



HAL
open science

Raisonner sur l'éthique avec un cadre d'argumentation fondé sur une logique modale normale

Christopher Leturc, Grégory Bonnet

► To cite this version:

Christopher Leturc, Grégory Bonnet. Raisonner sur l'éthique avec un cadre d'argumentation fondé sur une logique modale normale. Journées Francophones Planification, Décision et Apprentissage, Jun 2022, Saint-Etienne, France. pp.113-120. hal-03819902

HAL Id: hal-03819902

<https://hal.science/hal-03819902>

Submitted on 18 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Raisonner sur l'éthique avec un cadre d'argumentation fondé sur une logique modale normale

C. Leturc¹ et G. Bonnet²

¹ Inria, Université Côte d'Azur, CNRS, I3S, France

² Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, France

christopher.leturc@inria.fr, gregory.bonnet@unicaen.fr

Résumé

Le comportement d'agents artificiels autonomes peut soulever des questions éthiques lorsque leurs actions ont un impact sur les humains, questions que les agents doivent prendre en compte dans leur raisonnement. Si certaines approches représentent le raisonnement moral avec des logiques déontiques, utilisant des opérateurs modaux, d'autres décrivent le raisonnement éthique avec un cadre d'argumentation valué. Cependant, il n'existe pas à notre connaissance de cadre qui combine à la fois logique modale et argumentation pour raisonner sur l'éthique. Dans cet article, nous proposons donc un cadre d'argumentation où les arguments sont construits à partir d'un cadre général de logique modale normale. Celui-ci permet d'exprimer et de raisonner sur un ensemble de valeurs morales via des opérateurs déontiques dyadiques (voire n-aires). Toutefois comme n'y a pas de contradiction nécessaire entre les opérateurs modaux et certaines attaques classiques ne peuvent pas être déduites. Nous proposons alors un mécanisme pour décrire les formes que doivent prendre les contradictions.

Mots-clés

Argumentation, Éthique artificielle, Logique modale.

Abstract

When the actions of artificial agents impact humans, autonomous behaviors may rise ethical issues that agents must consider in their reasoning. Some approaches represent moral reasoning with deontic logics, which use modal operators, while reasoning about ethics may be represented with a value-based argumentation framework as it has been proposed in the literature. However, there is no work that combines both modal logic and argumentation to reason about ethics. In this article, we propose an argumentation framework where arguments are built from a general normal modal logic framework. It allows to express different kinds of modal operators of interest in ethical reasoning, such as moral worth operators, dyadic (or n-ary) deontic operators. However since there is no necessary contradiction between modal operators, some classical argumentative attacks cannot be characterized. To solve this problem, we define attacks based on modal logic thank to a description of how the contradictions must arise.

Keywords

Argumentation, Computational ethics, Modal logic.

1 Introduction

Les comportements des agents artificiels peuvent soulever des questions éthiques lorsque leurs actions ont un impact sur les humains, questions qu'ils doivent représenter et intégrer dans leur raisonnement. À cette fin, de nombreux cadres ont été développés pour permettre la conception d'agents autonomes intégrant des concepts éthiques explicites [5, 12, 18, 19]. Si ces approches modélisent certains aspects spécifiques du raisonnement éthique, elles ne rendent pas entièrement compte de la manière dont les êtres humains effectuent un tel raisonnement. En effet, lorsqu'un être humain prend une décision, il tient compte de ses désirs, de ses croyances sur une situation, de ses émotions ou de valeurs morales [15, 23]. Cependant, dans certaines situations, il n'est pas capable de donner une évaluation différente entre deux options, chacune soutenue par une valeur morale différente et chacune apportant un regret après leur exécution. Pour résoudre ces dilemmes moraux, les êtres humains recherchent des arguments qui soutiendront certaines valeurs morales, considérées comme importantes dans la situation [17]. Dans cet article, nous avons pour but de modéliser un tel processus de raisonnement éthique en considérant différentes valeurs morales et en les confrontant les unes aux autres à travers un cadre d'argumentation (AF) [14]. Plus précisément, nous utilisons un cadre d'argumentation fondé sur des valeurs (VAF) [6]. Dans un VAF, les arguments sont associés à une valeur qui peut correspondre à une valeur morale. Par exemple, un véhicule autonome dans une situation d'urgence peut se représenter deux arguments pour savoir si celui-ci doit dépasser la limitation vitesse pour arriver plus vite à l'hôpital. Un premier argument pourrait indiquer que selon la loi, « il est interdit de dépasser la limite de vitesse ». Cet argument promeut une valeur *respecter la loi*. Un second peut indiquer qu'« une passagère mourante doit arriver à l'hôpital le plus vite possible » et peut promouvoir une valeur *préserver la vie*. Cependant, dans un VAF, les associations entre arguments et valeurs sont données a priori. C'est pourquoi nous proposons d'aller plus loin en considérant un VAF

fondé sur une logique modale. En effet, nous affirmons que le raisonnement selon les valeurs peut être représenté par une logique modale puisque les opérateurs modaux sont bien adaptés pour exprimer des notions de désirs (valeurs égoïstes ou hédoniques) ou de mondes idéaux déontiques. De plus, la logique modale est également adaptée pour exprimer et raisonner sur les connaissances et les croyances d'un agent. Par exemple, dans l'exemple précédent, nous pouvons considérer un autre argument qui exprimerait que « la passagère n'est pas mourante pas car elle souffre d'une luxation de l'épaule ». Un tel argument ne promeut aucune valeur morale mais attaque clairement l'argument précédent. Ainsi, la principale contribution de cet article est de fournir un cadre d'argumentation où les arguments et les relations d'attaque sont sémantiquement définis du point de vue des logiques modales.

2 Application

L'exemple ci-après a pour objectif de modéliser le raisonnement éthique d'un agent devant décider d'un mode de transport. Pour prendre sa décision, nous faisons l'hypothèse que l'agent considère un ensemble de valeurs, a une préférence sur elles et connaît les différents itinéraires associés au mode de transport choisi : prendre sa voiture, son vélo, les transports publics ou aller à pied. Nous supposons aussi que l'agent ne prend que des décisions rationnelles fondées ses préférences. L'exemple 1 introduit quelques éléments formels qui seront repris dans la suite de cet article.

Exemple 1. *Nous considérons un ensemble de valeurs \mathcal{Val} , constitué d'une valeur écologique associée au respect de l'environnement (v_{env}), une valeur économique associée à la maximisation de son intérêt personnel (v_{eco}), une valeur associée au respect de consignes données par une autorité gouvernementale (v_{aut}) ainsi qu'une valeur associée au respect de sa santé (v_{health}). Nous considérons donc que l'agent doit décider entre plusieurs actions représentées par les symboles suivants :*

- $\Delta_{car} :=$ "prendre sa voiture"
- $\Delta_{p.t.} :=$ "prendre les transports publics"
- $\Delta_{bike} :=$ "prendre son vélo"
- $\Delta_{walk} :=$ "partir à pied"

Nous représentons par \mathcal{P} un ensemble de variables propositionnelles décrivant des informations sur la situation et l'agent. \mathcal{P} contient les éléments suivants :

- $has_{bike} :=$ "l'utilisateur possède un vélo"
- $has_{car} :=$ "l'utilisateur possède une voiture"
- $crisis_{health} :=$ "il y a une crise sanitaire"
- $congestion :=$ "il y a un embouteillage"
- $emergency :=$ "il s'agit d'une urgence médicale"

3 Notions préliminaires

Dans un premier temps, nous introduisons des notions élémentaires sur l'argumentation valuée. Dans un second temps, nous présentons la littérature qui combine à la fois logiques modales et argumentation.

3.1 Argumentation valuée

Un VAF étend le cadre de l'argumentation abstraite introduit par Dung [14]. Ce dernier l'étend en introduisant des concepts de valeurs, et de préférences sur les valeurs [6].

Définition 1 (VAF). *Un VAF est un tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{Val}, \mu, \succ \rangle$ tel que : \mathcal{A} est un ensemble non vide d'arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ est une relation binaire décrivant la notion d'attaque entre arguments de \mathcal{A} , \mathcal{Val} est un ensemble non vide de valeurs ; $\mu : \mathcal{A} \rightarrow \mathcal{Val}$ est une fonction associant à chaque argument une valeur, et $\succ \subseteq \mathcal{Val} \times \mathcal{Val}$ est une relation transitive de préférence entre les valeurs.*

Un argument en écarte un autre s'il l'attaque et que l'argument attaqué n'est à l'argument attaquant au regard des valeurs qui leur sont respectivement associées.

Définition 2 (Écart). *Soit $\langle \mathcal{A}, \mathcal{R}, \mathcal{Val}, \mu, \succ \rangle$ un VAF. Un argument $A \in \mathcal{A}$ écarte un argument $B \in \mathcal{A}$, noté $A \mathcal{R}_{vaf} B$ si, et seulement si, $A \mathcal{R} B$ i.e. A attaque B , et $\mu(B) \not\succeq \mu(A)$ i.e. la valeur $\mu(B)$ n'est pas préférée à la valeur $\mu(A)$.*

Définition 3 (Sans conflit et acceptabilité). *Soient \mathcal{R}_{vaf} une attaque VAF sur $\langle \mathcal{A}, \mathcal{R}, \mathcal{Val}, \mu, \succ \rangle$, et $S \subseteq \mathcal{A}$. S est un ensemble d'argument sans conflit par rapport à \mathcal{R}_{vaf} si, et seulement si, $\nexists A, B \in S$ t.q. $A \mathcal{R}_{vaf} B$. Pour tout $A \in \mathcal{A}$, A est acceptable pour S si, et seulement si, $\forall B \in \mathcal{A}$ si $B \mathcal{R}_{vaf} A$, alors $\exists C \in S, C \mathcal{R}_{vaf} B$.*

Pour caractériser des arguments dits « admissibles », nous utilisons la notion de sémantique d'admissibilité ou extensions [14]. Nous donnons ci-dessous les principales :

Définition 4 (Extensions). *Soit $S \subseteq \mathcal{A}$. S est une extension admissible ssi S est sans conflit et tout argument $A \in S$ est acceptable pour S . S est une extension complète ssi S est admissible et contient tous les arguments acceptables pour S . S est une extension préférée ssi S est une extension admissible maximale pour \subseteq . S est une extension stable ssi S est sans conflit, et $\forall A \in \mathcal{A} \setminus S, S \mathcal{R}_{vaf} A$.*

Dans la suite, nous ne faisons pas de supposition sur la sémantique utilisée.

3.2 L'argumentation fondée sur la logique

Pour combiner logique et argumentation, certains travaux utilisent la logique pour modéliser l'argumentation formelle [11, 16, 22] et d'autres considèrent l'argumentation comme une structure issue d'une logique [1, 4, 7, 8, 10, 20]. La première approche ne permet pas d'exprimer des arguments fondés sur des états mentaux ou des modalités déontiques car la relation d'accessibilité dans ces modèles représentent une relation d'attaque entre arguments. C'est pourquoi, nous nous intéressons uniquement à la seconde approche. Par exemple, [8] définit une argumentation fondée sur la logique propositionnelle : un argument est un couple (prémisse, conclusion); la relation d'attaque est définie sur les contradictions déduites dans la logique. Pour aller plus loin, [10] propose une argumentation fondée sur le calcul des hyperséquents et une logique modale S5. Ici,

$A = (\Gamma, \phi)$ est un argument avec Γ un ensemble de formules de S5 supportant la conclusion ϕ si, et seulement si, $\Gamma \Rightarrow \phi$ est un séquent prouvé. Cependant, il n'y a qu'un type de modalité et [10] ne fournit pas de sémantique à leur opérateur. Enfin, plutôt que de considérer une sémantique particulière associée à une logique, [1] propose une généralisation de l'argumentation en considérant une fonction \mathcal{C} représentant l'opérateur de conséquence syntaxique \vdash (i.e. ce qui peut être déduit d'un système de preuve). Cependant, toutes ces approches ne fournissent pas de sémantique associé aux opérateurs de preuves syntaxiques. À notre connaissance, si certains travaux introduisent une sémantique de logique du premier ordre pour les arguments comme [7], aucun ne propose un cadre d'argumentation fondé sur la sémantique des logiques modales normales.

4 Logique multimodale n-aire

Pour exprimer certaines logiques de la morale comme DL-MA [19] ou les logiques déontiques dyadiques [21], nous considérons un cadre de logique multimodal normal avec des opérateurs n -aires [9]. Pour ce faire, nous devons tout d'abord définir la notion de type de similarité.

Définition 5 (Type de similarité). *Un type de similarité est un couple $\tau = (\mathcal{O}, \rho)$ où \mathcal{O} représente un ensemble non vide de symboles pour chaque opérateur modal et $\rho : \mathcal{O} \rightarrow \mathbb{N}$ associe une arité à chaque opérateur modal.*

Un type de similarité représente un ensemble d'opérateurs modaux avec une arité ≥ 0 qui lui est assignée. Les éléments de \mathcal{O} sont usuellement représentés par un « triangle » comme $\Delta_0, \Delta_1, \dots$ pour les opérateurs modaux d'arité 0 (appelés opérateurs nullaires). Ces opérateurs peuvent être vus comme des variables propositionnelles au niveau du cadre de Kripke et pas au niveau d'un modèle.

Exemple 2. *Considérons $\Delta_{car}, \Delta_{p.t.}, \Delta_{bike}, \Delta_{walk}$ les choix de l'agent de respectivement prendre sa voiture, prendre les transports publics, partir en vélo, ou partir à pieds. Ces opérateurs nullaires représentent donc les choix.*

Pour les arités supérieures, par convention l'ensemble \mathcal{O} contient seulement des modalités de type « diamants ». Ainsi, si un Δ est un opérateur monadique, i.e. $\rho(\Delta) = 1$, alors il s'agit de l'opérateur standard \diamond , i.e. une modalité de type « existence » (noté \exists -type). Pour $\rho(\Delta) > 1$, un opérateur n -aire Δ signifie qu'il existe un N -uplet de mondes accessibles où chaque formule à l'intérieur des paramètres est vraie. Nous représentons par ∇ le dual de Δ . Quand ∇ est un opérateur monadique, il est équivalent à l'opérateur standard \square . Pour un τ donné, nous notons \mathcal{O}_{dual} l'ensemble des opérateurs duaux dans \mathcal{O} , i.e. tous les types « pour tout » (noté \forall -type). Pour les modalités d'arité supérieure, ∇ signifie que quelque soit les N -uplet de mondes accessibles, au moins une formule contenue dans les paramètres de la modalité est vérifiée. Notons que les valeurs peuvent être soit représentées par des opérateurs nullaires (pour exprimer des valeurs promues dans un monde) ou par des opérateurs d'arité supérieure pouvant être interprété comme des obligations déontiques au nom d'une valeur spécifique.

Exemple 3. *Nous pouvons représenter « prendre son vélo promeut la valeur environnementale » soit en considérant un opérateur nullaire Δ_{env} qui est vrai pour chaque monde où Δ_{bike} est vrai, ou avec un opérateur déontique monadique de SDL e.g. $\nabla_{env}(\Delta_{bike})$, ou de façon plus expressive, avec un opérateur n -aire. Dans la suite, nous considérons des opérateurs dyadiques [21] dont un paramètre représente le contexte d'application d'une règle déontique et le second paramètre représente ce qui devrait être dans un monde idéal associé à ce contexte. Intuitivement, ces opérateurs représentent une contrainte sur trois mondes (w, v, u) où w est un monde dans lequel une norme (v, u) peut s'appliquer, norme signifiant que u est un monde idéal pour un monde v où le contexte est vérifié.*

Contrairement à [9], nous considérons des sémantiques alternatives pour Δ et ∇ : un opérateur n -aire, noté \square , est défini comme pour tout N -uplet de mondes accessibles, chaque formule à l'intérieur de \square est vérifiée. \square a aussi sa modalité duale, notée \diamond , qui est définie comme il existe un N -uplet tel que au moins une formule est vraie dans son monde respectif. Les ensembles des \diamond et \square sont représentés respectivement par \mathcal{O}^* et \mathcal{O}_{dual}^* . Pour lier chaque opérateur modal avec son dual et ses sémantiques alternatives, nous notons $\mathcal{O}^\Omega = \mathcal{O} \cup \mathcal{O}_{dual} \cup \mathcal{O}^* \cup \mathcal{O}_{dual}^*$ l'ensemble de tous les opérateurs modaux et une fonction $\varrho : \mathcal{O}^\Omega \rightarrow \mathcal{O} \times \mathcal{O}_{dual} \times \mathcal{O}^* \times \mathcal{O}_{dual}^*$ qui associe à chaque modalité son dual et ses alternatives. Dans la suite, nous considérons que le domaine de ρ est étendu à \mathcal{O}^Ω , i.e. $\rho : \mathcal{O}^\Omega \rightarrow \mathbb{N}^*$ et ρ est telle que : $\forall \Delta \in \mathcal{O}$, si $\varrho(\Delta) = (\Delta, \nabla, \diamond, \square)$, alors $\rho(\Delta) = \rho(\nabla) = \rho(\diamond) = \rho(\square)$.

Définition 6 (Langage des logiques modales). *Soit \mathcal{P} un ensemble d'atomes propositionnels et $\tau = (\mathcal{O}, \rho)$ un type de similarité. Nous considérons le langage des formules bien formées (fbf) généré par la grammaire sous forme de Backus-Naur, pour tout $p \in \mathcal{P}$, $\Delta \in \mathcal{O}$, $\star \in \varrho(\Delta)$:*

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid \star(\phi_1, \dots, \phi_{\rho(\Delta)})$$

De façon standard, nous avons $\top := \neg\perp$, $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, et $\phi \Rightarrow \psi := \neg\phi \vee \psi$. De plus si $\rho(\Delta) > 0$:

$$\begin{aligned} \neg \nabla(\phi_1, \dots, \phi_{\rho(\Delta)}) &:= \neg\Delta(\neg\phi_1, \dots, \neg\phi_{\rho(\Delta)}) \\ \neg \square(\phi_1, \dots, \phi_{\rho(\Delta)}) &:= \neg\diamond(\neg\phi_1, \dots, \neg\phi_{\rho(\Delta)}) \end{aligned}$$

Nous ne considérons ici qu'une sémantique de système K car elle fournit un cadre minimaliste, sans contrainte sur les relations d'accessibilité. Cependant, elle pourrait être étendue en contraignant les relations d'accessibilité.

Définition 7 (Sémantique des logiques multi-modales normales n -aires). *Soit $\tau = (\mathcal{O}, \rho)$ un type de similarité. Un τ -modèle est un N -uplet $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_\Delta\}_{\Delta \in \mathcal{O}}, \mathcal{V} \rangle$ tel que :*

- \mathcal{W} est un ensemble non vide de mondes,
- $\forall \Delta \in \mathcal{O}$, \mathcal{R}_Δ est une relation $(\rho(\Delta) + 1)$ -aire¹,
- $\mathcal{V} : \mathcal{P} \rightarrow 2^{\mathcal{W}}$ est une fonction de valuation.

$\mathcal{C} = \langle \mathcal{W}, \{\mathcal{R}_\Delta\}_{\Delta \in \mathcal{O}} \rangle$ est un τ -cadre et \mathcal{M} est un modèle dans \mathcal{C} ssi \mathcal{M} appartient à la classe des modèles $\mathcal{C}_{\mathcal{M}} = \{\mathcal{M} : \exists \mathcal{V}, \mathcal{M} = \langle \mathcal{C}, \mathcal{V} \rangle\}$. Pour tout $w \in \mathcal{W}$, et $\phi, \psi \in \mathcal{L}$, $\Delta \in \mathcal{O}$, $\varrho(\Delta) = (\Delta, \nabla, \diamond, \square)$, et $p \in \mathcal{P}$:

1. Chaque $\star \in \varrho(\Delta)$ est défini sur la même relation $(\rho(\Delta) + 1)$ -aire.

1. $\mathcal{M}, w \not\models \perp$
2. $\mathcal{M}, w \models p$ ssi $w \in V(p)$
3. $\mathcal{M}, w \models \neg\phi$ ssi $\mathcal{M}, w \not\models \phi$
4. $\mathcal{M}, w \models \phi \wedge \psi$ ssi $\mathcal{M}, w \models \phi$ et $\mathcal{M}, w \models \psi$

Pour les cas où $\rho(\Delta) \geq 1$:

5. $\mathcal{M}, w \models \Delta(\phi_1, \dots, \phi_{\rho(\Delta)})$ ssi $\exists(v_1, \dots, v_{\rho(\Delta)}) \in \mathcal{W}^{\rho(\Delta)} : w\mathcal{R}_\Delta(v_1, \dots, v_{\rho(\Delta)})$ et $\forall k \in \mathbb{N}^*, k \leq \rho(\Delta), v_k \models \phi_k$
6. $\mathcal{M}, w \models \nabla(\phi_1, \dots, \phi_{\rho(\Delta)})$ ssi $\forall(v_1, \dots, v_{\rho(\Delta)}) \in \mathcal{W}^{\rho(\Delta)} : si w\mathcal{R}_\Delta(v_1, \dots, v_{\rho(\Delta)})$ alors $\exists k \in \mathbb{N}^*, k \leq \rho(\Delta), v_k \models \phi_k$
7. $\mathcal{M}, w \models \Box(\phi_1, \dots, \phi_{\rho(\Delta)})$ ssi $\forall(v_1, \dots, v_{\rho(\Delta)}) \in \mathcal{W}^{\rho(\Delta)} : si w\mathcal{R}_\Delta(v_1, \dots, v_{\rho(\Delta)})$ alors $\forall k \in \mathbb{N}^*, k \leq \rho(\Delta), v_k \models \phi_k$
8. $\mathcal{M}, w \models \Diamond(\phi_1, \dots, \phi_{\rho(\Delta)})$ ssi $\exists(v_1, \dots, v_{\rho(\Delta)}) \in \mathcal{W}^{\rho(\Delta)} : w\mathcal{R}_\Delta(v_1, \dots, v_{\rho(\Delta)})$ et $\exists k \in \mathbb{N}^*, k \leq \rho(\Delta), v_k \models \phi_k$

Pour les cas où $\rho(\Delta) = 0$:

9. $\mathcal{M}, w \models \Delta$ ssi $w \in \mathcal{R}_\Delta$
10. $\mathcal{M}, w \models \nabla$ ssi $\mathcal{R}_\Delta = \mathcal{W}$
11. $\mathcal{M}, w \models \Box$ ssi $\mathcal{R}_\Delta = \emptyset$
12. $\mathcal{M}, w \models \Diamond$ ssi $\mathcal{R}_\Delta \neq \emptyset$

Pour résumer, nous avons quatre types d'opérateurs modaux : Δ est une modalité (\exists, \wedge)-type appartenant à l'ensemble \mathcal{O} , ∇ est de (\forall, \vee)-type appartenant à l'ensemble \mathcal{O}_{dual} , \Diamond est de (\exists, \vee)-type appartenant à l'ensemble \mathcal{O}^* et \Box est de (\forall, \wedge)-type appartenant à l'ensemble \mathcal{O}_{dual}^* . Par (\exists, \wedge)-type, nous entendons les modalités ayant une sémantique où nous avons un quantificateur \exists sur les N -uplets accessibles et une « conjonction » sur la vérité des formules par rapport à leurs mondes respectifs du N -uplet, tandis que les (\forall, \vee)-types ont une sémantique fondée sur le quantificateur \forall pour les mondes accessibles et une « disjonction » sur la vérité des formules par rapport à leur monde respectif du N -uplet. Pour le cas où $\rho(\Delta) = 1$, i.e. Δ est monadique, il y a équivalence entre opérateurs : $\Delta \equiv \Diamond$ et $\Box \equiv \nabla$.

Exemple 4. *Considérons la formule $O(crisis_{health}, \neg\Delta_{p.t.})$ où O est un opérateur déontique dyadique [21]. Cette formule peut signifier que dans le cas où il y a une crise sanitaire, il est fortement déconseillé de prendre les transports publics. Dans notre cadre, cet opérateur peut être exprimé sémantiquement avec un opérateur ∇ ayant donc une sémantique de (\forall, \vee)-type, i.e. $\nabla_O(\neg crisis_{health}, \neg\Delta_{p.t.})$ car le contexte doit être représenté avec une négation dans les modalités (Qt, \vee)-type où $Qt \in \{\forall, \exists\}$. En effet, toutes les modalités (Qt, \vee)-type considèrent une « disjonction » sur les mondes accessibles du N -uplet, e.g. $w \models \nabla(\neg\phi, \psi)$ est vrai ssi $\forall(u, v) \in \mathcal{R}_\Delta(w), u \models \neg\phi$ ou $v \models \psi$. Ainsi, nous pouvons traduire cette disjonction, de manière équivalente en si $u \models \phi$, alors $v \models \psi$.*

Rappelons les notions de validité et satisfiabilité d'une formule. ϕ est valide dans \mathcal{M} (noté $\mathcal{M} \models \phi$) si, et seulement si, pour tous les mondes $w \in \mathcal{W}$, ϕ est satisfiable dans w i.e. $\mathcal{M}, w \models \phi$ est vraie. Une formule ϕ est valide dans un cadre \mathcal{C} (noté $\models_{\mathcal{C}} \phi$ ou $\mathcal{C} \models \phi$) si, et seulement si, pour tous les modèles \mathcal{M} construits sur \mathcal{C} , $\mathcal{M} \models \phi$. Dans ce cas, ϕ est une tautologie de \mathcal{C} . Soit $\Gamma \subseteq \mathcal{L}$ un ensemble de formules, $\Gamma \models_{\mathcal{C}} \phi$ signifie que ϕ est une *conséquence sémantique* de Γ si, et seulement si, pour tout modèle \mathcal{M} dans \mathcal{C} , si $\mathcal{M} \models \Gamma$ i.e. $\mathcal{M} \models \bigwedge_{\psi \in \Gamma} \psi$ alors $\mathcal{M} \models \phi$.

Exemple 5. *Considérons le type de similarité $\tau = (\mathcal{O}, \rho)$ où $\mathcal{O} = Dyadic_\Delta \cup Unary \cup Nullary_\Delta$. Ici, $Dyadic_\Delta = \{\Delta_{gov}, \Delta_{env}, \Delta_{sport}, \Delta_{time}, \}$ représente les opérateurs dyadiques qui sont respectivement associés aux conseils des autorités, les recommandations environnementales, les recommandations pour la santé, et ce qui permet de favoriser le gain de temps. $Unary = \{\Diamond\}$ contient un opérateur unaire pour décrire ce qui est possible et $Nullary = \{\Delta_{bike}, \Delta_{p.t.}, \Delta_{walk}, \Delta_{car}\}$ sont les choix. Ainsi, la fonction ρ est telle que : $\forall \star \in Dyadic_\Delta, \rho(\star) = 2, \rho(\Diamond) = 1, \forall \star \in Nullary, \rho(\star) = 0$. Par conséquent, nous considérerons le τ -cadre $\mathcal{C} = \langle \mathcal{W}, \{\mathcal{R}_\Delta\}_{\Delta \in \mathcal{O} \setminus \{\Diamond\}}, \mathcal{R}_\Diamond \rangle$ où la sémantique de chaque opérateur est donnée par le type de symbole utilisé dans la formule et \mathcal{O}^Ω représente l'ensemble de tous les symboles pour les opérateurs modaux Δ, ∇ et leur alternative \Box et \Diamond .*

En étendant [9] aux modalités n-aires \Box , nous proposons un système axiomatique **fortement correct et complet**, appelé GK, qui généralise les cadres de Kripke. Nous notons $\vdash \phi$ pour signifier que ϕ est un théorème. Nous considérons les axiomes classiques du calcul propositionnel (CP), le modus ponens (MP) et la substitution uniforme (SUB). Pour tout $\Delta \in \mathcal{O}$, t.q. $\rho(\Delta) \geq 1, \varrho(\Delta) = (\Delta, \nabla, \Diamond, \Box)$.

La règle de nécessité (*NEC*) :

Pour $\vdash \phi : \vdash \nabla(\perp, \dots, \phi, \dots, \perp) \wedge \Box(\top, \dots, \phi, \dots, \top)$

Les axiomes $\forall k \in [1, \rho(\Delta)]$, (K_{∇}^k) :

$\vdash \nabla(\phi_1, \dots, \phi_k \Rightarrow \psi_k, \dots, \phi_{\rho(\Delta)})$
 $\Rightarrow \nabla(\phi_1, \dots, \phi_k, \dots, \phi_{\rho(\Delta)}) \Rightarrow \nabla(\phi_1, \dots, \psi_k, \dots, \phi_{\rho(\Delta)})$

Les axiomes $\forall k \in [1, \rho(\Delta)]$, (K_{\Box}^k) :

$\vdash \Box(\phi_1, \dots, \phi_k \Rightarrow \psi_k, \dots, \phi_{\rho(\Delta)})$
 $\Rightarrow \Box(\phi_1, \dots, \phi_k, \dots, \phi_{\rho(\Delta)}) \Rightarrow \Box(\phi_1, \dots, \psi_k, \dots, \phi_{\rho(\Delta)})$

Les axiomes d'interaction (*Int*(\Box, ∇)) :

$\vdash \Box(\phi_1, \dots, \phi_{\rho(\Delta)}) \Rightarrow \nabla(\phi_1, \dots, \phi_{\rho(\Delta)})$

Les axiomes de dualité pour (*Dual* $_{\nabla}$) et (*Dual* $_{\Box}$) :

$\vdash \nabla(\phi_1, \dots, \phi_{\rho(\Delta)}) \Leftrightarrow \neg\Delta(\neg\phi_1, \dots, \neg\phi_{\rho(\Delta)})$

$\vdash \Box(\phi_1, \dots, \phi_{\rho(\Delta)}) \Leftrightarrow \neg\Diamond(\neg\phi_1, \dots, \neg\phi_{\rho(\Delta)})$

Les règles d'équivalence (*RE*), pour tout $\star \in \varrho(\Delta)$:

Pour $\vdash \phi_k \Leftrightarrow \psi_k : \vdash \star(\dots, \phi_k, \dots) \Leftrightarrow \star(\dots, \psi_k, \dots)$

Nous devons aussi considérer deux cas pour les opérateurs monadiques et nullaires. Si $\rho(\Delta) = 1$, nous ajoutons aux axiomes précédents : $\vdash \Box\phi \Leftrightarrow \nabla\phi$ et $\vdash \Diamond\phi \Leftrightarrow \Delta\phi$. Si $\rho(\Delta) = 0$, les axiomes précédents ne tiennent plus et nous avons seulement (*PC*), (*MP*), (*RE*), (*SUB*).

Définition 8 (Système de déduction). Soit Σ un ensemble de formules et ϕ une formule de \mathcal{L} . ϕ est déductible de Σ , noté $\Sigma \vdash \phi$, si, et seulement si : si $\Sigma = \emptyset$, alors $\vdash \phi$, sinon $\exists n \in \mathbb{N}^*, \exists \psi_1, \dots, \psi_n \in \Sigma, \vdash \psi_1 \wedge \dots \wedge \psi_n \Rightarrow \phi$.

Exemple 6. Dans notre exemple, pour exprimer une généralisation de l'axiome de vérité, notée T_τ , pour les modalités de type \Box et \Diamond , nous supposons que \mathcal{R}_\Diamond est telle que :

$$\forall w \in \mathcal{W}, \mathcal{R}_\Diamond(w) = \{w\} \cup \bigcup_{\Delta' \in \text{Dyadic}} \{u, v : (u, v) \in \mathcal{R}_{\Delta'}(w)\}$$

Cette contrainte exprime la tautologie suivante :

$$\models \Box\phi \Rightarrow \left(\phi \bigwedge_{\Delta' \in \text{Dyadic}: \varrho(\Delta') = (\Delta', \nabla', \diamond', \Box')} (\Box'(\top, \phi)) \wedge \Box'(\phi, \top) \wedge \Box'(\phi, \phi) \wedge \nabla'(\perp, \phi) \wedge \nabla'(\phi, \perp) \wedge \nabla'(\phi, \phi) \right)$$

Elle signifie que si une formule est nécessairement vraie, alors elle est aussi vraie dans tous les mondes accessibles à partir des autres relations.

Exemple 7. Nous considérons les hypothèses suivantes :

$$\begin{aligned} \Gamma_1 &\stackrel{\text{def}}{=} \{\nabla_{aut}(\neg \text{crisis}_{health}, \neg \Delta_{p.t.})\} \\ \Gamma_2 &\stackrel{\text{def}}{=} \{\Delta_{car} \rightarrow \text{has}_{car}, \Delta_{bike} \rightarrow \text{has}_{bike}\} \\ \Gamma_3 &\stackrel{\text{def}}{=} \{\nabla_{health}(\perp, \Delta_{bike})\} \\ \Gamma_4 &\stackrel{\text{def}}{=} \{\nabla_{time}(\neg \text{emergency} \vee \text{congestion}, \Delta_{car})\} \\ \Gamma_5 &\stackrel{\text{def}}{=} \{\neg \Delta_{time}(\text{emergency}, \Delta_{walk})\} \\ \Gamma_6 &\stackrel{\text{def}}{=} \{\nabla_{env}(\perp, \Delta_{bike} \vee \Delta_{walk} \vee \Delta_{p.t.})\} \\ \Gamma_7 &\stackrel{\text{def}}{=} \{\nabla_{env}(\neg \text{congestion}, \neg \Delta_{car})\} \\ \Gamma_8 &\stackrel{\text{def}}{=} \{\Box \text{XOR}(\Delta_{car}, \Delta_{p.t.}, \Delta_{bike}, \Delta_{walk})\} \\ \Gamma_9 &\stackrel{\text{def}}{=} \{\Box(\text{has}_{car} \wedge \neg \text{has}_{bike} \wedge \text{crisis}_{health} \wedge \neg \text{congestion} \wedge \neg \text{emergency})\} \end{aligned}$$

Γ_1 signifie que les autorités recommandent de ne pas prendre les transports publics quand il y a une crise sanitaire. Γ_2 représente la condition, faire du vélo ou conduire implique d'avoir un vélo ou une voiture. Γ_3 signifie que faire du vélo est obligatoire dans tous les contextes du point de vue de la santé. Γ_4 signifie que s'il y a une urgence et qu'il n'y a pas d'embouteillage, alors il est obligatoire de prendre sa voiture pour économiser du temps. Γ_5 signifie qu'il n'est pas possible qu'en cas d'urgence, marcher vous fasse gagner du temps. Γ_6 signifie que peu importe le contexte, il est obligatoire de prendre son vélo, marcher ou prendre les transports publics pour le bien de l'environnement. Γ_7 signifie qu'il est obligatoire de ne pas prendre sa voiture quand il y a des embouteillages. Γ_8 signifie qu'il est nécessaire que l'agent choisisse un seul choix au même moment (e.g. l'agent ne peut pas à la fois prendre sa voiture et son vélo). Γ_9 décrit la situation initiale de l'agent.

Exemple 8. Nous montrons à partir de l'exemple 7 :

$$\begin{aligned} \Gamma_5 &\vdash \nabla_{time}(\neg \text{emergency}, \neg \Delta_{walk}) \\ \Gamma_2 &\vdash \neg \text{has}_{bike} \Rightarrow \neg \Delta_{bike} \\ \Gamma_2 \cup \Gamma_9 &\vdash \neg \Delta_{bike} \\ \Gamma_6 \cup \Gamma_8 &\vdash \nabla_{env}(\perp, \neg \Delta_{car}) \\ \Gamma_8 &\vdash \text{XOR}(\Delta_{car}, \Delta_{p.t.}, \Delta_{bike}, \Delta_{walk}) \\ \Gamma_9 &\vdash \neg \text{emergency} \vee \text{congestion} \end{aligned}$$

Remarquons que nous déduisons de Γ_6 et Γ_8 que, peu importe le contexte, il n'est pas bon pour l'environnement de prendre sa voiture, ce qui est plus fort que l'hypothèse Γ_7 .

5 VAF fondé sur les logiques modales

Nous définissons les arguments puis les relations d'attaque pour les VAF fondés sur les logiques modales.

5.1 Arguments

Dans la littérature, il est usuel de considérer un argument comme un couple (Γ, ϕ) où Γ est un ensemble de formules, nommées *prémises* ou *support*, et implique une formule ϕ , appelée *conclusion* [1, 4, 8, 10, 20]. Usuellement, ces arguments sont définis sur la base des *implications syntaxiques* en supposant que la logique est connue et bien construite. Ici, nous donnons une définition pour les arguments déduits à partir de notre système de preuve : un *argument valide* est un couple (Γ, ϕ) où la conclusion ϕ est une *conséquence sémantique* des prémisses de Γ .

Définition 9 (Argument valide). Soit $\tau = (\mathcal{O}, \rho)$ un type de similarité, et \mathcal{C} un τ -cadre. Un argument valide dans \mathcal{C} est un couple $A = (\Phi, \alpha)$ t.q. :

1. $\Phi \subseteq \Delta$ où $\Delta \subseteq \mathcal{L}$ est fini²,
2. $\Phi \not\models_{\mathcal{C}} \perp$ i.e. Φ n'est pas inconsistant : $\exists \mathcal{M}, w \models \Phi$,
3. $\Phi \models_{\mathcal{C}} \alpha$,
4. Φ est minimal pour \subseteq t.q. 1, 2 et 3 sont respectés.

Un argument est *déductible* quand il est obtenu à partir d'un système de preuve syntaxique, i.e. avec l'opérateur de déduction \vdash plutôt que \models . Il y a généralement équivalence entre argument déductible et argument valide dans un système logique correct et complet. Par abus de langage, nous écrivons $A = (_, \phi)$ comme un raccourci signifiant que l'ensemble des prémisses Φ n'est pas spécifié. Cela correspond à $\exists \Phi \subseteq \mathcal{L}$ t.q. $A = (\Phi, \phi)$.

Exemple 9. Quelques exemples d'arguments déductibles :

$$\begin{aligned} A_1 &= (\Gamma_1, \neg \text{congestion}) \\ A_2 &= (\Gamma_7, \nabla_{env}(\neg \text{congestion}, \neg \Delta_{car})) \\ A_3 &= (\Gamma_9, \Box \neg \text{congestion}) \\ A_4 &= (\Gamma_8, \Box \text{XOR}(\Delta_{car}, \Delta_{p.t.}, \Delta_{bike}, \Delta_{walk})) \\ A_5 &= (\Gamma_3, \nabla_{sport}(\perp, \Delta_{bike})) \\ A_6 &= (\Gamma_2 \cup \Gamma_9, \Box \neg \Delta_{bike}) \\ A_7 &= (\Gamma_4, \nabla_{time}(\neg \text{emergency} \vee \text{congestion}, \Delta_{car})) \\ A_8 &= (\Gamma_8, \neg \text{emergency} \vee \text{congestion}) \\ A_9 &= (\Gamma_5, \nabla_{time}(\neg \text{emergency}, \neg \Delta_{walk})) \\ A_{10} &= (\Gamma_9, \neg \text{emergency}) \\ A_{11} &= (\Gamma_6 \cup \Gamma_8, \nabla_{env}(\perp, \neg \Delta_{car})) \\ A_{12} &= (\Gamma_1, \nabla_{aut}(\neg \text{crisis}_{health}, \neg \Delta_{p.t.})) \end{aligned}$$

D'autres arguments peuvent être déduits (ex. $(\Gamma_9, \text{crisis}_{health})$) mais ne sont pas considérés car ils ne supportent aucune décision ou n'attaquent aucun argument supportant une décision.

2. Puisque il y a un nombre infini de fbf sur \mathcal{L} , il est usuel de considérer Δ un ensemble de fbf sur lesquelles les agents peuvent raisonner [8].

5.2 Attaques modales

Rappelons tout d'abord la sémantique des relations d'attaque les plus utilisées : le *rebuttal*, l'*undercut* et le *defeater* [2, 8].

Définition 10 (Attaques). $A = (\Phi, \alpha)$ est un rebuttal pour $B = (\Psi, \beta)$ (noté $A \text{ Reb } B$) ssi $\models \alpha \Leftrightarrow \neg \beta$.

A est un undercut pour B (noté $A \text{ Und } B$) ssi il existe :

$$\Psi' = \{\psi_1, \dots, \psi_n\} \subseteq \Psi \text{ t.q. } \models \alpha \Leftrightarrow \bigwedge_{\psi_i \in \Psi'} \neg \psi_i$$

A est un defeater pour B (noté $A \text{ Def } B$) ssi il existe :

$$\Psi' = \{\psi_1, \dots, \psi_n\} \subseteq \Psi \text{ t.q. } \alpha \models \bigwedge_{\psi_i \in \Psi'} \neg \psi_i$$

Cependant cela ne caractérise pas les attaques entre formules avec des opérateurs modaux différents car elles n'aboutissent pas nécessairement à des incohérences. Ainsi, nous proposons un nouveau type d'attaque pour toute logique modale, que nous appelons *attaque modale*. Considérons une formule déontique dyadique $\nabla_O(\neg\phi, \psi)$ signifiant que quand ϕ est vraie, il est obligatoire d'avoir ψ à vrai, et une formule monadique déontique $\Box_K\neg\psi$ signifiant qu'il est obligatoire pour une certaine doctrine que ψ soit au contraire fausse. Intuitivement une attaque entre deux arguments $C = (_, \nabla_O(\neg\phi, \psi))$ et $K = (_, \Box_K\neg\psi)$ serait fondée sur une contradiction entre le second paramètre de ∇_O et la formule contenue dans \Box_K . Cependant, C devrait être attaqué si ϕ est faux, i.e. le contexte n'est pas vérifié. Ainsi, un argument $F = (_, \neg\phi)$ serait un *defeater* de C . De plus, puisque l'opérateur \Box_K est un opérateur monadique et ∇_O est un opérateur dyadique, l'attaque entre K et C est fondée sur une contradiction entre le premier paramètre de \Box_K et le second paramètre de ∇_O . Nous disons ici que les paramètres sont *non alignés*. D'autre part, l'attaque entre F et C est fondée sur une **équivalence** entre $\neg\phi$ et le premier paramètre de ∇_O . Nous disons que ces paramètres sont *alignés*. Pour spécifier sur quels paramètres se fondent l'attaque et quelle est sa nature, nous définissons $d_{\star_1}^{*2}$ -mapping une fonction qui associe les paramètres de deux modalités distinctes \star_1 et \star_2 pour caractériser une attaque. Dans l'exemple précédent, $K = (_, \Box_K\psi_1)$ avec $\psi_1 := \neg\psi$, et $C = (_, \nabla_O(\phi_1, \phi_2))$ avec $\phi_1 := \neg\phi$ et $\phi_2 := \psi$, s'attaque mutuellement car $\models \phi_2 \equiv \neg(\psi_1)$. Ainsi, le d -mapping est défini comme $d_{\Box_K}^{\nabla_O} = \{(1, 2)\}$, i.e. le premier paramètre de \Box_K est associé avec le second paramètre de ∇_O et il doit y avoir une contradiction pour définir une relation d'attaque. Nous définissons aussi un $a_{\star_1}^{*2}$ -mapping qui est une fonction qui associe des paramètres devant être équivalents pour définir une attaque. Considérons le \Box standard et la formule $\Box\theta_1$ où $\theta_1 := \neg\phi$. Considérons $\nabla_O(\phi_1, \phi_2)$, θ_1 est nécessairement vraie, i.e. $\neg\phi$, il ne peut pas être possible que le contexte de $\nabla_O(\phi_1, \phi_2)$ soit vérifié, i.e. ϕ est vraie. Ainsi, le a -mapping est défini comme $a_{\Box}^{\nabla_O} = \{(1, 1)\}$ et est défini lorsque $\models \phi_1 \equiv \theta_1$ est une tautologie. S'il pourrait sembler contre intuitif de considérer une équivalence pour une attaque, cela fait sens pour les modalités de type (\forall, \vee) et (\exists, \vee) , i.e. ∇ et \diamond , comme elles

représentent le contexte avec une négation. Enfin, nous définissons un $d_{\star_0}^{*2}$ -mapping (resp. $a_{\star_0}^{*2}$ -mapping) correspondant aux paramètres de \star_2 qui doivent être incohérents (reps. équivalents) par rapport aux vérités factuelles, i.e. un argument dont la conclusion n'est pas précédée par un opérateur modal. Par exemple, $d_{\star_0}^{\nabla_O} = \{1\}$ signifie que le paramètre contextuel de l'opérateur dyadique déontique ∇_O peut être attaqué par une vérité factuelle. Pour représenter ces mapping de manière compacte, nous utilisons un ensemble d'ensembles de règles, où les règles décrivent quels paramètres attaquent quels paramètres pour n'importe quel type de mapping. Par exemple, un $\langle d, a \rangle_{\star_1}^{*2}$ -mapping peut être $\{\{(1, 2, d), (2, 3, a)\}, \{(3, 4, d)\}\}$. Ici, $\langle d, a \rangle_{\star_1}^{*2}$ peut être vue comme une formule normale conjonctive (FNC) définissant les règles devant être satisfaites pour avoir une attaque, i.e. il y a une attaque si le paramètre 1 dans \star_1 et le paramètre 2 dans \star_2 se contredisent, ou le paramètre 2 dans \star_1 et le paramètre 3 dans \star_2 sont équivalents, et le paramètre 3 dans \star_1 et le paramètre 4 dans \star_2 se contredisent.

Définition 11 ($\langle d, a \rangle$ -mappings modaux). Soient $\tau = (\mathcal{O}, \rho)$, \mathcal{O}^Ω un ensemble de modalités de τ , $\{d, a\}$ un ensemble de symboles, et $(\star_1, \star_2) \in \mathcal{O}^\Omega \times \mathcal{O}^\Omega$ deux modalités. $\langle d, a \rangle_{\star_1}^{*2}$ est un mapping modal pour (\star_1, \star_2) ssi : $\langle d, a \rangle_{\star_1}^{*2} \subseteq 2^{[|1, \rho(\star_1)|] \times [|\rho(\star_2)|] \times \{d, a\}}$. $\langle d, a \rangle_{\star_0}^{*2}$ est un mapping modal pour \star_2 ssi : $\langle d, a \rangle_{\star_0}^{*2} \subseteq 2^{[|1, \rho(\star_2)|] \times \{d, a\}}$. $\langle d, a \rangle_\tau$ est un mapping modal pour le τ -cadre, une fonction qui associe pour chaque paire de modalités $(\star_1, \star_2) \in \mathcal{O}^\Omega \times \mathcal{O}^\Omega$ un mapping modal pour (\star_1, \star_2) et pour chaque modalité $\star_2 \in \mathcal{O}^\Omega$ assigne un mapping modal pour \star_2 , i.e. $\langle d, a \rangle_\tau$ est tel que :

$$\begin{aligned} & - \langle d, a \rangle_\tau : (\mathcal{O}^\Omega \cup \{\emptyset\}) \times \mathcal{O}^\Omega \rightarrow 2^{2^{((\mathbb{N}^* \times \mathbb{N}^*) \cup \mathbb{N}^*) \times \{d, a\}}} \\ & - \forall (\star_1, \star_2) \in \mathcal{O}^\Omega \times \mathcal{O}^\Omega, \langle d, a \rangle_\tau(\star_1, \star_2) \subseteq 2^{[|1, \rho(\star_1)|] \times [|\rho(\star_2)|] \times \{d, a\}} \\ & - \forall \star_2 \in \mathcal{O}^\Omega, \langle d, a \rangle_\tau(\emptyset, \star_2) \subseteq 2^{[|1, \rho(\star_2)|] \times \{d, a\}} \end{aligned}$$

Dans la suite, nous considérons les mapping modaux *consistants*. En effet, si pour deux modalités nous avons $\langle d, a \rangle_{\star_1}^{*2} = \{\{(i, j, d)\}, \{(i, j, a)\}\}$, alors cela signifierait que les paramètres devraient être à la fois incohérents et équivalents, ce qui n'a pas de sens.

Définition 12 (Mapping modal consistant). Soit $\tau = (\mathcal{O}, \rho)$ un type de similarité, \mathcal{O}^Ω l'ensemble des symboles de τ , et $\langle d, a \rangle_\tau$ un mapping modal défini sur le τ -cadre. Nous appelons $\langle d, a \rangle_\tau$ un mapping modal consistant ssi (1) et (2) sont vérifiés :

(1) Pour tout $\star_1, \star_2 \in \mathcal{O}^\Omega$, il existe une fonction $I : [|\rho(\star_1)|] \times [|\rho(\star_2)|] \rightarrow \{\top, \perp\}$ t.q. :

$$\forall S \in \langle d, a \rangle_\tau(\star_1, \star_2), \exists (i, j, t) \in S, I(i, j) = \begin{cases} \top, & \text{if } t = a \\ \perp, & \text{if } t = d \end{cases}$$

(2) Pour tout $\star_2 \in \mathcal{O}^\Omega$, il existe une fonction $I : [|\rho(\star_2)|] \rightarrow \{\top, \perp\}$ t.q. :

$$\forall S \in \langle d, a \rangle_\tau(\emptyset, \star_2), \exists (j, t) \in S, I(j) = \begin{cases} \top, & \text{if } t = a \\ \perp, & \text{if } t = d \end{cases}$$

Vérifier si un mapping modal est consistant est un problème NP-complet car cela revient à résoudre un problème de satisfiabilité pour la logique propositionnelle. La preuve de NP-complétude découle du théorème suivant :

Théorème 1. Soit $\langle d, a \rangle_\tau$ un mapping modal consistant ssi la formule propositionnelle suivante ϕ est satisfiable, avec $\mathcal{P} = \{p_{i,j} : (i, j) \in [1, \rho(\star_1)] \times [1, \rho(\star_2)], (\star_1, \star_2) \in (\mathcal{O}^\Omega)^2\} \cup \{p_j : j \in [1, \rho(\star_2)], \star_2 \in \mathcal{O}^\Omega\}$ un ensemble d'atomes propositionnels :

$$\phi = \bigwedge_{(\star_1, \star_2) \in \mathcal{O}^\Omega \times \mathcal{O}^\Omega} \left(\bigwedge_{S \in \langle d, a \rangle_{\star_1}^{\star_2}} \left(\bigvee_{\substack{(i,j,t) \in S: \\ t=a}} p_{i,j} \bigvee_{\substack{(i,j,t) \in S: \\ t=d}} \neg p_{i,j} \right) \right. \\ \left. \bigwedge_{\star_2 \in \mathcal{O}^\Omega} \left(\bigwedge_{S \in \langle d, a \rangle_0^{\star_2}} \left(\bigvee_{(j,t) \in S: t=a} p_j \bigvee_{(j,t) \in S: t=d} \neg p_j \right) \right) \right)$$

La preuve est évidente en remarquant qu'elle est une traduction directe d'un $(i, j)_{\star_1}^{\star_2}$ variable propositionnelle $p_{i,j}$ quand il s'agit d'un mapping de type aligné et la négation $\neg p_{i,j}$ quand il est de type non aligné. Nous faisons de même pour les mappings de la forme $(j)_{\star_0}^{\star_2}$. Prouver la réciproque correspond à construire un bon modal mapping correspondant à une formule ϕ en FNC.

Exemple 10. Soient un mapping non aligné entre chaque second paramètre des opérateurs dyadiques et un mapping aligné pour chaque premier paramètre des opérateurs dyadiques et les formules factuelles et les formules nécessairement vraies. Ainsi, $\langle d, a \rangle_\tau$ est tel que :

$$\begin{aligned} & - \forall \star_1, \star_2 \in \text{Dyadic}_\nabla, \langle d, a \rangle_\tau(\star_1, \star_2) = \{\{(2, 2, d)\}\} \\ & - \forall \star_2 \in \text{Dyadic}_\nabla, \langle d, a \rangle_\tau(\emptyset, \star_2) = \{\{(1, a)\}\} \\ & - \forall \star_2 \in \text{Dyadic}_\nabla, \langle d, a \rangle_\tau(\square, \star_2) = \{\{(1, 1, a)\}\} \end{aligned}$$

En raison du nombre d'interactions entre modalités et mappings, décrire chaque attaques est fastidieux. Nous définissons donc une fonction Υ qui exprime des règles pour décrire les attaques. Nous avons les règles pour les attaques entre deux modalités, et les règles pour les faits et les modalités. De plus, un paramètre dans Υ sert à décrire le type, distinguant rebuttals et undercuts (\Leftrightarrow), et defeaters (\Rightarrow).

Définition 13 (Règle Υ -attaque pour $\langle d, a \rangle$ -mapping). Soient $\tau = (\mathcal{O}, \rho)$ un type de similarité, \mathcal{O}^Ω un ensemble de symboles fondés sur τ , $\langle d, a \rangle$ un mapping modal sur le τ -cadre, et $\Upsilon = \{\Upsilon_{\star_1}^{\star_2}\}_{(\star_1, \star_2) \in M}$ un ensemble de fonctions où $M = (\mathcal{O}^\Omega \cup \{\emptyset\}) \times \mathcal{O}^\Omega$:

$$\forall (\star_1, \star_2) \in M, \Upsilon_{\star_1}^{\star_2} : \mathcal{L} \times \mathcal{L} \times \{\Rightarrow, \Leftrightarrow\} \rightarrow \{\perp, \top\}$$

Pour tout $(\phi, \psi) \in \mathcal{L}^2$, pour tout $(\star_1, \star_2) \in \mathcal{O}^\Omega \times \mathcal{O}^\Omega$. Υ est l'ensemble des règles d'attaque pour le $\langle d, a \rangle$ -mapping ssi Υ est t.q. (1) et (2) sont respectées :

1. si $\phi = \star_1(\phi_1, \dots, \phi_{\rho(\star_1)})$, et pour tout $i \in [1, \rho(\star_1)]$, $\phi_i \equiv \phi_i^1 \wedge \dots \wedge \phi_i^{n_i}$ sa forme canonique normale conjonctive (FCNC), $\Upsilon_{\star_1}^{\star_2}(\phi, \psi, \text{type}) = \top$ si, et seulement si, $\forall S \in \langle d, a \rangle_{\star_1}^{\star_2}, \exists (i, j, t) \in S, \exists K \subseteq [1, n_i]$:
 - si $t = d$, alors :

$$\left\{ \bigwedge_{k \in K} \phi_i^k \right\} \models \neg \psi_j \text{ et (si type } \Leftrightarrow, \{\neg \psi_j\} \models \bigwedge_{k \in K} \phi_i^k)$$

- si $t = a$, alors :

$$\left\{ \bigwedge_{k \in K} \phi_i^k \right\} \models \psi_j \text{ et (si type } \Leftrightarrow, \{\neg \psi_j\} \models \bigwedge_{k \in K} \phi_i^k)$$

2. si $\phi \equiv \Phi$ où $\Phi = \phi^1 \wedge \dots \wedge \phi^n$ sa FCNC de ϕ , et $\psi = \star_2(\psi_1, \dots, \psi_{\rho(\star_2)})$, alors $\Upsilon_{\star_0}^{\star_2}(\phi, \psi, \text{type}) = \top$ si, et seulement si, $\forall S \in \langle d, a \rangle_{\star_0}^{\star_2}, \exists (j, t) \in S, \exists K \subseteq [1, n]$:

- si $t = d$, alors :

$$\left\{ \bigwedge_{k \in K} \phi^k \right\} \models \neg \psi_j \text{ et (if type } \Leftrightarrow, \{\neg \psi_j\} \models \bigwedge_{k \in K} \phi^k)$$

- si $t = a$, alors :

$$\left\{ \bigwedge_{k \in K} \phi^k \right\} \models \psi_j \text{ et (if type } \Leftrightarrow, \{\psi_j\} \models \bigwedge_{k \in K} \phi^k)$$

Nous pouvons désormais définir les attaques modales standards de l'argumentation fondée sur les logiques modales.

Définition 14 (Types d'attaques modales). Soit $\tau = (\mathcal{O}, \rho)$ un type de similarité, et $\langle d, a \rangle$ un mapping modal consistant sur le τ -cadre, et $(\star, \star_2) \in (\mathcal{O}^\Omega \cup \{\emptyset\}) \times \mathcal{O}^\Omega$,

Un argument A est un $\langle d, a \rangle_{\star}^{\star_2}$ rebuttal modal direct pour B , noté $(A, B) \in \text{Reb} \langle d, a \rangle_{\star}^{\star_2}$ si, et seulement si, $A = (_, \phi)$, $B = (_, \star_2(\psi_1, \dots, \psi_{\rho(\star_2)}))$ et $\Upsilon_{\star}^{\star_2}(\phi, \star_2(\psi_1, \dots, \psi_{\rho(\star_2)}), \Leftrightarrow) = \top$.

Un argument A est $\langle d, a \rangle_{\star}^{\star_2}$ undercut modal direct pour B , noté $(A, B) \in \text{Und} \langle d, a \rangle_{\star}^{\star_2}$ si, et seulement si, $A = (_, \phi)$, $B = (\Psi, _)$, il existe $\psi \in \Psi$, $\psi = \star_2(\psi_1, \dots, \psi_{\rho(\star_2)})$, $\Upsilon_{\star}^{\star_2}(\phi, \star_2(\psi_1, \dots, \psi_{\rho(\star_2)}), \Leftrightarrow) = \top$.

Un argument A est un defeater modal direct $\langle d, a \rangle_{\star}^{\star_2}$ pour B , noté $(A, B) \in \text{Def} \langle d, a \rangle_{\star}^{\star_2}$ si, et seulement si, $A = (_, \phi)$, $B = (\Psi, _)$, il existe $\psi \in \Psi$, $\psi = \star_2(\psi_1, \dots, \psi_{\rho(\star_2)})$, $\Upsilon_{\star}^{\star_2}(\phi, \star_2(\psi_1, \dots, \psi_{\rho(\star_2)}), \Rightarrow) = \top$.

Nous définissons une $\langle d, a \rangle_{\star}^{\star_2}$ attaque modale comme :

$$\text{Att} \langle d, a \rangle_{\star}^{\star_2} = \text{Reb} \langle d, a \rangle_{\star}^{\star_2} \cup \text{Und} \langle d, a \rangle_{\star}^{\star_2} \cup \text{Def} \langle d, a \rangle_{\star}^{\star_2}$$

Les attaques sur les opérateurs nullaires sont considérées comme pour les formules factuelles. En effet, les opérateurs nullaires sont vrais dans des mondes spécifiques quelque soit les modèles du cadre. Nous pouvons donc les considérer comme des formules propositionnelles, i.e. factuelles.

5.3 VAF fondés sur les logiques modales

Nous appelons τ -VAF un VAF fondé sur le système GK.

Définition 15 (AF fondé sur la logique modale). Soit $\tau = (\mathcal{O}, \rho)$. Un τ -AF est un tuple $\langle \mathcal{C}, \langle d, a \rangle, \mathcal{A}, \mathcal{R} \rangle$ où \mathcal{C} est un τ -cadre, et $\langle \mathcal{A}, \mathcal{R} \rangle$ est un AF t.q. \mathcal{A} contient seulement des arguments valides sur \mathcal{C} , $\langle d, a \rangle$ est un mapping modal consistant, et $\mathcal{R} = \text{Reb} \cup \text{Und} \cup \text{Def} \bigcup_{(\star, \star_2) \in (\mathcal{O}^\Omega \cup \{\emptyset\}) \times \mathcal{O}^\Omega} \text{Att} \langle d, a \rangle_{\star}^{\star_2}$.

Nous appelons un τ -VAF un N -uplet $\langle \mathcal{C}, \langle d, a \rangle, \mathcal{A}, \mathcal{R}, \text{Val}, \mu, \succ \rangle$ qui est un τ -AF et $\langle \mathcal{A}, \mathcal{R}, \text{Val}, \mu, \succ \rangle$ est un VAF.

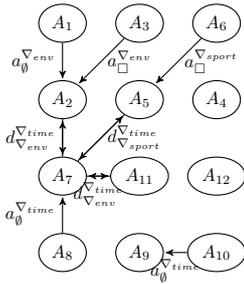


FIGURE 1 – Graphe d’argumentation

Exemple 11. Notre exemple est décrit par le τ -VAF de la figure 1. Les attaques modales sont données avec leur type de mapping modal. Nous supposons μ est t.q. pour tout $\star \in \{\square, \diamond, \nabla, \Delta\}$ et $\forall \phi, \psi \in \mathcal{L}$: (1) $\mu(_, \star_{sport}(\phi, \psi)) = \nu_{health}$; (2) $\mu(_, \star_{env}(\phi, \psi)) = \nu_{env}$; (3) $\mu(_, \star_{time}(\phi, \psi)) = \nu_{eco}$; (4) $\forall \phi, \psi \in \mathcal{L}$, $\mu(_, \star_{gov}(\phi, \psi)) = \nu_{aut}$; (5) $\mu(_, \square\phi) = \emptyset$; (6) $\forall A \in \mathcal{A}$ qui n’est pas de la forme $(_, \star(\dots))$, $\mu(A) = \emptyset$. Ainsi, $\mu = \{(A_1, \emptyset), (A_2, \nu_{env}), (A_3, \emptyset), (A_4, \emptyset), (A_5, \nu_{health}), (A_6, \emptyset), (A_7, \nu_{eco}), (A_8, \emptyset), (A_9, \nu_{env}), (A_{10}, \emptyset), (A_{11}, \nu_{env}), (A_{12}, \nu_{aut})\}$. Considérons ensuite les préférences suivantes : $\emptyset \succ \nu_{health} \succ \nu_{aut} \succ \nu_{env} \succ \nu_{eco}$. Ainsi, certains arguments comme A_7 sont défait par A_2, A_5, A_{11} et sont associés à une valeur préférée. La VAF-attaque = $\{(A_1, A_2), (A_3, A_2), (A_6, A_5), (A_5, A_7), (A_2, A_7), (A_{11}, A_7), (A_8, A_7), (A_{10}, A_9)\}$. Enfin, considérons l’extension stable $\mathcal{E} : \mathcal{E} = \{(A_1, A_3, A_4, A_6, A_8, A_{10}, A_{11}, A_{12})\}$. Le résultat conclut que l’utilisateur ne peut pas prendre son vélo puisqu’il n’a pas de vélo. Puisqu’il n’y a pas d’embouteillage et pas d’urgence, l’utilisateur n’a pas besoin de prendre sa voiture pour réduire son temps de trajet. Pour l’environnement, il est meilleur de prendre les transports publics ou marcher. Cependant, puisqu’il y a une crise sanitaire, il est recommandé de ne pas prendre les transports publics. Ainsi, la décision éthique est marcher.

Même si nous n’explicitons pas le modèle de décision, nous pourrions utiliser une fonction \mathcal{F}_f (resp. \mathcal{F}_c) assignant une décision à un ensemble d’arguments qui la supporte (resp. ne la supporte pas) [3] (e.g. une décision peut être éthique si elle est supportée crêdulement, i.e. au moins un argument supporte la décision et appartient à au moins une extension).

6 Conclusion

Nous avons proposé de modéliser le raisonnement éthique avec un cadre d’argumentation valué, appelé τ -VAF, fondé sur une logique modale normale généralisée aux opérateurs n-aires (pour raisonner avec des opérateurs déontiques dyadiques). Cela nous a conduit à proposer de nouveaux types d’attaques, appelées les attaques modales, qui généralisent les attaques logiques standard. Les travaux futurs consisteront à instancier un τ -VAF avec une logique modale comme DL-MA [19] ou avec une logique de description déontique comme celle proposée dans [13].

Références

- [1] L. Amgoud. Postulates for logic-based argumentation systems. *IJAR*, 55(9) :2028–2048, 2014.
- [2] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *4th SUM*, pages 42–55, 2010.
- [3] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *AI*, 173(3) :413–436, 2009.
- [4] O. Arieli and C. Straßer. Sequent-based logical argumentation. *A & C*, 6(1) :73–99, 2015.
- [5] K. Atkinson and J. M. Bench-Capon. Addressing moral problems through practical reasoning. *J. Appl. Log.*, 6(2) :135–151, 2008.
- [6] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *JLC*, 13(3) :429–448, 2003.
- [7] P. Besnard and A. Hunter. Practical first-order argumentation. In *20th AAAI*, pages 590–595, 2005.
- [8] P. Besnard and A. Hunter. Argumentation based on classical logic. In Guillermo Simari and Iyad Rahwan, editors, *Argumentation in artificial intelligence*, pages 133–152. Springer, 2009.
- [9] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic : graph. Darst*, volume 53. Cambridge University Press, 2002.
- [10] A.-M. Borg and O. Arieli. Hypersequential argumentation frameworks : An instantiation in the modal logic S5. In *17th AAMAS*, pages 1097–1104, 2018.
- [11] M. Caminada and D. Gabbay. A logical account of formal argumentation. *Stud. Log.*, 93(2-3) :109, 2009.
- [12] N. Cointe, G. Bonnet, and O. Boissier. Ethical judgment of agents’ behaviors in multi-agent systems. In *15th AAMAS*, pages 1106–1114, 2016.
- [13] T. Dalmonte, A. Mazzullo, and A. Ozaki. On non-normal modal description logics. In *32nd DL*, page 14, 2019.
- [14] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AI*, 77(2) :321–357, 1995.
- [15] J. Greene and J. Haidt. How (and where) does moral judgment work? *Trends Cogn. Sci.*, 6(12) :517–523, 2002.
- [16] D. Grossi. On the logic of argumentation theory. In *9th AAMAS*, pages 409–416, 2010.
- [17] J. Haidt. The emotional dog and its rational tail : a social intuitionist approach to moral judgment. *Psychol. Rev.*, 108(4) :814–834, 2001.
- [18] J. F. Horty. Moral dilemmas and nonmonotonic logic. *J. Philos. Log.*, 23(1) :35–65, 1994.
- [19] E. Lorini. On the logical foundations of moral agency. In *11th DEON*, pages 108–122, 2012.
- [20] H. Prakken. An abstract framework for argumentation with structured arguments. *A & C*, 1(2) :93–124, 2010.
- [21] H. Prakken and M. Sergot. Dyadic deontic logic and contrary-to-duty obligations. In *Defeasible deontic logic*, pages 223–262. Springer, 1997.
- [22] C. Proietti, D. Grossi, S. Smets, and F. R. Velázquez-Quesada. Bipolar argumentation frameworks, modal logic and semantic paradoxes. In *7th LORI*, pages 214–229, 2019.
- [23] M. Timmons. *Moral theory : an introduction*. Rowman & Littlefield Publishers, 2012.