that the mixed model is a compromise of the warm and cold discussion models. Deliberation not only contributes to obtaining correct answers but also to a slight decrease in the variance of opinion and in the proportion of extremists. However, it does not do better than the cold discussion model on coherence and produces less shifts of opinion.

Procedural parameters of interest					**Other parameters**						
t_D	$\underline{m}$	n_D	α	$focused$	$\overline{m}$	T	U	p_a	p_r	γ	k
{1,2,...,6}	{1,2,...5}	{0.01,0.02,...,0.05}	$\{0, \frac{1}{2}, \frac{2}{3}\}$	$\{True, False\}$	{7}	{1.6}	{0.2}	{0.3}	{0.05}	{0.2}	{12}

Table 3: Domains and types for procedural and behavior parameters in sensitivity analysis.

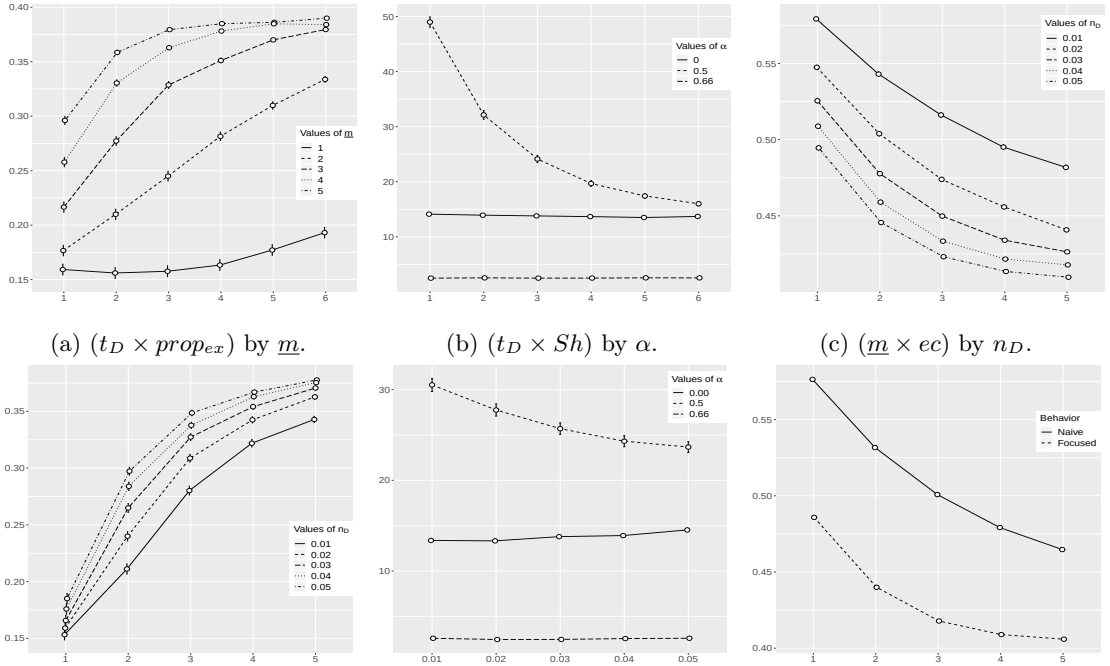
A first reading of the result concludes that there is a trade-off between judgment accuracy and variance in opinion which may be interesting to explore. More accuracy is related to slightly less extremism, which points to the fact that extremism may not contribute to successful deliberation.

Minimum number of debates ($\underline{m}$). We observe that minimum number of debates has a significant, well-observed effect on all of our metrics excluding *coherence* (ir). Taking variance of opinions, we notice that the more debates there are, the bigger the value of the metric is, and the higher n_D and t_D are, the weaker is the overall effect (Fig. 4d and Fig. 4a). Moreover, the marginal increase of the minimal number of debates on the variance of opinions is decreasing. In contrast, shifts in opinion are less and less likely as $\underline{m}$ grows and this independently of other parameters. Again, the effect is marginally decreasing and is only truly significant when $\alpha = \frac{1}{2}$. An explanation of these effects may be linked to the design of the system. First, variance of opinions are higher when deliberation is asked for because the more deliberation steps there are the higher the chances that the proposal argument is deemed unacceptable. 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 an argument is considered undecided, t_D time steps are added to the process. So, unless the debate yields decisive labels for proposal arguments (less likely considering that σ = grounded semantics), more non-deliberation steps take place in the decision process and the higher the variance of opinions will be. Concerning shifts, when $\alpha \neq \frac{1}{2}$, 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 the process (Fig. 4b).

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 more linear; when they are focused, the strongest effects of adding more deliberation are found when levels of deliberation are already low (Fig. 4f). This can be explained by the fact that the more debates there are in the decision process, the closer one gets to the consensual argumentation framework. The effect is stronger for the focused agents because, when reconstructing the framework, they take into account the deliberated proposal and advance the most pertinent arguments they have.

Proportion of the population in deliberation steps (n_D). Like with $\underline{m}$, n_D has a significant effect on the proportion of extremists and on the variance and shift of opinions (Fig. 4e). This may result from the fact that being able to put more arguments in play at the same debate step can increase the chances of revealing the cycles around the proposal argument. Given that we use grounded semantics, the arguments in the cycles are labeled **UND** postponing debates more often than if n_D was lower. Postponing debates, in turn, increases the number of informal interactions in the decision process, which increases the variance in opinion and limits the effect of deliberation.

Moreover, the effect of this parameter is very dependent on the value of α (Fig. 4e). For shifts, for instance, $\alpha = \frac{1}{2}$ makes the effect of n_D negative, while $\alpha = 0$ makes it positive to a lesser



(a) $(t_D \times prop_{ex})$ by $\underline{m}$. (b) $(t_D \times Sh)$ by α. (c) $(\underline{m} \times ec)$ by n_D.

(d) $(\underline{m} \times prop_{ex})$ by n_D. (e) $(n_D \times Sh)$ by α. (f) $(\underline{m} \times ec)$ by $focused$.

Fig. 4: Curves of mean observations for 36,000 runs on metrics (0.95 confidence intervals).

degree. For higher requirements for 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 whether they are focused or not. This is a surprising result as one would have expected more focused agents in an arena to heavily impact the proportion of extremists. They play to knock out opposing proposal arguments and thereof hinder the opinion-moderating effects of deliberation. Finally, similar to $\underline{m}$, adding more people into the deliberation process increases judgment accuracy (Fig. 4c) and has no effect on coherence (ir).

Steps between deliberation steps (t_D). In all configurations, t_D increases the proportion of extremists (variance) and decrease the shifts in opinion. The shifts and the effect on the variance of opinions are only observable for $\alpha = \frac{1}{2}$ (Fig. 4b). 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 $\underline{m}$ increases, the curve becomes more and more concave, which means that t_D has a more important effect on the opinion distribution as collective decision-making processes are longer. This seems counter-intuitive yet it reflects the multiplicative relationship between deliberation and pair-wise interactions. If $\underline{m}$ is low, and t_D high, the effective number of pair-wise interactions are, on average, fewer in the deliberation process, which limits the increase of the variance in opinions. Additionally, since the grounded semantics yields few **IN** arguments w.r.t. other admissibility-based semantics, getting closer to the consensual argumentation framework may lessen the number of opinion updates due to deliberation. Last, a lower $\underline{m}$ makes deliberation more influential on opinions.

Acceptability voting quota (α). By far, the most influential parameter in our study. It changes the direction and the intensity of the effect of other procedural parameters and, by construction, heavily constrains the road to accepting a proposal. In few words, α constrains the world to warm discussion, or throws it into a process in which cold discussion is a lot more important. One either gives too much weight to deliberated results ($\alpha = 0$) and the effect of pair-wise interactions becomes negligible, or too much weight to pair-wise interaction ($\alpha = \frac{2}{3}$). It follows that updates due to deliberation happen rarely and opinions do not moderate. Concerning labeling-based metrics, α entirely determines the coherence statistic. For $\alpha \neq \frac{1}{2}$, there is no difference in coherence, because of how coherence is defined: it is maximal when $\alpha = 0$ and strangely maximal when $\alpha = \frac{2}{3}$. The reason for the latter is that latitudes of acceptance (rejection) are too low (high) and thus pair-wise interactions are unable to unevenly polarize the population in such way that deliberated proposal arguments are accepted by the two-third majority.

4 Related Work

We see our model as a contribution to the influence and opinion dynamics field in agent-based modeling (ABM) and a pragmatic application of abstract argumentation theory. To our knowledge, we are unaware of existing literature on ABM that explicitly relates collective decision-making, deliberation by abstract argumentation, and opinion diffusion as we have done it. This said, many models in the literature of opinion diffusion are interested in opinions because they influence collective decisions and can be used to reveal certain types of social phenomena. For instance, in [15] the authors are interested in consensus and in how a group collectively decides on an action when it is given two alternatives. In other models, authors are interested in the emergence of extremism [19] and on the distribution of opinions when extremists are introduced in the population [10], while other authors coin the notion of opinion polarization as an emergent property [19]. They show, using models of "bounded confidence" and opinion diffusion with trust, that three different kinds of steady states (unipolar, bipolar and central) were possible depending on whether agents were sufficiently uncertain about their opinions, sufficiently connected, and/or a certain proportion of individuals were already extreme. Similarly, work on collective cognitive convergence and opinion sharing [22] show that consensus towards a certain opinion or cognitive state is always possible yet dependent on noise, variability and awareness of agents. Closer to opinion formation and argumentation, authors in [14] define an agent's opinion as a function of the arguments she holds and their relationship (logical). They device a peer-to-peer dialog system (NetArg) that uses only abstract argumentation to study opinion polarization and opinion dynamics. When it comes to abstract argumentation theory, we take an approach that wires two type of dialogues that are well-studied in the literature: persuasion dialogues [21] and deliberation dialogues [1]. The line of work that might be closest to ours is the one on mechanism design [23], or the problem of devising an argumentation protocol where strategic argumentation has no benefit. We tackle mechanism design in a different way. Instead of considering strategy-proofness, we are interested in how differences in protocol can result in "better" collective choices and guarantee that opinion distributions are favorable for deliberation ("reasonable" level of variance). For a survey on persuasion dialogue, see [21].

Work on agent-based argumentation usually assumes that the semantic relationship between arguments is fixed [23,24]. Other models which do not make this restrictive assumption can also be found in the literature and derive from the class or family of opponent models [6] in which two opposing sides attempt to win the dialogue. Our model is in the intersection of these, but the framework that combines opinion diffusion of the kind and argumentation seems original.

The idea of mixing interpersonal influence and vertical communication is not new. For instance, [10] and [11] describe and implement such ideas in innovation diffusion. In both cases, vertical communication is modeled as exogenous information. The originality of our work is in that 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 [4], where the authors control for the design of vertical communication, we control for the process generating vertical information.

5 Conclusion

The main objective of this article is to build a bridge in which decision-making, argumentation, and opinion diffusion can meet. We propose a model that combines abstract argumentation theory and a bounded confidence opinion diffusion model and showed to what extent it could explain variability in opinion and correctness of collective decisions. The model reveals that (1) to ask for more deliberation, (2) to allow for more agents to participate in deliberative instances, and (3) to make deliberative interactions less frequent in time guarantees an increase in the variance of opinion and in the proportion of extremists in a group. These results are consistent with results found in [18] and in [17,28], which stress that deliberation may polarize groups and may have a meager effect on shifts in opinion; and inconsistent with [13] where it is argued that deliberation moderates opinions. Deliberation alone does moderate opinion as noted in [20] yet, when integrated into a complex system in which individuals are allowed to interact with one another, its influence is overshadowed by other individual-based dynamics. Undeniably, grounded semantics play an important role in the weakness of the effect of deliberation since it models a skeptical way of reasoning over arguments. Although accepting an argument happens less often, deliberation still increases judgment accuracy in a marginally decreasing fashion. We show that voting within the deliberation protocol not only increases the proportion of extremists and the variance of opinions but also determines how coherent deliberation and voting are with one another. The most influential parameter found was the voting quota for proposal acceptability because it determined which part of the mixed model (deliberation if small, pair-wise influence if big) dominated the dynamics.

Extensions of this model include better-thought deliberation protocols where one may consider deliberation as having only an impact on people that actually debate, and where observing how deliberation changes opinion distributions equates to observing how it spreads within a group. Taking into account trust, network effects, multi-dimensionality in opinions, new processes of argument exchange or learning are ways to further extend the model and relax unrealistic assumptions. To conclude, exploring different argumentation ontologies and opinion dynamics and finding case studies to apply the model are essential points to build on in future work.

References

1. 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*, 11(2):153–171, 2005.
2. E. Awad, M. Caminada, G. Pigozzi, M. Podlaszewski, and I. Rahwan. Pareto optimality and strategy-proofness in group argument evaluation. *Journal of Logic and Computation*, 27(8):2581–2609, 2017.
3. R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.
4. V. Barbet and J. Rouchier. Tension between stability and representativeness in democracy. 2018.

5. P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The knowledge engineering review*, 26(4):365–410, 2011.
6. E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Vol.1*, pages 47–54. International Foundation for Autonomous Agents and Multiagent Systems, 2011.
7. M. Caminada. On the issue of reinstatement in argumentation. In *European Workshop on Logics in Artificial Intelligence*, pages 111–123. Springer, 2006.
8. S. Chambers. Deliberative democratic theory. *Annual review of political science*, 6(1):307–326, 2003.
9. 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*, 5(4), 2002.
10. G. Deffuant, S. Huet, and F. Amblard. An individual-based model of innovation diffusion mixing social value and individual benefit. *American Journal of Sociology*, 110(4):1041–1069, 2005.
11. 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*, 60(8):826–835, 2007.
12. P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.
13. J. Fishkin and R. Luskin. Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta politica*, 40(3):284–298, 2005.
14. S. Gabbriellini and Paolo Torroni. A new framework for ABMs based on argumentative reasoning. In *Advances in Social Simulation*, volume 229, pages 25–36. Springer Berlin Heidelberg, 2014.
15. S. Galam and S. Moscovici. Towards a theory of collective phenomena: Consensus and attitude changes in groups. 21(1):49–74, 1991.
16. R. Glinton, O. Scerri, and K. Sycara. An investigation of the vulnerabilities of scale invariant dynamics in large teams. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Vol.2*, pages 677–684. International Foundation for Autonomous Agents and Multiagent Systems, 2011.
17. K. Hansen. *Deliberative democracy and opinion formation*. University Press of Denmark Odense, 2004.
18. W. Jager and F. Amblard. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational & Mathematical Organization Theory*, 10(4):295–303, 2005.
19. M. Meadows and D. Cliff. The relative disagreement model of opinion dynamics: Where do extremists come from? In *International Workshop on Self-Organizing Systems*, pages 66–77. Springer, 2013.
20. S. Moscovici and W. Doise. *Dissensions et consensus: une théorie générale des décisions collectives*. Presses Universitaires de France-PUF, 1992.
21. H. Prakken. Systems for persuasion dialogue. *The knowledge engineering review*, 21(2):163–188, 2006.
22. O. Pryymak, A. Rogers, and N. Jennings. Efficient opinion sharing in large decentralised teams. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Vol.1*, pages 543–550. International Foundation for Autonomous Agents and Multiagent Systems, 2012.
23. I. Rahwan and K. Larson. Mechanism design for abstract argumentation. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Vol.2*, pages 1031–1038. International Foundation for Autonomous Agents and Multiagent Systems, 2008.
24. T. Rienstra, M. Thimm, and N. Oren. Opponent models with uncertainty for strategic argumentation. In *IJCAI*, pages 332–338, 2013.
25. J. Rouchier and E. Tanimura. Learning with communication barriers due to overconfidence. *Journal of artificial societies and social simulation*, 19(2), 2016.
26. D. Rousseau. Reinforcing the micro/macro bridge:organizational thinking and pluralistic vehicles. 2011.
27. M. Sherif and C. Hovland. Social judgment: Assimilation and contrast effects in communication and attitude change. 1961.
28. C. Sunstein. The law of group polarization. *Journal of political philosophy*, 10(2):175–195, 2002.
29. G. Vreeswijk. An algorithm to compute minimally grounded and admissible defence sets in argument systems. In *COMMA*, pages 109–120, 2006.
30. Y. Zhang, H. Duan, and Z. Geng. Evolutionary mechanism of frangibility in social consensus system based on negative emotions spread. *Complexity*, 2017, 2017.