Spécificités de l'argumentation scientifique dans un débat

Louise Dupuis de Tarlé[a] louise.dupuis@dauphine.eu
Gabriella Pigozzi[a] gabriella.pigozzi@lamsade.dauphine.fr
Juliette Rouchier[a] juliette.rouchier@lamsade.dauphine.fr

[a]Universite Paris-Dauphine, CNRS, LAMSADE, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France

Résumé

Depuis l'introduction des Abstract Argumentation Frameworks (cadres d'argumentation abstraits), de nombreux jeux d'argumentation basés sur des agents ont été développés pour étudier les débats. La question de la génération collective de connaissances a été étudiée par des modèles multi-agents, notamment des modèles bayesiens. Notre objectif est d'étudier une communauté épistémique d'agents qui construit un corpus de connaissances communes en générant et en échangeant des arguments. Nous caractérisons les conditions sous lesquelles ces agents convergent vers un résultat cohérent avec leur environnement, divergent ou se polarisent.

Mots-clés : *Argumentation, Multi-Agent, Epistémologie Sociale*

Abstract

Since the introduction of Abstract Argumentation Frameworks, many agent-based argumentation games have been developed to study debates. The question of the collective generation of knowledge has been studied by many agent-based models, such as Bayesian ones. Our goal is to study an epistemic community of agents that builds a body of common knowledge by generating and exchanging arguments. We characterize the conditions under which these agents converge towards a result consistent with their environment, diverge or polarize.

Keywords: *Argumentation, Muti-Agent, Social Epistemology*

1 Introduction

Nous nous intéressons dans cet article à modéliser la discussion scientifique à travers deux de ses spécificités. Pour cela nous faisons des hypothèses sur le but de ces discussions (de découvrir la vérité sur certains aspects de la réalité auxquels l'expérience nous donne accès) et leur méthode (l'échange répété d'arguments contradictoires). Nous utilisons les frameworks d'argumentation abstraite, munis de sémantiques graduelles, pour baser cette modélisation, que nous utilisons pour tester quelques hypothèses sur l'influence d'un biais cognitif : le biais de confirmation [17]. Nous avons choisi de modéliser une communauté scientifique car il s'agit d'un archétype de communauté épistémique publique (les arguments échangés sont visibles de tous). Notre modèle peut permettre d'étudier d'autre types de communautés épistémiques, pourvu qu'elles vérifient les caractéristiques précédentes.

Les fondements de l'epistémologie sociale, qui étudie la connaissance et les croyances de groupes d'individus, ont été introduit par [12]. A la suite, de nombreux travaux ont utilisé les systèmes Multi-Agent pour étudier les communautés scientifiques comme exemples de communautés épistémiques. [23] et [18] utilisent des modèles de réseaux d'agents qui intéragissent avec un environnement et échangent les résultats de leurs expériences. [23] met en évidence l'"Effet Zollman" : il est parfois désavantageux pour les communautés épistémiques de beaucoup communiquer. Une plus grande communication donne lieu à un consensus plus rapide mais empêche certaines idées d'être explorées, ce qui nuit au résultat collectif. Nous obtenons un comportement similaire avec notre modèle.

Notre travail est directement fondé sur [11] qui présentent une étude de la dynamique d'une sémantique graduelle, un type de sémantique appliqués aux frameworks d'argumentation abstraite qui a récemment été proposée dans la littérature [5, 16, 14, 9]. Nous reprenons de nombreux éléments du protocole de [11] : la sémantique utilisée, la représentation de la connaissance chez les agents et la création d'un débat publique. Nous y introduisons les notions de génération de graphe et d'intéraction entre agents et environnement. [11] montrent que l'apprentissage d'arguments par les agents permet à leurs opinions de converger. Notre modèle reproduit ces résultats.

[17] ont également étudié les spécificités du raisonnement et de l'argumentation dans les

milieux scientifiques. Ils mettent en évidence l'existence d'un biais de confirmation qui affecte les communautés de scientifiques autant que celles de profanes. Le biais de confirmation est un biais cognitif qui consiste à privilégier ses propres idées préconçues ou hypothèses, lors de la production d'arguments ou de l'évaluation des informations disponibles. Nous introduisons le biais de confirmation comme paramètre dans notre modèle et étudions son effet.

Notre travail s'inscrit dans une tendance plus large à l'utilisation de modèles d'argumentation pour étudier la diffusion d'opinion. Ainsi, [15] et par la suite [4] utilisent des échanges d'arguments pour modéliser le changement d'opinion et la polarisation des agents, mais ils n'utilisent pas de structure de cadre d'argumentation pour modéliser explicitement le lien entre les arguments. [6, 22] sont deux approches récentes qui mélangent l'argumentation abstraite et la diffusion d'opinion. Notre modèle permet une forme d'évolution dynamique des opinions des agents, aussi nous aimerions caractériser son lien avec les modèles classiques de diffusion d'opinion. Cela explique certains de nos choix expérimentaux, mais cette étude dépasse le cadre de notre article. A notre connaissance, nous sommes les premiers à utiliser les cadres d'argumentation abstraite pour modéliser explicitement une communauté scientifique.

La première section de cet article présente quelques éléments de théorie de l'argumentation abstraite, et la seconde présente notre modèle. Nous exposons dans la dernière section les simulations menées et leurs résultats.

2 Background

2.1 Cadres d'argumentation abstraite

Un framework d'argumentation abstraite (AAF pour *Abstract Argumentation Frameworks*) est constitué d'un ensemble fini d'arguments et d'une relation binaire sur cet ensemble, nommée la relation d'attaque. Les AAF peuvent être représentés par des graphes dont les noeuds sont les arguments et les arrêtes la relation d'attaque. Dans la suite de cet article, nous utiliserons indépendamment les termes AAF et graphes d'argumentation. La Figure 1 présente un petit débat formé de cinq arguments, et leur représentation en tant que graphe d'argumentation avec la relation d'attaque qui les lie. On peut noter que les graphes d'argumentation *abstraite* ne disent rien dans le cas général du contenu des arguments, mais s'occupent de représenter la structure générale que forme le débat.

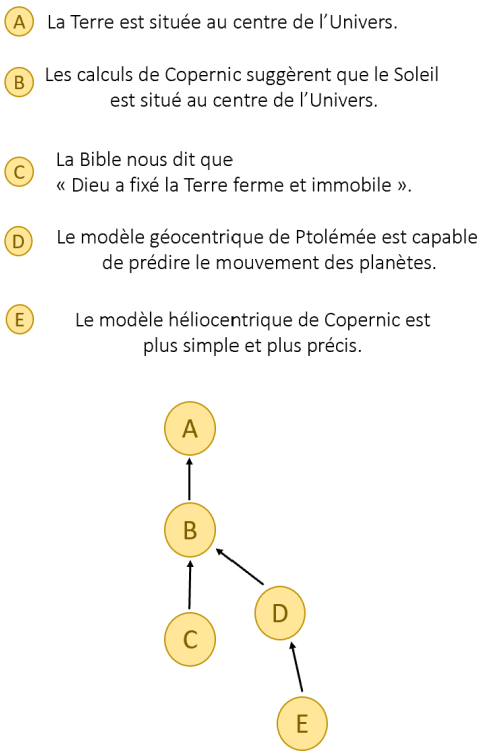


FIGURE 1 – Exemples de débat et sa représentation sous forme d'AAF.

L'argumentation abstraite s'intéresse particulièrement à la question de l'*acceptabilité* des arguments. A partir d'un graphe d'argumentation donné, les **sémantiques** sont des fonctions qui déterminent l'acceptabilité des arguments. Les sémantiques proposées par [10] évaluent les arguments en tant qu'ensembles, qui peuvent être acceptables ou non : ici, acceptable peut être compris comme "rationnellement défendable". Récemment, de nouvelles sémantiques ont été proposées pour permettre une évaluation plus fine de la notion d'acceptabilité. Ainsi, les ranking-based semantics (voir [8, 1, 13, 20, 21, 3, 19]) ordonnent les arguments d'un AAF afin de comparer leurs forces respectives. Un autre type de sémantique sont les sémantiques *graduelles* (voir [5, 16, 14, 9]), qui assignent un score à chaque argument.

2.2 Sémantiques Graduelles

Une **sémantique graduelle** est une fonction qui affecte à chaque argument d'un graphe d'argumentation un score numérique. Le plus souvent,

ce score est compris entre 0 et 1 et représente le *degré d'acceptabilité* de l'argument. Les sémantiques classiques introduites par [10] permettent de déterminer l'acceptabilité d'un argument ou d'un groupe d'argument, mais ne proposent que trois valeurs : accepté, non accepté, ou non décidé. Les sémantiques graduelles permettent de représenter les débats plus finement et présentent des dynamiques intéressantes. Une des propriétés désirables des sémantiques graduelles est qu'une attaque ne détruit pas complètement un argument, mais diminue simplement son acceptabilité. De plus, la présence de valeurs continues permet de faire le lien entre l'argumentation abstraite et le domaine de la diffusion d'opinion, où il est assez courant que les opinions des agents soient des nombres réels compris entre 0 et 1 (ou bien -1 et 1). Ce sont pour ces raisons que nous avons choisit de travailler avec une sémantique graduelle.

Parmi ces sémantiques, nous utilisons une sémantique graduelle particulière, la *h-categorizer* introduite par [5]. Une version de cette sémantique étendue à des graphes munis de poids a été étudiée par [2] qui montre qu'elle satisfait certaines propriétés désirables.

La sémantique **h-categorizer** affecte à chaque argument le *degré d'acceptabilité* suivant :

$$Hbs(a) = \frac{1}{1 + \sum_{b \in Att(a)} Hbs(b)}$$

où $Att(a)$ représente l'ensemble des arguments qui attaquent a.

La Figure 2 montre les degrés d'acceptabilités donnés par *h-categorizer* à chaque argument du graphe d'argumentation vu précédemment. Les arguments non attaqués ont un degré d'acceptabilité maximum de 1.

3 Le Modèle

Notre modèle utilise un type d'AAF introduit par [11] : les graphes d'argumentation orientés vers une *issue* (IOAG pour *issue-oriented argumentation graph*). Un argument spécial, l'*issue*, constitue la racine du graphe et tous les autres arguments font partie d'un chemin d'attaque vers cette *issue*. Intuitivement, l'*issue* représente la proposition centrale du débat, et tous les coups des agents ont pour but d'attaquer ou de défendre cette *issue*.

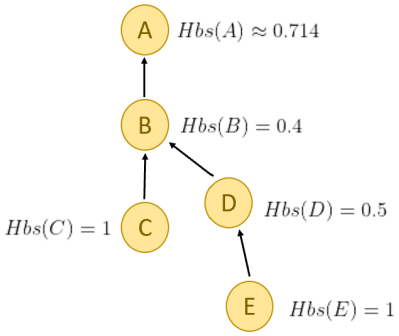


FIGURE 2 – Exemples d'application de la sémantique *h-categorizer* à un graphe d'argumentation.

Comme dans le protocole de [11], les agents construisent ainsi un IOAG commun et visible de tous, en publiant des arguments les uns après les autres. Nous nommons cet arbre le *graphe publique scientifique* (PSG pour Public Scientific Graph). Il représente dans notre modélisation une sorte d'"état de l'art", la somme des connaissances publiés par les agents scientifiques étudiant le problème. Le degré d'acceptabilité, calculé par la sémantique, de l'*issue* au sein de ce graphe correspond à une conclusion collective, qui découle de la production commune des agents. On la nomme V_P pour valeur publique.

Chaque agent possède également son propre AAF, qui contient l'*issue* et les arguments auquel l'agent adhère. Ce graphe correspond à la base de croyance de l'agent, et le degré d'acceptabilité de l'*issue* au sein de ce graphe est appelé **opinion** de l'agent. Les arguments d'un graphe d'opinion ne sont pas nécéssairement tous reliés à l'*issue* par une chaîne d'attaques car au cours du jeu, les agents ajoutent à leur graphe d'opinion les arguments qu'ils produisent ainsi que certains arguments publiés par leurs pairs.

L'environnement extérieur est représenté par une valeur $\mathcal{WV}$ (pour World Value) comprise entre 0 et 1, qui correspond au degré d'acceptabilité (ou degré de croyance) que les agents devraient affecter à l'*issue* du débat si ils avaient un accès total à la "réalité". Bien sûr, les agents n'ont jamais accès directement à cette valeur ; celle-ci guide leur production d'arguments. Les agents ont également plus de probabilités de générer des arguments qui vont dans le sens de leur propre opinion. Le but de notre modèle est d'étudier les conditions dans lesquels la valeur publique produite par les agents, et les opinions individuelles des agents, convergent ou non vers cette valeur

de l'environnement.

Au début du jeu, le PSG et les graphes des agents sont initialisés avec l'*issue* uniquement. Les agents jouent chacun à leur tour, une fois par itération du modèle, dans un ordre aléatoire. A chaque fois qu'il joue, un agent essaie de générer un argument, avec une certaine probabilité : il s'agit d'un simple test aléatoire, cette probabilité corresponant à sa "productivité". Si l'agent réussit, l'argument qu'il produit est généré aléatoirement en prenant en compte l'environnement et l'opinion de l'agent, puis il est publié dans le PSG.

A la fin de chaque tour, les autres agents prennent connaissance du nouvel argument, l'évaluent et peuvent l'ajouter à leur graphe d'opinion.

3.1 Génération d'arguments

A chaque tour, un agent a une certaine probabilité $p_{productivity}$ de générer un nouvel argument. Ce nouvel argument est un contre-argument qui attaque un des arguments du PSG. L'argument attaqué dépend de deux facteurs relatif à l'impact qu'aura cette attaque sur la valeur V_P :

• La **World Value** $\mathcal{WV}$: les arguments qui ont pour impact de rapprocher V_P de $\mathcal{WV}$ ont plus de probabilité d'être générés. Un paramètre α nommé précision détermine à quel point la génération d'argument est précise, c'est à dire correspond à la réalité.

• L'**opinion** de l'agent : les arguments qui ont pour impact de rapprocher V_P de l'opinion de l'agent ont plus de probabilité d'être générés par cet agent. Le paramètre β, ou paramètre de biais, détermine à quel point la génération d'argument est biaisée et favorise l'opinion de l'agent.

Formellement à chaque tour, pour chaque argument i du PSG, on calcule la valeur V_P si on ajoute un contre-argument à i. Notons V_P et V_{Pi} les valeurs du PSG sans et avec l'attaque contre i respectivement, et O l'opinion de l'agent.

Alors, on associe à i un poids $w_i = a * b$:
- Si $|\mathcal{WV} - V_P| \leq |\mathcal{WV} - V_{Pi}|$, $a = \alpha$, sinon $a = 1 - \alpha$. On fixe $\alpha \geq 0.5$.
- Si$|O - V_P| \leq |O - V_{Pi}|$, $b = \beta$, sinon $b = 1 - \beta$. De même, $\beta \geq 0.5$.

Alors, pour N arguments, la probabilité de générer un contre-argument à l'argument j est :

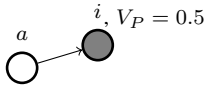
$$P_j = \frac{e^{\gamma w_j}}{\sum_{i=1}^{N} e^{\gamma w_i}}$$

Cette fonction est la fonction logit, qui est souvent utilisée dans les modèles de choix dans l'incertain (voir [7], page 835). Elle permet d'affecter à chaque argument une probabilité proportionnelle au rapport entre son poids et le reste des possibilités. De plus, on peut modifier la façon dont les différences de poids entre arguments affectent les différences de probabilités grâce à γ, le paramètre de diffusion. Ainsi, l'aspect tautologique du modèle est réduit, car on peut diminuer la préférence forte pour les arguments de poids plus grand. Nous recherchions une valeur du paramètre de diffusion qui permette de faire varier la précision $alpha$ sur une plage assez grande tout en conservant une certaine convergence. Après quelques essais sur des graphes simples, nous avons fixé $\gamma = 8$.

On peut remarquer que les agents n'ont jamais accès à la World Value en elle même, cette valeur influence simplement la probabilité de génération des arguments. L'idée est que, malgré les biais, la présence d'un environnement extérieur influence indirectement le type de connaissance généré par les agents.

Example

Soit un jeu à un agent : dans ce cas, le graphe publique PSG et le graphe d'opinion de l'agent sont identiques, représentés par la figure ci-dessous. $\mathcal{WV} = 0.7$*, et les caractéristiques de l'agent sont* $\alpha = 0.9$ *et* $\beta = 0.7$.



L'agent commence par effectuer un test qu'il réussit avec la probabilité $P_{productivity}$*. S'il réussit, il peut générer un argument qui attaque l'*issue i *ou l'argument* a*. La valeur actuelle de l'état de l'art* V_P *et de l'opinion de l'agent* O *sont les mêmes, puisque les gaphes sont identiques :* $V_P = O = 0.5$*.* $V_{Pi} \approx 0,33$ *et* $V_{Pa} \approx 0,67$*. Ainsi, attaquer* a *rapprocherait la valeur du graphe publique de la World Value, contrairement à une attaque envers* i*. Les deux attaques éloignent cette valeur de l'opinion de l'agent. On a donc* $w_i = (1 - \alpha)(1 - \beta) = 0,03$ *et* $w_a = \alpha(1 - \beta) = 0,27$*. Finalement :*

— $P_i \approx 0,13$
— $P_a \approx 0,87$

Ainsi, l'agent a plus de chance de générer un argument qui attaque a *que* i*, mais il n'est pas impossible qu'il génère une attaque contre* i*.*

3.2 Evaluation des arguments

Après chaque publication d'arguments, l'agent qui a produit cet argument l'ajoute à son graphe d'opinion, et les autres agents évaluent chacun l'argument pour déterminer si ils l'acceptent (et l'ajoutent à leur propre graphe) ou non. Pour cela, ils effectuent un test de probabilité P_{accept}. Ainsi, chaque agent a la même probabilité d'accepter tous les arguments produits par ses pairs [1].

3.3 Déroulé d'un jeu

On observe la création d'un arbre d'argumentation orienté vers l'*issue* (IOAG) dont la forme dépend fortement de la World Value. Lorsque α est haut et β bas, c'est à dire le cas d'agents précis et peu biaisés, on observe que les arbres générés pour une $\mathcal{WV}$ proche de 1 sont souvent constitués d'un attaquant direct de l'$issue$ et de nombreux contre attaquants à cet argument. Inversement, les arbres générés lorsque $\mathcal{WV}$ est proche de 0 présentent généralement de nombreux attaquants directs de l'issue. La Figure 3 présente deux exemples de ces types d'arbres. La taille des noeuds est proportionnelle au degré d'acceptabilité des arguments.

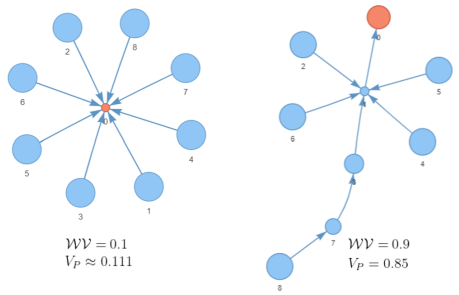


FIGURE 3 – Exemples d'arbres générés par les agents.

4 Simulations

Les questions de recherche qu'on se pose dans cet article sont les suivantes :

1. Est ce qu'une plus grande précision α permet bien au résultat du débat d'être plus proche de la World Value ?
2. Est ce que plus d'agents ou bien plus de pas de temps permettent une convergence plus rapide ?
3. [11] montrent expérimentalement que lorsque la probabilité d'acceptation des arguments augmente, les opinions des agents convergent. Dans leur modèle, les agents ne génèrent pas d'arguments mais débutent le jeu avec un graphe d'opinion dont ils peuvent jouer les arguments. Notre modèle génératif reproduit-il ces résultats ?
4. Quel est l'effet du biais de génération ? Est-il similaire au niveau collectif et individuel ?

La question 1 peut sembler évidente au premier abord, mais [11] montre que la sémantique utilisée exhibe certaines propriétés non intuitives, ce qui nous pousse à être précautionneux. Les questions 1 et 2 permettent ainsi de valider que notre modèle fonctionne correctement, et les questions 3 et 4 sont des questions ouvertes sur les phénomènes que le modèle fait émerger. Pour répondre à ces questions, nous introduisons les trois métriques suivantes :

L'**Erreur collective (CE)** correspond à la distance entre $\mathcal{WV}$ et V_p. Elle mesure la réussite des agents en tant que groupe construisant une somme de connaissances.

L'**Ecart type des opinions des agents (STD)**, qui mesure la diversité des opinions.

L'**Erreur de la moyenne (EA)** qui correspond à la distance entre la moyenne des opinions des agents et la $\mathcal{WV}$. On peut voir la moyenne des opinions des agents comme une représentation du consensus. Cette métrique permet de caractériser la réussite individuelle des agents.

4.1 Description des Tests Expérimentaux

Nous avons généré aléatoirement $\mathcal{WV}$ à chaque itération. Le paramètre de productivité $p_{productivity}$ est fixé à 0.5 pour toutes nos expériences. Les valeurs par défaut des paramètres sont : $\alpha = 0.95$, $\beta = 0.6$, $P_{accept} = 0.7$, le nombre de pas de temps $S = 20$ et le nombre d'agents $N = 3$. Chaque résultat correspond à la moyenne de la métrique correspondante sur 150 débats.

Nous présentons les résultats de simulation à trois agents pour limiter le temps d'expérimentation, tout en nous permettant d'observer des phénomèmes intéressant qui diffèrent du cas avec un

1. [11] introduisent un biais de confirmation dans l'évaluation des arguments par les agents (les agents ont plus de chance d'accepter les arguments qui confirment leur opinion). Pour l'instant, nous restreignons notre étude au biais de confirmation appliqué à la *génération* d'arguments.

seul agent. Nous avons observé les mêmes phénomènes lors d'une étude plus qualitative des jeux avec 6 agents.

5 Résultats

5.1 Effet de la précision

On fait varier la précision α en laissant les valeurs par défaut. Les résultats présentés dans la Table 1 et Figure 4 montrent qu'une plus grande précision permet bien de diminuer l'erreur collective et les erreurs individuelles des agents. On observe aussi une légère diminution de l'écart type des opinions des agents. Nous supposons que cela est dû au fait que les agents convergent tous vers une opinion plus proche de la $\mathcal{WV}$.

	STD	CE	EA
0,5	0,074	0,257	0,260
0,6	0,079	0,181	0,186
0,7	0,068	0,133	0,146
0,8	0,065	0,080	0,108
0,9	0,060	0,051	0,082
1	0,064	0,036	0,068

TABLE 1 – Variation de l'écart type des opinions (STD), l'erreur collective (CE), et l'erreur de la moyenne (EA) en faisant varier la précision α.

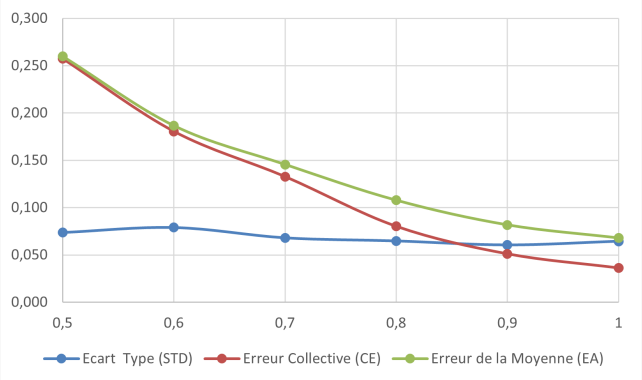


FIGURE 4 – Graphe représentant la variation des trois métriques en faisant varier la précision α.

5.2 Rôle du nombre d'arguments générés

Les règles du protocole nous donnent une approximation du nombre d'argument générés : $N_{arguments} \approx N_{agents} * P_{productivity} * S$, avec S le nombre de pas de temps du modèle. La Table 3 montre l'erreur collective (CE) en fonction du nombre d'agents et de pas de temps du modèle. Les résultats confirment que la convergence à la $\mathcal{WV}$ de la valeur publique est plus rapide lorsque le nombre d'agent augmente, et lorsque le nombre de pas de temps du modèle augmente. On peut en conclure que le nombre d'argument générés influence positivement la convergence. Une légère augmentation de l'erreur pour 6 agents et 50 arguments nous indique qu'il pourrait exister un nombre optimal de nombre d'arguments produits.

5.3 Effet de la probabilité d'acceptation

Nous faisons varier la probabilité d'acceptation des arguments. Les résultats présentés en Table 2 confirment les résultats expérimentaux de [11] : plus la probabilité d'acceptation est haute, plus l'opinion des agents converge. On observe que l'erreur de la moyenne des agents diminue également lorsque P_{accept} augmente. Cela est dû au phénomène identifié plus haut : avoir accès à plus d'arguments permet aux agents de mieux converger, dans le cas où ces agents sont précis et peu biaisés. L'erreur collective augmente légèrement, ce qui peut être dû à l'effet néfaste du biais de confirmation lorsque les communautés communiquent plus, qui est identifié à la section suivante.

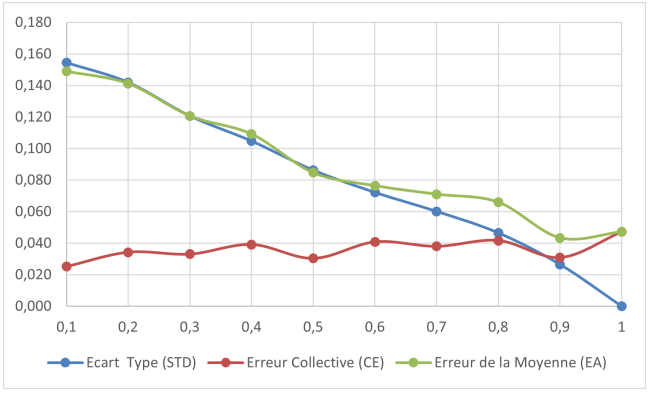


FIGURE 5 – Graphe représentant la variation des trois métriques en fonction de la probabilité d'acceptation P_{accept}.

5.4 Effet du biais de génération

On fait varier le biais β dans le cas de trois agents avec les paramètres par défaut, et dans un deuxième cas où les trois agents n'acceptent aucun argument de leurs pairs : $P_{accept} = 0$.

Les résultats présentés en Table 4 et les Figures 6 et 8 nous montrent que le biais a un effet néfaste sur les succès individuels et collectifs des

P_{accept}	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Ecart Type	0,154	0,142	0,121	0,105	0,086	0,072	0,060	0,047	0,027	0,000
Erreur Collective	0,025	0,034	0,033	0,039	0,030	0,041	0,038	0,042	0,031	0,047
Erreur de la Moyenne	0,149	0,141	0,121	0,109	0,085	0,076	0,071	0,066	0,043	0,047

TABLE 2 – Variation de l'écart type des opinions (STD), l'erreur collective (CE), et l'erreur de la moyenne (EA) en faisant varier la probabilité d'acceptation P_{accept}.

Agents/P. de Temps	10	20	30	50
1	0,094	0,075	0,064	0,051
2	0,054	0,053	0,045	0,029
6	0,041	0,033	0,024	0,026

TABLE 3 – Erreur collective (CE) en fonction du nombre d'agents et de pas de temps du modèle.

	β	$P_{accept} = 0$	$P_{accept} = 0.7$
CE	0,5	0,029	0,032
	0,6	0,035	0,043
	0,7	0,039	0,049
	0,8	0,071	0,097
	0,9	0,107	0,156
	0,9999	0,162	0,213
STD	0,5	0,155	0,058
	0,6	0,174	0,064
	0,7	0,174	0,064
	0,8	0,19	0,07
	0,9	0,224	0,08
	0,9999	0,248	0,073
EA	0,5	0,179	0,068
	0,6	0,18	0,073
	0,7	0,194	0,081
	0,8	0,215	0,116
	0,9	0,224	0,174
	0,9999	0,273	0,23

TABLE 4 – Variation de l'écart type des opinions (STD), l'erreur collective (CE), et l'erreur de la moyenne (EA) en faisant varier le biais β, dans le cas avec acceptation $P_{accept} = 0.7$ et sans $P_{accept} = 0$.

agents. Cela est cohérent avec la vision classique du biais de confirmation comme un obstacle à la rationnalité des raisonnements individuels. Le groupe d'agents à $P_{accept} = 0$ montre une plus grande diversité d'opinion que celui à $P_{accept} = 0.7$, comme illustré en Figure 7, ce qui est cohérent avec les résultats de la section précédente. On observe également que les agents qui n'acceptent pas d'autres arguments que les leurs semblent protégés des effets du biais, et commettent moins d'erreurs que ceux qui acceptent

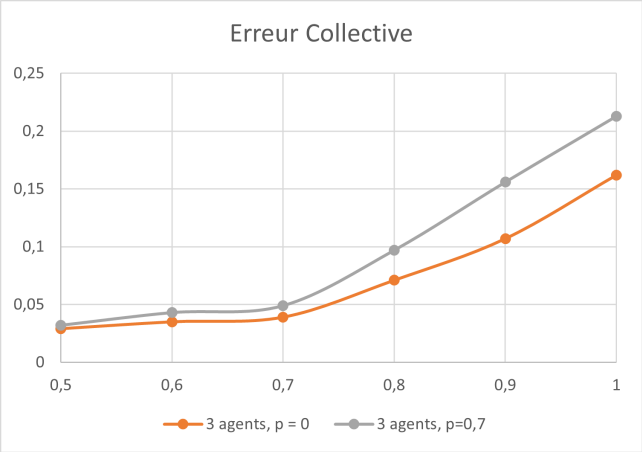


FIGURE 6 – Graphe représentant la variation de l'erreur collective (CE) en fonction du biais β dans le cas avec acceptation $P_{accept} = 0.7$ et sans $P_{accept} = 0$.

certains arguments de leurs pairs. Ce résultat qui pourrait apparaître contre-intuitif à première vue semble pointer vers un facteur protectif de la diversité d'opinion : les agents ayant chacun des opinions différentes, le résultat de leurs efforts collectifs reflète mieux la "réalité", c'est à dire la $\mathcal{WV}$. Au contraire, lorsque les agents apprennent les uns des autres, le résultat collectif est plus biaisé vers leur opinion commune. Cette intuition semble confirmée par le fait que l'erreur de la moyenne des opinions des agents est plus grande dans le cas sans acceptation d'arguments. Ainsi, lorsque les agents biaisés ne communiquent pas, ils font plus d'erreur individuellement mais ils parviennent collectivement à mieux approcher la vérité. Ces résultats sont proches de ceux obtenus par [23] et montrent que la communication peut parfois s'avérer désavantageuse pour une communauté à la recherche de la vérité.

6 Conclusion et Discussion

Les résultats semblent montrer que notre modèle permet bien de modéliser la génération d'argu-

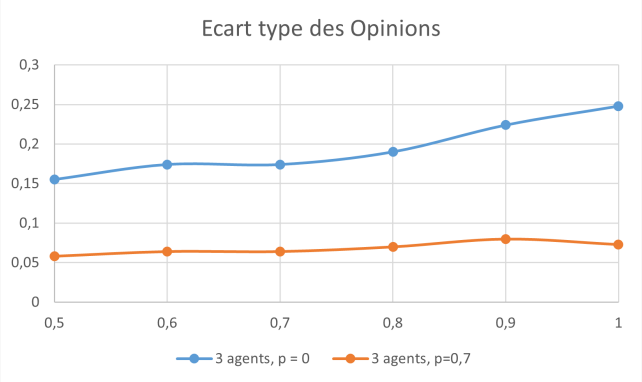


FIGURE 7 – Graphe représentant la variation de l'écart type des opinions (STD) en fonction du biais β dans le cas avec acceptation $P_{accept} = 0.7$ et sans $P_{accept} = 0$.

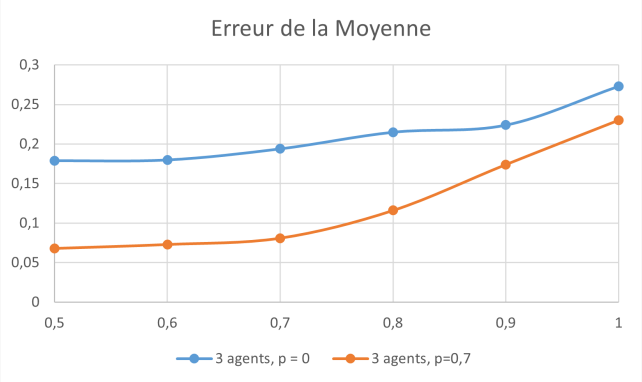


FIGURE 8 – Graphe représentant la variation de l'erreur de la moyenne (EA) en fonction du biais β dans le cas avec acceptation $P_{accept} = 0.7$ et sans $P_{accept} = 0$.

ments influencés par un environnement extérieur à l'aide d'une sémantique graduelle. Nous modélisons de façon satisfaisante les notions de précision par rapport à l'environnement et de biais de confirmation dans la génération d'arguments. A notre connaissance, il s'agit du premier travail de recherche qui utilise les sémantiques graduelles pour représenter la construction d'un débat influencé par une notion de réalité extérieure et la convergence associée.

Nous confirmons les résultats de [11] qui semblent ainsi robuste aux détails du protocole utilisé : partager des arguments permet aux opinions des agents de converger. Cela permet de créer un lien entre les frameworks d'argumentation abstraite et la diffusion d'opinion. Nous observons également que la diversité d'opinion semble protéger les groupes d'agents contre les effets du biais de confirmation. Ce résultat rappelle ceux obtenus par [23] : une communication réduite est parfois bénéfique pour les communautés épistémiques.

Notre modèle présente toutefois des limites : ainsi, nous n'avons pas présenté d'interprétation satisfaisante de la notion de World Value, et la façon dont les arguments sont générés est ambigue et pourrait donner à penser que les agents ont connaissance de cette valeur. Pour dépasser cela, nous aimerions définir une nouvelle méthode de génération d'arguments qui serait fondée sur le principe d'apprentissage par expérience. De plus, notre modèle présente une certaine assymétrie entre les valeurs basses de la WV, qui sont plus faciles à atteindre que les valeurs hautes pour les agents. Cela est dû au fait que les graphes utilisés contiennent uniquement des attaques et qu'il est donc plus facile d'attaquer un argument (diminuer sa valeur) que de le défendre (l'augmenter). Nous proposons par la suite d'utiliser des graphes d'argumentation *bi-directionnels*, c'est à dire contenant des relations de support aussi bien que d'attaque entre arguments. Le modèle pourrait également être augmenté d'un biais de confirmation lors de *l'évaluation* des arguments, comme chez [11].

Ainsi, ce premier modèle est une étape qui nous permet de valider notre concept, avant de le perfectionner. Notre but est ensuite étendre le modèle et d'inclure, en s'inspirant du travail de [18], la diffusion de la production de connaissances scientifiques auprès d'une communauté de profanes.

Références

[1] Leila AMGOUD et Jonathan BEN-NAIM. "Ranking-Based Semantics for Argumentation Frameworks". In : *Proc. of the 7th Int. Conference on Scalable Uncertainty Management, (SUM'13)*. 2013, p. 134-147.

[2] Leila AMGOUD et al. "Acceptability semantics for weighted argumentation frameworks". In : *Proceedings of the Twenty-Sixth Int. Joint Conference on Artificial Intelligence, IJCAI-17*. 2017, p. 56-62. DOI : 10.24963/ijcai.2017/9.

[3] Leila AMGOUD et al. "Ranking Arguments With Compensation-Based Semantics". In : *Proc. of the 15th Int. Conference on Principles of Knowledge Representation and Reasoning, (KR'16)*. 2016, p. 12-21.

[4] Sven Banisch, Eckehard Olbrich et al. "An Argument communication model of polarization and ideological alignment". In : *Journal of Artificial Societies and Social Simulation* 24.1 (2021), p. 1-1.

[5] Philippe Besnard et Anthony Hunter. "A logic-based theory of deductive arguments". In : *Artificial Intelligence* 128.1-2 (2001), p. 203-235.

[6] George Butler, Gabriella Pigozzi et Juliette Rouchier. "An opinion diffusion model with deliberation". In : *20th Int. Workshop on Multi-Agent-Based Simulation (MABS 2019)*. Montreal, Canada, mai 2019.

[7] Colin Camerer et Teck Hua Ho. "Experience-weighted attraction learning in normal form games". In : *Econometrica* 67.4 (1999), p. 827-874.

[8] Claudette Cayrol et Marie-Christine Lagasquie-Schiex. "Graduality in Argumentation". In : *Journal of Artificial Intelligence Research* 23 (2005), p. 245-297.

[9] Célia da Costa Pereira, Andrea Tettamanzi et Serena Villata. "Changing one's mind : erase or rewind ?" In : *Proc. of the 22nd Int. Joint Conference on Artificial Intelligence, (IJCAI'11)*. 2011, p. 164-171.

[10] Phan Minh Dung. "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and N-persons games". In : *Artificial Intelligence* 77 (1995), p. 321-357.

[11] Louise Dupuis de Tarlé, Elise Bonzon et Nicolas Maudet. "Multiagent dynamics of gradual argumentation semantics". In : *Proceedings of the 21th Int. Conference on Autonomous Agents and Multi Agent Systems (AAMAS)* (à para.).

[12] Alvin I Goldman. *Knowledge in a Social World*. Oxford University Press, 1999.

[13] Davide Grossi et Sanjay Modgil. "On the Graded Acceptability of Arguments". In : *Proc. of the 24th Int. Joint Conference on Artificial Intelligence, (IJCAI'15)*. 2015, p. 868-874.

[14] João Leite et João Martins. "Social abstract argumentation". In : *Proc. of the 22nd Int. Joint Conference on Artificial Intelligence, (IJCAI'11),* 2011, p. 2287-2292.

[15] Michael Mäs et Andreas Flache. "Differentiation without distancing. Explaining bi-polarization of opinions without negative influence". In : *PloS one* 8.11 (2013), e74516.

[16] Paul-Amaury Matt et Francesca Toni. "A Game-theoretic measure of argument strength for abstract argumentation". In : *Proc. of the 11th European Conference on Logics in Artificial Intelligence, (JELIA'08)*. 2008, p. 285-297.

[17] Hugo Mercier et Christophe Heintz. "Scientists' argumentative reasoning". In : *Topoi* 33.2 (2014), p. 513-524.

[18] Cailin O'Connor et James Owen Weatherall. "Scientific polarization". In : *European Journal for Philosophy of Science* 8.3 (2018), p. 855-875.

[19] Theodore Patkos, Antonis Bikakis et Giorgos Flouris. "A Multi-aspect evaluation framework for comments on the social web". In : *Proc. of the 15th Int.l Conference on Principles of Knowledge Representation and Reasoning (KR'16)*. 2016, p. 593-596.

[20] Fuan Pu et al. "Argument Ranking with Categoriser Function". In : *Proc. of the 7th Int. Conference on Knowledge Science, Engineering and Management, (KSEM'14)*. 2014, p. 290-301.

[21] Fuan Pu et al. "Attacker and defender counting approach for abstract argumentation". In : *Proc. of the 37th Annual Meeting of the Cognitive Science Society, (CogSci'15)*. 2015.

[22] Patrick Taillandier, Nicolas Salliou et 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), p. 6. issn : 1460-7425. doi : 10.18564/jasss.4531. url : http://jasss.soc.surrey.ac.uk/24/2/6.html.

[23] Kevin JS Zollman. "The communication structure of epistemic communities". In : *Philosophy of Science* 74.5 (2007), p. 574-587.