

# Description of the model using the ODD protocol

immediate

October 16, 2023

The model description follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models [Gri+06], as updated by Grimm et al. 2020.

## 1 Purpose and patterns

Recent advances in cognitive sciences include the introduction of “myside bias”, a specific form of confirmation bias, in a series of publications by Mercier, Sperber and colleagues ([Mer17; MH14; MS17]). They define myside bias as a tendency of individuals to predominantly produce reasons in support of their own claims, opinions and beliefs while neglecting (the search for) defeating reasons against their own claims. Myside bias is construed as an adaptive feature of reasoning that affects various contexts and has a harmful effect on the epistemic abilities of individuals and communities (e.g. individuals and communities are less likely to approach the truth). According to the authors, scientific reasoning is not exempt from myside bias. They specifically study scientific communities as examples of successful epistemic communities, where myside bias is kept in check, and hypothesize that specific socio-epistemic mechanisms may mitigate it.

The purpose of our model is twofold : first, to simulate a community of scientists choosing between alternative research programs, who exhibit a form of myside bias similar to the one described by Mercier and Heinz [MH14]. Secondly, to try to determine mechanisms which may mitigate a harmful effect of this bias. In particular, we test their hypotheses that social mechanisms such as a form of shared beliefs, understood as shared standards of evaluation to filter arguments helps to reduce the harmful effect of the bias in communities. Since our model aims at testing a theoretical hypothesis, its purpose fits best in the *theoretical exposition* category of [Edm+19]. In our model, we are concerned with describing a community looking for the truth: one research program is considered objectively better than the other, and thus, we denote *epistemic success* the agent’s ability to choose the best alternative. A higher purpose of this work is to show that abstract argumentation is a promising tool for the study of epistemic communities.

We will test three patterns. The first one allows us to validate a baseline model of the scientific community. The second and third one test Mercier and Heinz’ hypotheses about the effect of myside bias and shared standards of evaluation on epistemic communities.

**Pattern 1:** Unbiased communities are good at finding the best alternative.

**Pattern 2:** Myside bias is harmful for the agent’s epistemic success.

**Pattern 3:** Shared beliefs mitigate the harmful effect of the bias.

Here, being good at finding the best alternative means that all or a majority of the agents of the community prefer the superior research program. A harmful effect would manifest as a lower number of agents selecting the best alternative in a biased community than in an unbiased one, all other parameters being equal. Shared beliefs are understood as shared epistemic standards, or methods, which allow a community to filter out weak arguments.

## 2 Entities, state variables, and scales

### 2.1 Entities

The entities in the model are the environment, which we call the *observer*, and the agents. In order to define these entities, we have to introduce the objects used in our model.

**Definition 1** (Argumentation framework [Dun95]). An **argumentation framework** (AF) is a pair  $\langle \mathcal{A}, \mathcal{R} \rangle$  where  $\mathcal{A}$  is a finite and non-empty set of (abstract) arguments, and  $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$  is a binary relation on  $\mathcal{A}$  called the attack relation. For two arguments  $a, b \in \mathcal{A}$  we say that  $a$  attacks  $b$  in case  $(a, b) \in \mathcal{R}$ .

**Definition 2** (Research Program). A **research program** (*ResProg*) is a set of arguments of size  $N_{CA} (\geq 1)$ .

**Definition 3** (View of Debate). The view of the debate  $V_i^t$  of agent  $i$  at step  $t$  is an argumentation framework  $\langle \mathcal{A}_i^t, \mathcal{R}_i^t \rangle$  such that for any  $t, i$ :  $ResProg_1, ResProg_2 \subset \mathcal{A}_i^t$ .

Our model represents scientists debating and producing arguments about two alternative research programs. Thus, our model only contains two kinds of entities: the observer which describes the environment of the scientists and keeps track of the model’s data, and the agents themselves.

The **observer** is the entity which keeps track of the model’s state: the observer is characterized by two research programs, combined in a view.

Research programs represent scientific theories and are characterized by a fixed number of central arguments. The view is composed of these central arguments and of arguments which are generated by the agents during the simulation. Due to the way the views are populated, each central argument is the root of an in-tree, that is a tree where all edges point towards the root. Each argument  $a \in A_i$  belongs to one and only one of the in-trees that can be found in the framework, that is every non-central argument can be linked by a path with one and only one of the central arguments. In each instantiation of the model two research programs are created, but different versions (“views”) are considered by the agents, who may be aware only of a subset of the arguments created throughout the simulation.

**Agents** represent the members of the scientific community. There is only one type of agent in the model. Their state variables include their current preferred research program, their view of the debate and three variables which track the process of their investigation: their current state, their investigated argument and their current result. An agent’s view of the debate is an argumentation framework that contains among its arguments the central arguments of each research programs. An agent’s view of debate is populated only by the arguments that the agent is aware of.

An **argument** is an object characterized by a name and a strength.

### 2.2 State Variables

Table 1 and 2 respectively describe the state variables of the observer and agent entities. All of the state variables of the observer are dynamic. The only static state variables of an agent is its name.

### 2.3 Scales

As this model is abstract and highly idealized, the discrete time steps of the simulation are not meant to represent any actual time range. A time step in the simulation corresponds to agents choosing an argument, investigating it, or publishing their result. “Arguments” do not necessarily have realistic counterparts, but if we liken them to scientific papers, a time step in our model

Table 1: State variables of the Observer

<b>Variable name</b>	<b>Variable type</b>	<b>Description</b>
<i>general view</i>	View, dynamic	A view of the two research programs which contains all arguments generated by the agents.
<i>new arguments</i>	List of arguments, dynamic	A list storing the argument which have been generated at each turn.
<i>acceptable arguments</i>	List of arguments, dynamic	A list storing the arguments which have been deemed acceptable by the community at each turn.

Table 2: State variables of an agent

<b>Variable name</b>	<b>Variable type</b>	<b>Description</b>
<i>name</i>	Integer, static	Name identifying the agent
<i>view</i>	View, dynamic	View of the debate characterizing the agent's position.
<i>state</i>	FREE or SUCCESS	Specifies the current state of the agent.
<i>current argument</i>	Argument, dynamic	Stores the argument investigated by the agent.
<i>result</i>	Argument, dynamic	Stores the argument generated by the agent.
<i>preference</i>	Integer, {1,2}, dynamic	The research program preferred by the agent.

would be the time necessary to investigate a research question and publish the results. The average time varies across disciplines, the order of magnitude could be about a year. Of course, scientists sometimes investigate more than one question at the same time, a fact which is simplified in our model. Space, which is not a feature of the theories that we are modelling, is not represented in our model.

### 3 Process overview and scheduling

In our model, one time step is divided into four phases: **choice**, when the agents choose an argument to investigate, **study**, when they investigate it and possibly generate a counter-argument, **publish** when they add this counter-argument to their own view and the observer’s and **update** when all agents determine, collectively and individually, whether they add their peer’s arguments to their views. These phases do not apply to all agents at every turn: their *state* variable is used to determine which phase they enter. Here, we describe each phase in detail. The corresponding submodels will be detailed in Section 7. Figure 1 summarizes the link between every value of *state* and each phase.

1. **Choice phase:** All agents execute their “choice” submodel. They update their *preference* for a research program, and their *current argument* variable to a random chosen argument.
2. **Study phase:** All agents execute their “investigate” submodel. Then, each scientist proceeds as follows.
  - If the “investigate” submodel succeeds for them, they update their *state* to SUCCESS and *result* to the new argument generated.
  - Else, their *state* variable becomes FREE.
3. **Publish phase:** Only the agents whose *state* variable is SUCCESS execute their “publish” submodel (the other agents do nothing). This adds the new argument to their *view*, and the observer adds the new argument to its *general view* and *new arguments* list. These agents update their *state* to FREE and *result* and *current argument* to *None*.
4. **Update phase:**
  - The observer executes the “shared beliefs” submodel on the arguments of *new arguments*. This submodel returns a boolean value specifying whether each argument is accepted or rejected by the community.
  - For each accepted argument, the agents execute their “update” submodel, performing a probabilistic test of fixed success rate  $p_{see}$ , and adding the argument (and associated attack relation) to their *view*, or not.

This protocol aims at representing the research and publication process of scientists. We simplified it to the few elements that were of interest for our model purpose: we model a community of scientists who tackle an issue by producing a body of knowledge. That is why our agents first choose a related topic of study, and then investigate it in order to produce new research. The dynamic production of arguments was central to our purpose, as our work is grounded on the argumentative theory of reasoning of Mercier and Sperber [MS11], which states that people are biased when they *produce* new arguments. Finally, the publish and update phases allow us to study not only individuals but interconnected communities whose members influence each other. This allows us to

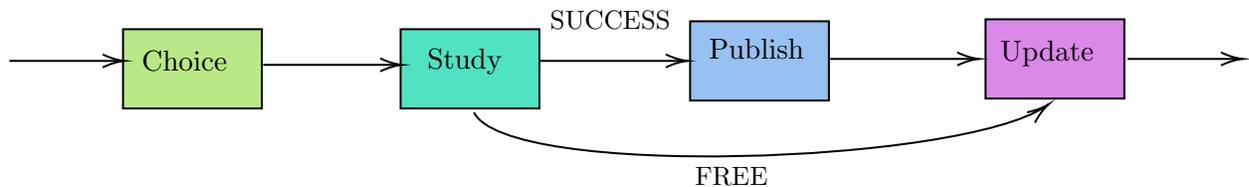


Figure 1: Process sequence and associated values of *state*.

assess whether our object of study, myside bias, behaves differently at the level of the community than at that of the individual, and whether it is mitigated by social mechanisms.

Our scheduler simulates the simultaneous activation of the agents: all of the agent’s actions are performed in a fixed order, but the agent’s actions and state updates are separated. This is to avoid artifacts of execution order, and also to account for the fact in real life, many scientists may be working on similar problems in parallel.

## 4 Design concepts

### 4.1 Basic principles

Our model aims at representing and studying a community of scientists who exhibit a cognitive bias: myside bias. This bias is described as a tendency to produce reasons in favor of one’s own theory rather than against [MH14]. Since reasoning is constituted by the production of arguments (or “reasons”) in Mercier and Sperber’s work [MS17], we investigate this issue in a model which explicitly represents an exchange of arguments. For this, we use tools from abstract argumentation theory. Introduced in the 90s, [Dun95], abstract argumentation aims at representing argumentative discussions as mathematical objects: the argumentation frameworks (also called graphs). In our model, argumentation frameworks represent the arguments produced and exchanged by the agents, thus equipping each agent with their own view on the current debate. This simple and flexible approach to argumentation is used in a growing number of agent-based models to investigate opinion dynamics [DBM22; BPR19; TST21].

Because we study a community of scientists, our model is part of an area of epistemology - the study of knowledge and justified beliefs - called computational social epistemology. This branch aims at studying how communities relate to the truth by using agent-based models and computer simulations. These simulations are useful to understand how social characteristics such as groups of opinion and interconnection can influence the pursuit of truth. They can shed light on phenomenons such as opinion polarization, explain the outcome of scientific controversies, and inform us on how to improve epistemic communities.

Our novel argumentative approach provides a new framework to model and study epistemic communities. It is well suited to explore phenomenons that are explicitly argumentative, e.g. the effect of myside bias as described by Mercier and Sperber. Additionally, this framework enables the assessment of the replicability of outcomes achieved through alternative models of epistemic communities. This contribution enhances the depth and quality of discussions pertaining to these models.

### 4.2 Emergence

We define the epistemic success of the community of agents as the outcome where a majority of the agents prefer the strongest research program at the end of the simulation. We study and compare

three types of communities: a baseline community, a community where agents exhibit myside bias, and a community where agents are biased and use a common mechanism to reject weak arguments, the “shared beliefs” mechanism. We observe five main results.

1. Agents almost never change their preference after about 100 steps.
2. Baseline communities (without bias and shared beliefs) are fully epistemically successful.
3. Bias reduces epistemic success, except when communities are equally distributed at the beginning.
4. The more biased a community is, the lower its polarization ratio.
5. Shared beliefs reduce the interval in which biased communities are not epistemically successful.

The processes of the system are complex enough that a complete mathematical analysis of its behavior is not realistic. As such, each of these phenomenon can be said to emerge from the mechanisms of the model. For example, result 1 was not implemented but rather derives from the construction of large argumentation graphs and the update process of the agents, which makes them less and less likely to add new arguments as time progresses.

### 4.3 Adaptation

Our model represents the way agents “choose” their preferred research program using an adaptive behavior which can be described as “direct objective seeking”. The objective of the agents is to choose the research program which is the most supported given the arguments that they are aware of. For this, they apply a special function called the grounded semantics to the argumentation graphs of their *view*, and compute a score for each research program. They choose the research program with the highest score. In case of equality, they have a 0.5 probability of choosing either one.

The agents exhibit three other adaptive behaviors, all of which are modeled as “indirect objective seeking” with stochastic rules.

The first one is argument generation: when investigating an argument, the agents have a certain probability of generating a counter-argument which attacks it. This probability depends on the strength of the investigated argument and, in order to represent the action of myside bias, on the “bias” parameter of the agent. The more biased an agent is, the lower the probability that they will generate a counter-argument to their preferred research program, and conversely, the higher the probability that they will generate an argument against the other research program.

The second adaptive behavior, the shared beliefs test, is performed at the level of the community. We represent the fact that a community of scientists shares some common methodology which allows them to reject certain arguments. We model this by a probabilistic test determining whether each new argument can be considered by the the other members of the community. This test’s chance of success increases with the argument’s strength; the shared belief parameter controls how discriminating it is.

The third adaptive behavior of the agents is the update of their *view* by adding the arguments generated by others. Agents only consider arguments which have been deemed acceptable by the shared beliefs test. An agent can add a new argument to their *view* only if their *view* contains the argument attacked by the new argument. This condition ensures that the new argument is relevant to the debate for the agent. If this condition is met, the agent performs a test with a fixed probability which decides whether they add the new argument. This probability, which we call

$p_{see}$ , is the same for all agents and corresponds to how likely they are to be aware of their peers work.

#### 4.4 Objectives

Here we describe in detail the function used by the agents to determine their preferred research program.

Each agent is equipped with a *view* composed of two research programs, which are both composed of the same number of central arguments and associated argumentation graphs. The grounded semantics is a function applied to abstract argumentation graphs which returns a set of arguments which are deemed “acceptable”. The idea is to provide a way to compute which arguments can be accepted by rational agents, given the attack relations between them.

**Definition 4** (Grounded set). Let  $\text{AF} = \langle \mathcal{A}, \mathcal{R} \rangle$  be an AF. An argument  $a \in \mathcal{A}$  is **acceptable with respect to a set of arguments**  $E$  in case for every  $b$  for which  $(b, a) \in \mathcal{R}$ , there is a  $c \in E$  for which  $(c, b) \in \mathcal{R}$ . The **grounded set**  $G$  for AF is iteratively defined as follows:  $G = \bigcup_{i=0}^{\infty} G_i$  where  $G_0$  is the set of all arguments which have no attackers in  $\mathcal{A}$  and  $G_{i+1}$  is the set of all arguments that are acceptable with respect to  $G_i$ . We call an argument **acceptable** if it is part of the grounded set.

We refer the reader to [Dun95] for further details about the grounded semantics. By applying the grounded semantics on each argumentation graph of their *view*, the agents can determine whether each central argument is part of the set of acceptable arguments. Then, they compute the score of the research programs by adding the number of their central arguments which are acceptable. This allows them to choose the research program which is the most rationally supported by the evidence that they are aware of. Thus, their preference reflects the dynamics of the generation of arguments, which may in turn reflect the intrinsic quality of each research program. This creates a link between agent’s knowledge and their epistemic success, which depends on whether the arguments that they are aware of support the best research program.

#### 4.5 Learning

Our model does not feature learning.

#### 4.6 Prediction

Our model does not contain prediction processes.

#### 4.7 Sensing

Agents are aware of their own *view*. In particular, they are aware of the arguments which they produce themselves. They are not aware of their own bias nor of the arguments’ strengths. They become aware of the arguments produced by others only if a certain number of conditions are met, as described in the “Adaptation” subsection.

#### 4.8 Interaction

Interaction between agents in our model takes the form of one-to-many transmission of information. Each agent sends her new arguments to all the other agents, who may or may not add it to their view. Our goal was to represent one way that scientists access information: not through dyadic interaction

but by being part of a public space where information is broadcast, potentially reaching everyone. One could think about the internet and its network of freely accessible journals, conferences and archives as an example of such a space. The probability that agents actually “see” the information represents their awareness of this space.

## 4.9 Stochasticity

Stochasticity is used to model three adaptive behaviors of the agents (see Section 4.3): the argument generation process which depends on a probability derived from the argument’s strength and agent’s preference and bias; the shared belief test, and the agents’ update process. By introducing a stochastic method we aim at providing a more realistic picture, which limits the agent’s rationality and creates a variety of different outcomes arising from the same initial conditions.

## 4.10 Collectives

No collective is explicitly represented as an entity in our model. The two groups of agents who support each research program can be seen as emergent collectives, but there is no action of these groups themselves nor coordination of the agents within.

## 4.11 Observation

The observed output of our model are the preferences of the agents. We define the support for each research program as the number of agents who prefer it after 100 steps <sup>1</sup>. Formally, the support at time  $t$  for a research program *ResProg*  $i$  is denoted  $n_i^t$ .

By comparing the average support for each research program to the difference between their strength, we define the notion of **epistemic success** of the community of agents. The community is epistemically successful if, on average, the support for the stronger research program is higher.

We also use the agent’s preferences to compute the **polarization ratio**. This is the ratio between the largest group of agents and the smallest group, i.e. at time  $t$

$$PolRat^t = \frac{\min(n_1^t, n_2^t)}{\max(n_1^t, n_2^t)}. \quad (1)$$

This allows us to track situations in which the population is divided between groups of similar sizes (high polarization ratio) to those situations of near or total consensus (low polarization ratio).

## 5 Initialization

The following parameters are initialized at the beginning of a simulation.

- $T$  the number of steps of the simulation.
- $N$  the number of agents.
- $n_1^0$  the number of agents who prefer the first research program at the beginning of the simulation.
- $n_C$  the number of central arguments in each research program.

---

<sup>1</sup>We stop our simulations after 100 steps because we found empirically that there was very little variation of the agent’s preferences after this threshold.

- $S_1$  and  $S_2$ , respectively the strengths of the central arguments of  $ResProg_1$  and  $ResProg_2$ .
- $p_{see} \in [0, 1]$ , the probability for agents to be aware of the arguments of others.
- $\sigma$  the standard deviation of the Gaussian distribution used to sample the strength of generated arguments.
- $bias \in [0, 1]$  the bias parameter.
- $shBel \in [0, 1]$  parameter which controls the degree of shared beliefs.

The initialization of our model starts with the creation of the two research programs which will compose the *general view* of the observer. The  $n_C$  central arguments are created in each research program  $i$ , receive a unique id and their corresponding strength  $S_i$ , and they initialize the tree structures which compose the research programs. Then, the  $N$  agents are created, and each is given a unique id; their *view* is composed of two research programs which are initialized with the same central arguments as the ones from the *general view*. Their *state* is “FREE”, *current argument* and *result* are empty. Their initial *preference* depends on a parameter specified by the user: the number of agents preferring  $ResProg_1$ , which we denote  $n_1^0$ . Precisely, the first  $n_1^0$  agents to be created are initialized with a *preference* for  $ResProg_1$ , the others for  $ResProg_2$ . Note that since the agents act in a way which mimics simultaneous activation, the order of creation has no impact on the outcome of the simulation.

## 6 Input data

The model does not use input data to represent time-varying processes.

## 7 Submodels

In this section we describe the five submodels mentioned in Section 3: “choice”, “investigate”, “publish”, “shared beliefs” and “update”.

### 7.1 Choice

The choice submodel of the agents consists of two steps: the first one is the determination of their preferred research program, and the second one is the choice of an argument to investigate.

**Preference determination:** The principle for the determination of the agent’s preferred research program has been described in the “Objectives” subsection. As explained above, the agents apply the grounded semantics function to determine whether each central argument of the research programs in their *view* is acceptable. This allows them to compute a score for each research program, equal to their number of acceptable central arguments. They prefer the *ResProg* with the highest score, and do not change their opinion in case of equality.

We use Algorithm 1 to determine whether the central argument of an argumentation framework is in the set of acceptable arguments. This algorithm is a modification of the algorithm to compute the grounded extension in trees presented by [MC09]. Our modification is faster on average because it does not compute the entire extension but stops once the central argument is assigned either to the extension or to its complement.

---

**Algorithm 1** Algorithm to determine whether the central argument of a CAAT is in the set of acceptable arguments

---

Let  $T$  be a CAAT, and  $CA$  its central argument. We call successors of an argument  $a \in T$  the arguments of  $T$  attacked by  $a$ , and predecessors the arguments of  $T$  which attack  $a$ .

```

 $G \leftarrow \text{leafs}(T)$ 
 $O \leftarrow \emptyset$ 
 $flag \leftarrow \text{True}$ 
while  $flag$  do
   $O \leftarrow O \cup \text{successors}(G)$ 
  if  $CA \in O$  then
    return  $\text{False}$ 
  end if
  for  $o \in O$  do
    for  $a \in \text{successors}(o)$  do
      if  $a \notin O$  and  $\text{predecessors}(a) \subset O$  then
        if  $a == CA$  then
          return  $\text{True}$ 
        end if
         $G \leftarrow G \cup \{a\}$ 
      end if
    end for
  end for
end while

```

---

**Choosing an argument:** The process for choosing an argument to investigate is very simple: the agents select one of the arguments contained in their *view*, with equal probability. Note that this includes the central arguments, but that the probability of investigating one of the central arguments decreases with time as the number of arguments in the agent's *view* increases.

## 7.2 Investigate

The agents investigate an argument by trying to generate an attack against it.

If successful, the new argument  $a'_i$  obtains a strength  $S(a'_i)$  drawn from a normal distribution centered on  $1 - S(a_i)$  with a standard deviation  $\sigma$  chosen by the user at initialization.

**Biased Community** In line with Mercier and Sperber's account [MH14], myside bias is modelled as a factor that makes it easier to *produce* arguments in favor of one's standpoint. Supposing two rival research programs, an argument in favor of one's preferred research program can either defend the program or attack the opposing one.

**Definition 5** (Indirect attackers and defenders). Let  $\langle C, \mathcal{A}, \mathcal{R} \rangle$  a CAAT with  $C$  its central argument. We say that  $a \in \mathcal{A}$  is an indirect attacker of  $C$  if the length of the path from  $a$  to  $C$  is odd (in the number of attack relations), and that  $a$  is an indirect defender of  $C$  if it attacks one of its indirect attackers, or equivalently if the length of the path from  $a$  to  $C$  is even.

**Definition 6** (Against - In favour of). An argument is said to be in favour of a research program  $ResProg_i$  and against the other  $ResProg_j$  if it is an indirect defender of one of the central arguments of  $ResProg_i$ , or an indirect attacker of one of the central arguments of  $ResProg_j$ .

We model a biased community by changing the process of argument generation as follows: each agent is equipped with a bias parameter  $bias \in [0, 1]$ , chosen during initialization. If an agent is investigating an argument  $a$  *in favor* of her preferred research program, then

$$p_{attack} = \begin{cases} 1 - S(a) - bias & \text{if } 1 - S(a) - bias \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

If she is investigating an argument that is *against* her preferred *ResProg* then

$$p_{attack} = \begin{cases} 1 - S(a) + bias & \text{if } 1 - S(a) + bias \leq 1, \\ 1 & \text{otherwise.} \end{cases} \quad (3)$$

The bias parameter changes the probability an argument is attacked based on how it is viewed by the agent. When  $bias = 0$  for all agents, we are in the baseline case of an unbiased community, described below.

**Baseline Community** The success rate is a probability  $p_{attack}$ . In the case of a baseline community without bias,  $bias = 0$  for all agents, and the success rate of generating a counterargument only depends on the strength of the investigated argument:  $p_{attack} = 1 - S(a_i)$  where  $S(a_i)$  is the strength of  $a_i$ .

### 7.3 Publish

When this submodel is executed, agents add the new argument that they have generated to their *view*, and the observer adds the new argument to its *general view* and *new arguments* list. These agents update their *state* to FREE and *result* and *current argument* to None.

New arguments are identified with a unique id, and the id of the argument that they are attacking. Adding an argument to a view consists in adding the new argument to the set of arguments as well as an attack relation from this argument to the one that they attack.

Note that new arguments can only be added to views if they contain the argument that they attack. In the case of the general view, it is always the case. This is an important condition in the update submodel, because agent’s views contain different arguments.

### 7.4 Shared beliefs

Mercier and Heinz argue that *shared beliefs* among scientists can mitigate the detrimental effect of the bias [MH14]. They define shared beliefs as knowledge that is common to all scientists of a community. We hypothesize that the most important shared beliefs are those that concern epistemic standards, i. e. what counts as a good argument for a community. Thus, we model shared beliefs as shared methods: a community with many shared beliefs is a community that has clear standards for what a good argument is. Before a newly generated argument is added to the individual views of other agents, the community decides whether it is worthy of consideration (e.g., whether it fits the generally accepted methodological standards of the field, etc.). If not, it is immediately rejected.

The probability for an argument  $a$  to be discarded at this stage depends on its strength  $S(a)$  and the shared belief parameter  $shBel \in [0, 1]$  which is fixed at the initialization:

$$p_{discard} = shBel \cdot (1 - S(a)).$$

The higher  $shBel$ , the higher the probability that the community will discard weak arguments; for  $shBel = 0$ , we are in a community without shared beliefs. Arguments which pass the shared belief test are added to the *acceptable arguments* list by the observer.

## 7.5 Update

With this submodel, an agent determines whether they will add each accepted new argument in their own view. For each new argument in the *acceptable arguments* list, the agent:

- Checks whether this argument attacks an argument which is part of their own view.
- Performs a probabilistic test of probability  $p_{see}$ .

If the first condition is met, and the test succeeds, the new argument is added to the agent's view.

If  $p_{see}$  is initialized with a different value than 1, this process guarantees that the agents will have different views, and thus groups with diverse preferences can emerge.

## References

- [BPR19] George Butler, Gabriella Pigozzi, and Juliette Rouchier. “An opinion diffusion model with deliberation”. In: *20th International Workshop on Multi-Agent-Based Simulation (MABS 2019)*. 2019.
- [DBM22] Louise Dupuis de Tarlé, Elise Bonzon, and Nicolas Maudet. “Multiagent Dynamics of Gradual Argumentation Semantics”. In: *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 2022, pp. 363–371.
- [Dun95] Phan Minh Dung. “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games”. In: *Artificial intelligence* 77.2 (1995), pp. 321–357.
- [Edm+19] Bruce Edmonds et al. “Different Modelling Purposes”. In: *Journal of Artificial Societies and Social Simulation* 22.3 (2019), pp. 1–6.
- [Gri+06] Volker Grimm et al. “A standard protocol for describing individual-based and agent-based models”. In: *Ecological Modelling* 198.1 (2006), pp. 115–126.
- [Gri+20] Volker Grimm et al. “The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism”. In: *Journal of Artificial Societies and Social Simulation* 23.2 (2020). URL: <https://doi.org/10.18564/jasss.4259>.
- [MC09] Sanjay Modgil and Martin Caminada. “Proof Theories and Algorithms for Abstract Argumentation Frameworks”. In: *Argumentation in Artificial Intelligence*. Ed. by Guillermo Simari and Iyad Rahwan. Springer US, 2009. DOI: 10.1007/978-0-387-98197-0\_6. URL: [https://doi.org/10.1007/978-0-387-98197-0\\_6](https://doi.org/10.1007/978-0-387-98197-0_6).
- [Mer17] Hugo Mercier. “Confirmation bias–myside bias”. In: *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*. Ed. by R. F. Pohl. Routledge/Taylor & Francis Group, 2017, pp. 99–114.
- [MH14] Hugo Mercier and Christophe Heintz. “Scientists’ argumentative reasoning”. In: *Topoi* 33.2 (2014), pp. 513–524.
- [MS11] Hugo Mercier and Dan Sperber. “Why do humans reason? Arguments for an argumentative theory.” In: *Behavioral and brain sciences* 34.2 (2011), pp. 57–74.
- [MS17] Hugo Mercier and Dan Sperber. *The enigma of reason*. Harvard University Press, 2017.

- [TST21] Patrick Taillandier, Nicolas Salliou, and Rallou Thomopoulos. “Introducing the Argumentation Framework Within Agent-Based Models to Better Simulate Agents’ Cognition in Opinion Dynamics: Application to Vegetarian Diet Diffusion”. In: *Journal of Artificial Societies and Social Simulation* 24.2 (2021).