Une logique modale normale de la confiance

Christopher Leturc **Grégory Bonnet**

Normandie Université, UNICAEN, GREYC, CNRS UMR 6072, France
{prénom.nom}@unicaen.fr

Résumé

Nous proposons dans cet article une logique modale permettant de raisonner sur la confiance d'un agent envers le discours d'un autre agent. Notre modèle considère une modalité de confiance doxastique similaire à celle initialement introduite par Dundua et Uridia. En combinant cette formulation générale avec un formalisme STIT, nous exprimons aussi la confiance qu'un agent accorde à un autre lorsque ce dernier annonce qu'il veille à réaliser un énoncé. Nous retrouvons alors la notion de confiance dispositionnelle formalisée par Herzig *et al.* et caractérisons certaines propriétés qui avaient été axiomatisées par Singh. Enfin, nous montrons que notre modèle permet d'exprimer des notions de confiance collective.

Abstract

We propose in this paper a modal logic to reason on the trust of an agent towards a statement formulated by another agent. Our model considers a doxastic trust modality inspired from the work of Dundua and Uridia. We combine this modality with a STIT formalism that allows us to express that an agent trust another one when this latter states it will sees to it that a statement is true. Furthermore we can express a dispositional trust like Herzig *et al.* and we show that our formalism allows to deduce some axioms given by Singh. Finally, we propose some ways to express collective trust.

1 Introduction

Dans le domaine des systèmes multi-agents, une attention particulière a été accordée à la notion de confiance. En effet, au sein d'un tel système, les agents doivent coopérer les uns avec les autres pour satisfaire leurs buts. Toutefois, tous les agents ne sont pas nécessairement fiables ou coopératifs et une des techniques privilégiées pour déterminer si un agent est de confiance ou non est l'utilisation de systèmes de réputation [17]. Un système de réputation agrège des valeurs de *confiance* individuelles entre les agents afin de définir une valeur collective appelée *réputation*.

Dans la littérature, si de nombreuses approches s'appuient sur une représentation statistique de la confiance [11, 12, 16], d'autres travaillent à représenter les aspects socio-cognitifs de la confiance [2]. Certains travaux reposent sur des logiques floues [8, 13, 21] et introduisent une notion subjective de degré de confiance et de vraisemblance d'une source d'information. D'autres travaux s'intéressent à une confiance modélisée par des logiques modales [7, 10, 18, 19]. Ces logiques modales permettent d'exprimer la confiance par le biais d'une ou plusieurs modalités comme une intention, une croyance, une action ou un but qu'un agent possède. Si ces approches permettent d'exprimer facilement certains aspects de la confiance comme *la délégation*, elles s'intéressent principalement à la confiance dans l'action des autres agents. Or, dans le cadre des systèmes de réputation, les agents sont amenés à communiquer avec d'autres pour les informer, par exemple, de la qualité des services proposés par des agents tiers. Si certains travaux sont plutôt consacrés aux aspects de révision des connaissances fondés sur la confiance [15, 9], d'autres sont consacrés à la modélisation de la confiance qu'exprime un agent envers le discours d'un autre agent [4, 6, 14, 5].

Dans cet article, nous proposons une logique modale de la confiance permettant d'exprimer *la confiance accordée par un agent i envers un énoncé ϕ proposé par un autre agent j*. Cette formulation est générale car, en combinant notre modèle avec un formalisme STIT, nous pouvons aussi exprimer la confiance qu'un agent accorde à un autre lorsque ce dernier annonce qu'il veille à réaliser une action $[stit]_j\phi$. Nous montrons alors que nous retrouvons une notion de confiance dispositionnelle comme [10, 18] puis nous étendons notre formalisme à la confiance de groupe modélisant une notion de réputation. Cet article est structuré comme suit. Dans la section 2, nous présentons les logiques modales de la confiance et, dans la section 3, le formalisme STIT [1, 3] sur lequel repose notre approche. Notre axiomatique de la confiance est présentée en section 4 et nous montrons quelques propriétés de la confiance en section 5. Enfin, nous présentons en section 6 la confiance collective.

2 Logiques modales de la confiance

Castelfranchi et Falcone [2] ont étudié la hiérarchie des différents composants fondamentaux de la confiance mais aussi ses aspects dynamiques, en particulier dans le cadre de la décision, de la construction d'intentions, de l'acte de faire confiance en soi ou de s'autoriser à déléguer des actions. Modéliser la confiance dans son entièreté est très complexe et la plupart des approches logiques se restreignent à des aspects précis de la confiance. Nous distinguons ici deux types d'approches : les logiques modales qui utilisent un prédicat pour représenter la confiance et les logiques qui s'appuient sur une modalité de confiance.

2.1 La confiance comme un prédicat

Herzig *et al.* [10] considèrent la confiance comme un prédicat signifiant qu'un agent i fait confiance à un autre agent j par rapport à une action α aboutissant à une proposition ϕ, si et seulement si toutes les propriétés suivantes sont vraies :

1. i a le but que ϕ soit réalisé,
2. i croit que :
   (a) j est capable de réaliser l'action α,
   (b) j, après avoir réalisé α assurera ϕ,
   (c) j a l'intention de faire α.

Ceci permet de définir un prédicat de *confiance immédiate* [1]. Cette notion traduit un aspect de la confiance dans le présent, à savoir qu'un agent j s'apprête bien à réaliser l'action pour laquelle i lui fait confiance. Une seconde notion de confiance, la *confiance dispositionnelle* [2], est la confiance accordée par un agent i à un agent j que cet agent j réalisera le but ϕ de i dans un contexte spécifique.

Smith *et al.* [19] considèrent aussi une confiance immédiate signifiant qu'un agent i fait confiance à un autre agent j pour ϕ si, et seulement si toutes les propriétés suivantes sont vraies :

1. i a le but ϕ,
2. i croit que j réalise ϕ ,
3. i a l'intention que :
   (a) j réalise ϕ,
   (b) i ne fasse pas ϕ.
4. i a le but que j ait l'intention de faire ϕ,
5. i croit que j a l'intention de ϕ.

1. Occurrent trust.
2. Dispositional trust.

Remarquons que ces deux manières de définir la confiance ne suffisent pas pour considérer, par exemple, les mécanismes de transmission de la confiance car, ici, un agent ne peut pas avoir confiance dans *la confiance d'autrui*. De plus, nous ne pouvons pas caractériser un autre aspect de la confiance qui est la *confiance dans le discours d'un agent*. En effet, supposons qu'un agent j communiquant ψ à un agent i, notée $\alpha_{j,i}$, a pour conséquence $\phi = B_i\psi$ sur l'agent i. Si nous appliquons la définition de Herzig *et al.*, dire que i fait confiance à j pour l'action $\alpha_{j,i}$ implique que i a pour objectif de croire la proposition ψ. Cela exprime le fait que l'agent a confiance parce qu'il croit mais ne permet pas d'exprimer qu'il croit parce qu'il a confiance. Des travaux comme ceux de Christianson et Harbison [4] ou Demolombe [6] ont proposés de capturer d'autres aspects de la confiance en définissant des propriétés qui doivent être satisfaites pour voir apparaître une confiance. Par exemple, Demolombe modélise la confiance en la sincérité (un agent a confiance en la sincérité de son interlocuteur s'il croit que ce dernier croit son discours), la confiance en la crédibilité ou l'honnêteté [6]. Christianson et Harbison ont appliqué cette approche à la confiance en un discours [4]. Par exemple, leur confiance en l'honnêteté se modélise comme la confiance en la sincérité de Demolombe. Ainsi, il existe plusieurs aspects de la confiance pouvant être modélisés de façon indépendante. Dans cet article, nous considérons de manière abstraite la confiance en un énoncé produit par un autre agent sans pour autant préciser de quel aspect il s'agit (cela dépend uniquement de l'application considérée).

2.2 La confiance comme une modalité

Une modalité de confiance d'un agent i envers un agent j peut être considérée soit relativement à un contexte [18], soit relativement à des croyances sur l'état du monde [7] ou encore comme un moyen pour inférer de la croyance [14, 5].

La première approche, proposée par Singh [18], exprime une confiance dispositionnelle par l'intermédiaire d'une modalité $T_{i,j}(\phi, \psi)$ signifiant que, dans un contexte ϕ, un agent i fait confiance à un autre agent j pour réaliser ψ. Si ϕ est vraie, la confiance de l'agent i envers j est active. Une forme de confiance immédiate peut alors être exprimée par $T_{i,j}(\top, \psi)$ signifiant qu'à tout instant (et donc dans l'instant présent) i fait confiance à j pour réaliser ψ. Singh utilise pour cela une vingtaine d'axiomes. Par exemple, si ψ est déjà vraie alors l'agent i n'a pas confiance en j pour que, dans le contexte ϕ, ψ soit vraie.

La seconde approche, proposée par Dundua et Uridia [7], introduit une modalité $T_{i,j}$ représentant la confiance comme une forme de croyance qu'un agent entretient sur un autre agent. La confiance est alors axiomatisée au regard d'une modalité de croyance comme par exemple :

$T_{i,j}\phi \Leftrightarrow B_i T_{i,j}\phi$.

Toutefois, ces travaux n'expriment que la capacité d'un agent à effectuer une tâche [18] ou encore le lien qu'il peut exister entre une notion de confiance dispositionnelle ou immédiate et la croyance d'un agent à bien effectuer l'action [7]. De plus, ces dernières approches ne s'intéressent pas aux discours des agents. Les travaux de Dastani *et al.* et Liau se sont intéressés à l'influence de la confiance dans l'assimilation de nouvelles informations au travers du formalisme BIT [5, 14]. Ce formalisme est une logique non-normale qui propose une modalité de confiance $T_{i,j}$ entre deux agents, en y incorporant la notion de thème (ou *topics*) modélisant une transitivité entre les confiances accordées à un discours relevant du même thème. Cette approche est pertinente mais, en tant que logique non-normale, elle se passe de certaines propriétés, comme $T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\phi$, nécessaires pour exprimer les propriétés intuitives des confiances dispositionnelles.

Nous proposons alors une *logique normale* avec une modalité permettant de représenter *une confiance abstraite dans un énoncé produit par des agents*. Par exemple, un agent peut avoir confiance en un autre lorsque celui-ci déclare qu'il *veille à ce que quelque chose soit vrai lorsqu'un contexte est vérifié* ou encore qu'il a *confiance dans l'énoncé d'un tiers*. Cette formulation générale nous permet d'exprimer une notion de confiance dispositionnelle mais aussi les aspects doxastiques de la confiance. Comme certains aspects de la notion de confiance peuvent s'appuyer sur la notion d'action, nous considérons le formalisme STIT.

3 Formalisme STIT

3.1 Langage et règles d'inférences

Présentons le langage de STIT [1, 3].

$$\phi ::= p|\neg\phi|\phi \wedge \phi|\phi \vee \phi|\phi \Rightarrow \phi|B_i\phi|[stit]_i\phi|X\phi|Y\phi$$

Ce langage utilise plusieurs modalités et règles d'inférence. La modalité $[stit]_i\phi$ signifie que l'agent i veille à ce que ϕ soit vraie. Les modalités X et Y sont des modalités temporelles représentant respectivement le futur et le passé. Enfin, B_i est la modalité de croyance associée à un agent i.

Le formalisme STIT considère les règles d'inférence classiques suivantes : **PC** (toutes les tautologies du calcul propositionnel notée $\mathbf{R_1}$ à $\mathbf{R_8}$ et rappelée en Figure 1), le modus ponens **MP**, la substitution **Sub** et la nécessitation **Nec** pour toutes les modalités. De plus, le formalisme STIT considère **KD45** pour les modalités B_i et **S5** pour les $[stit]_i$, **KD** pour X et **K** pour Y de type $\Box$. Les règles d'inférence spécifiques à ce formalisme sont données en Figure 2.

$$\phi \Rightarrow (\psi \Rightarrow \phi) \quad (R_1)$$

$$(\phi \Rightarrow (\psi \Rightarrow \theta)) \Rightarrow ((\phi \Rightarrow \psi) \Rightarrow (\phi \Rightarrow \theta)) \quad (R_2)$$

$$(\neg\psi \Rightarrow \neg\phi) \Rightarrow (\phi \Rightarrow \psi) \quad (R_3)$$

$$(\neg\phi \Rightarrow \neg\psi) \Rightarrow ((\neg\phi \Rightarrow \psi) \Rightarrow \phi) \quad (R_4)$$

$$\phi \Rightarrow (\phi \vee \psi), \psi \Rightarrow (\phi \vee \psi) \quad (R_5)$$

$$(\phi \vee \psi) \Rightarrow ((\phi \Rightarrow \theta) \Rightarrow ((\psi \Rightarrow \theta) \Rightarrow \theta) \quad (R_6)$$

$$\phi \Rightarrow (\psi \Rightarrow \phi \wedge \psi) \quad (R_7)$$

$$\phi \wedge \psi \Rightarrow \phi, \phi \wedge \psi \Rightarrow \psi \quad (R_8)$$

FIGURE 1 – Tautologies du calcul propositionnel

($\mathbf{Alt}_X$)	$\neg X\phi \Rightarrow X\neg\phi$
($\mathbf{Alt}_Y$)	$\neg Y\phi \Rightarrow Y\neg\phi$
($\mathbf{Conv}_{X,Y}$)	$\phi \Rightarrow XY\phi$
($\mathbf{Conv}_{Y,X}$)	$\phi \Rightarrow YX\phi$
($\mathbf{Rel}_{B_i,[stit]_i}$)	$B_i\phi \Rightarrow [stit]_j\phi$
(**AIA**)	$\neg B_i\neg[stit]_j\phi_1 \wedge \neg B_i\neg[stit]_k\phi_2$ $\Rightarrow \neg B_i\neg([stit]_j\phi_1 \wedge [stit]_k\phi_2)$
($\mathbf{Rel}_{[stit]_i,[stit]_N}$)	$\bigwedge_{n\in N}[stit]_n\phi_n \Rightarrow [stit]_N \bigwedge_{n\in N}\phi_n$
(**NCUH**)	$X\neg B_i\neg\phi \Rightarrow \langle stit\rangle_N X\phi$

FIGURE 2 – Règles d'inférence de STIT

4 Une logique de la confiance

Nous proposons une logique modale de la confiance où cette dernière est représentée par un ensemble de modalités $T_{i,j}\phi$ exprimant qu'un agent i accorde sa confiance à un agent j *pour l'énoncé* ϕ. Cette proposition ϕ peut exprimer le pouvoir de l'agent j à agir sur le monde, un état mental de cet agent, ou simplement un état du monde. Par exemple, une proposition ϕ peut être *l'agent j veille à* ψ que nous pouvons écrire à l'aide du formalisme STIT, $T_{i,j}[stit]_j\psi$. Cette formule signifie alors que l'agent i a confiance en j lorsque j déclare veiller à ce que ψ soit vraie. Nous pouvons aussi exprimer une confiance envers l'énoncé d'un état de croyance d'un agent avec la formule $T_{i,j}B_j\phi$ signifiant que l'agent i a confiance en j lorsque j déclare croire ϕ. L'intérêt de cette modalité doublement indicée est donc d'exprimer clairement la disposition d'un agent à faire confiance à un agent tiers donné.

4.1 Langage de la confiance

Précisons le langage de notre modèle. Soit un ensemble fini de variables propositionnelles $\mathcal{A} = \{a, b, c, ...\}$, deux agents $i, j \in \mathcal{N}$, un sous-ensemble d'agents $N \subseteq \mathcal{N}$, $p \in \mathcal{A}$ une variable propositionnelle et ϕ une formule du langage

STIT. Le langage que nous employons est le suivant :

$$\psi ::= p|\neg\psi|\psi \wedge \psi|\psi \vee \psi|\psi \Rightarrow \psi|T_{i,j}\psi|\phi$$

Remarquons que la modalité B_i est bien différente de $T_{i,i}$ car elle signifie que l'agent i a confiance en lui pour l'énoncé ϕ.

4.2 Une logique modale normale

Nous considérons les axiomes suivants : les tautologies du calcul propositionnel ($\mathbf{R_1}$ à $\mathbf{R_7}$), les règles d'inférence classiques des logiques modales (**K**, **Nec**, **Sub**) et une règle de non-inconsistance entre confiances (**D**). Notre logique de la confiance est donc une logique normale qui satisfait la nécessitation, la substitution, le modus ponens et l'axiome de Kripke K.

La **nécessitation** signifie que si une proposition ϕ est un théorème ($\vdash \phi$) alors n'importe quel agent i peut avoir confiance en n'importe quel autre agent j pour ce théorème ($\vdash T_{i,j}\phi$). La **substitution** signifie que si un théorème est prouvé et que nous substituons uniformément une formule quelconque à une lettre de proposition, la formule résultante est aussi un théorème. Le **modus ponens** signifie que si une proposition ϕ est prouvée et qu'il est aussi prouvé que $\phi \Rightarrow \psi$ alors la formule ψ est prouvée.

Enfin, notre modalité de confiance satisfait l'axiome **K** : si un agent i a confiance en un agent j sur ϕ qui implique ψ alors, si i a confiance en j pour ϕ alors i a aussi confiance en j pour ψ.

$$T_{i,j}(\phi \Rightarrow \psi) \Rightarrow T_{i,j}\phi \Rightarrow T_{i,j}\psi \quad (K)$$

4.3 Non-inconsistance de la confiance

Nous voulons exprimer le fait que si un agent i fait confiance à j pour un énoncé, il ne peut pas lui faire confiance pour son contraire pour des raisons de cohérence de discours : il n'est pas possible de faire confiance à un agent qui raisonne de manière inconsistante.

$$T_{i,j}\phi \Rightarrow \neg T_{i,j}\neg\phi \quad (D)$$

Ceci traduit le fait que si un agent i a confiance en un agent j pour ϕ alors cet agent i n'a pas confiance en j pour $\neg\phi$. Cependant, nous ne pouvons généraliser cet axiome pour tout autre agent $k \in \mathcal{N}$, $T_{i,j}\phi \Rightarrow \neg T_{i,k}\neg\phi$. En effet, considérons un agent i qui a confiance en j lorsque j déclare veiller à ce que ϕ soit vrai, c'est-à-dire $T_{i,j}[stit]_j\phi$, et qui a aussi confiance en un agent k lorsque ce dernier déclare veiller à ce que $\neg\phi$ soit vrai, c'est-à-dire $T_{i,k}[stit]_k\neg\phi$. Ici, la confiance ne préjuge pas de la vérité de la formule sur laquelle elle s'applique : deux agents peuvent croire être dans deux mondes différents et donc tenir deux discours inconsistants entre eux tout en étant, par exemple, sincères et honnêtes. Nous désirons alors capturer le fait qu'un agent tiers puisse tout de même accorder une forme de confiance à ces deux agents, tout en sachant que les formules ϕ et $\neg\phi$ ne pourront pas toutes les deux être vraies. Or, comme $[stit]$ est une modalité **S5** alors – selon l'axiome **T** – $[stit]_j\phi \Rightarrow \phi$ et $[stit]_k\neg\phi \Rightarrow \neg\phi$. En appliquant la **nécessitation** sur ces deux dernières formules qui sont des théorèmes, nous obtenons $T_{i,j}([stit]_j\phi \Rightarrow \phi)$ et $T_{i,k}([stit]_k\neg\phi \Rightarrow \neg\phi)$ puis par **K**, nous déduisons $T_{i,j}[stit]_j\phi \Rightarrow T_{i,j}\phi$ et $T_{i,k}[stit]_k\neg\phi \Rightarrow T_{i,k}\neg\phi$. Enfin, par **modus ponens**, nous déduisons alors $T_{i,j}\phi$ et $T_{i,k}\neg\phi$. Ainsi, si $T_{i,j}\phi \Rightarrow \neg T_{i,k}\neg\phi$ était un axiome, nous déduirions aussi $\neg T_{i,k}\neg\phi$ ce qui est contradictoire avec $T_{i,k}\neg\phi$.

4.4 Lien entre confiance et croyance

Dundua et Uridia [7] ont axiomatisé un lien entre la confiance et la croyance. Notre modalité de confiance étant une forme de croyance, nous posons que pour tout agent $i, j \in \mathcal{N}$, si un agent i croit qu'il a confiance en j sur l'énoncé ϕ alors il est vrai que l'agent i a confiance en l'agent j sur l'énoncé ϕ :

$$B_iT_{i,j}\phi \Rightarrow T_{i,j}\phi \quad (C4_{T,B})$$

De la même manière, nous posons la réciproque signifiant que si un agent i a confiance en un agent j sur l'énoncé ϕ alors l'agent i croit qu'il a confiance en l'agent j sur l'énoncé ϕ :

$$T_{i,j}\phi \Rightarrow B_iT_{i,j}\phi \quad (4_{T,B})$$

De plus, nous remarquons que nous ne pouvons pas considérer un axiome tel que si un agent croit que quelque chose est vrai alors il ne peut pas faire confiance dans un discours qui annonce le contraire : $\forall i, j \in \mathcal{N}, B_i\phi \Rightarrow \neg T_{i,j}\neg\phi \quad (D_{B,T})$. En effet, lorsque nous considérons la confiance en la sincérité, il est tout à fait possible de croire quelque chose en contradiction avec la confiance qu'on accorde à quelqu'un.

5 Propriétés

Comme indiqué en Section 2, plusieurs travaux se sont déjà intéressés à modéliser les aspects de la confiance, comme la notion de *confiance immédiate* ou encore la notion de *confiance dispositionnelle* [10, 19]. D'autres se sont intéressés aux propriétés de la confiance [18] ou son lien avec la notion de croyance [7]. Dans cette section, nous montrons que notre modalité de confiance permet d'exprimer ces aspects.

5.1 Distributivité de la confiance

Proposition 5.1. *Soit* $i, j \in \mathcal{N}$,
i) $T_{i,j}\phi \wedge T_{i,j}\psi \equiv T_{i,j}(\phi \wedge \psi) \quad (\wedge_T)$
ii) $(T_{i,j}\phi \vee T_{i,j}\psi) \Rightarrow T_{i,j}(\phi \vee \psi) \quad (\vee_T)$

Démonstration. i) Soit $i, j \in \mathcal{N}$,

($\Rightarrow$) Supposons $T_{i,j}\phi \wedge T_{i,j}\psi$, c'est-à-dire $T_{i,j}\phi$ et $T_{i,j}\psi$. Or, $\phi \Rightarrow (\psi \Rightarrow (\phi \wedge \psi))$ est une tautologie et donc un théorème. Par conséquent, en appliquant la nécessitation, nous déduisons $T_{i,j}(\phi \Rightarrow (\psi \Rightarrow (\phi \wedge \psi)))$. Par application de K, nous déduisons $T_{i,j}\phi \Rightarrow T_{i,j}(\psi \Rightarrow (\phi \wedge \psi))$. Par modus ponens, nous déduisons $T_{i,j}(\psi \Rightarrow (\phi \wedge \psi))$ puis, de nouveau par application de K, nous obtenons $T_{i,j}\psi \Rightarrow T_{i,j}(\phi \wedge \psi)$. Enfin, par modus ponens, nous prouvons $T_{i,j}(\phi \wedge \psi)$ et, par synthèse, nous prouvons $T_{i,j}\phi \wedge T_{i,j}\psi \Rightarrow T_{i,j}(\phi \wedge \psi)$

($\Leftarrow$) Supposons $T_{i,j}(\phi \wedge \psi)$. Or, $\phi \wedge \psi \Rightarrow \phi$ et $\phi \wedge \psi \Rightarrow \psi$ sont des tautologies et donc des théorèmes. En appliquant la nécessitation et l'axiome K, nous déduisons $T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\phi$ et $T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\psi$. Or, $(T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\phi) \wedge (T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\psi) \Rightarrow (T_{i,j}(\phi \wedge \psi) \Rightarrow T_{i,j}\phi \wedge T_{i,j}\psi)$. Enfin, par modus ponens, nous prouvons $T_{i,j}\phi \wedge T_{i,j}\psi$ et, par synthèse, nous prouvons $T_{i,j}\phi \wedge T_{i,j}\psi \Leftarrow T_{i,j}(\phi \wedge \psi)$.

ii) Soit $i, j \in \mathcal{N}$. Supposons $T_{i,j}\phi \vee T_{i,j}\psi$. Or, $\phi \Rightarrow (\phi \vee \psi)$ et $\psi \Rightarrow (\phi \vee \psi)$ sont des tautologies et donc des théorèmes.En appliquant la nécessitation et l'axiome K, nous obtenons $T_{i,j}\phi \Rightarrow T_{i,j}(\phi \vee \psi))$ et $T_{i,j}\psi \Rightarrow T_{i,j}(\phi \vee \psi))$. Or, $(T_{i,j}\phi \vee T_{i,j}\psi) \Rightarrow ((T_{i,j}\phi \Rightarrow T_{i,j}(\phi \vee \psi)) \Rightarrow (T_{i,j}\psi \Rightarrow T_{i,j}(\phi \vee \psi)) \Rightarrow T_{i,j}(\phi \vee \psi))$ est aussi une tautologie du calcul propositionnel. Par modus ponens, nous prouvons $T_{i,j}(\phi \vee \psi)$ et, par synthèse, nous prouvons $(T_{i,j}\phi \vee T_{i,j}\psi) \Rightarrow T_{i,j}(\phi \vee \psi)$. □

Ces propositions se fondent sur le fait que nous utilisons une logique normale : un agent i ne peut pas faire confiance en un discours inconsistant (l'ensemble des propositions énoncées par l'agent j) car cela le conduirait à avoir confiance en toute proposition émanant de j. Par exemple, un agent ne peut pas faire confiance à un autre pour l'enrichir et l'appauvrir (en admettant que les deux propositions soient contraires) mais il n'exprimera pas pour autant une non confiance à son égard. Un agent peut bien entendu faire confiance à un autre pour l'enrichir ou l'appauvrir. Toutefois, il s'agit d'une tautologie et un agent a toujours confiance dans une tautologie, par nécessitation.

5.2 Confiance dispositionnelle

Rappelons que la confiance dispositionnelle présentée dans [10] exprime la confiance en un autre agent pour réaliser une action lorsqu'un contexte donné est vérifié. Exprimons cette notion en utilisant le formalisme STIT : la confiance dispositionnelle est la confiance (dans le présent) que, lorsque le contexte ϕ est vérifié, l'agent veille bien à ce que ψ soit réalisé.

$$\forall i, j \in \mathcal{N}, Ts_{i,j}(\phi, \psi) \triangleq T_{i,j}(\phi \Rightarrow [stit]_j\psi)$$

Contrairement à [10], nous n'introduisons pas la notion d'intention car celle-ci est déjà exprimée dans la modalité STIT. En effet, si un agent veille à ce que quelque chose devienne vrai, c'est qu'il a l'intention de rendre cette chose vraie. Mais intéressons-nous désormais aux propriétés algébriques liées aux opérateurs $\wedge$ et $\vee$ de notre confiance dispositionnelle. Que pouvons-nous déduire si un agent i a confiance en un agent j pour que, lorsque le contexte ϕ est réalisé, l'agent j veille à ce que $\psi \wedge \theta$ soit vrai, ou que dans un contexte $\phi \vee \psi$, l'agent j veille à ce que θ soit vrai.

Proposition 5.2. *Soit* $i, j \in \mathcal{N}$,
i) $Ts_{i,j}(\phi, \psi \wedge \theta) \equiv Ts_{i,j}(\phi, \psi) \wedge Ts_{i,j}(\phi, \theta) \quad (Ts_1)$
ii) $(Ts_{i,j}(\phi, \theta) \vee Ts_{i,j}(\psi, \theta)) \Rightarrow Ts_{i,j}(\phi \wedge \psi, \theta) \quad (Ts_2)$
iii) $Ts_{i,j}(\phi \vee \psi, \theta) \equiv Ts_{i,j}(\phi, \theta) \wedge Ts_{i,j}(\psi, \theta) \quad (Ts_3)$
iv) $(Ts_{i,j}(\phi, \psi) \vee Ts_{i,j}(\phi, \theta)) \Rightarrow Ts_{i,j}(\phi, \psi \vee \theta) \quad (Ts_4)$

Démonstration. Soit $i, j \in \mathcal{N}$,

i) Supposons $Ts_{i,j}(\phi, \psi \wedge \theta)$. Par définition, $T_{i,j}(\phi \Rightarrow [stit]_j(\psi \wedge \theta)) \equiv T_{i,j}(\phi \Rightarrow ([stit]_j\psi \wedge [stit]_j\theta)) \equiv T_{i,j}((\phi \Rightarrow [stit]_j\psi) \wedge (\phi \Rightarrow [stit]_j\theta)) \equiv T_{i,j}(\phi \Rightarrow [stit]_j\psi) \wedge T_{i,j}(\phi \Rightarrow [stit]_j\theta) \equiv Ts_{i,j}(\phi, \psi) \wedge Ts_{i,j}(\phi, \theta)$.

ii) Supposons $Ts_{i,j}(\phi, \theta) \vee Ts_{i,j}(\psi, \theta)$. Pr définition, $T_{i,j}(\phi \Rightarrow [stit]_j\theta) \vee T_{i,j}(\psi \Rightarrow [stit]_j\theta)$. Donc, par modus ponens sur $\vee_T$, nous déduisons $T_{i,j}((\phi \Rightarrow [stit]_j\theta) \vee (\psi \Rightarrow [stit]_j\theta)) \equiv T_{i,j}((\phi \wedge \psi \Rightarrow [stit]_j\theta)) \equiv Ts_{i,j}(\phi \wedge \psi, \theta)$. Par synthèse, nous prouvons alors $(Ts_{i,j}(\phi, \theta) \vee Ts_{i,j}(\psi, \theta)) \Rightarrow Ts_{i,j}(\phi \wedge \psi, \theta)$.

iii) Supposons $Ts_{i,j}(\phi \vee \psi, \theta)$. Par définition, $T_{i,j}((\phi \vee \psi) \Rightarrow [stit]_j\theta) \equiv T_{i,j}((\phi \Rightarrow [stit]_j\theta) \wedge (\psi \Rightarrow [stit]_j\theta)) \equiv T_{i,j}(\phi \Rightarrow [stit]_j\theta) \wedge T_{i,j}(\psi \Rightarrow [stit]_j\theta)$.

iv) Supposons $Ts_{i,j}(\phi, \psi) \vee Ts_{i,j}(\phi, \theta)$. Par définition, $T_{i,j}(\phi \Rightarrow [stit]_j\psi) \vee T_{i,j}(\phi \Rightarrow [stit]_j\theta)$. Donc, par modus ponens sur $\vee_T$, nous déduisons $T_{i,j}((\phi \Rightarrow [stit]_j\psi) \vee (\phi \Rightarrow [stit]_j\theta)) \equiv T_{i,j}(\phi \Rightarrow ([stit]_j\psi \vee [stit]_j\theta))$. Par nécessitation, nous obtenons $T_{i,j}[(\phi \Rightarrow (([stit]_j\psi \vee [stit]_j\theta) \Rightarrow [stit]_j(\psi \vee \theta)) \Rightarrow ((\phi \Rightarrow ([stit]_j\psi \vee [stit]_j\theta)) \Rightarrow (\phi \Rightarrow [stit]_j(\psi \vee \theta))]$. Par K et modus ponens, nous déduisons $T_{i,j}(\phi \Rightarrow [stit]_j(\psi \vee \theta))$. Enfin, par synthèse, nous obtenons $(Ts_{i,j}(\phi, \psi) \vee Ts_{i,j}(\phi, \theta)) \Rightarrow Ts_{i,j}(\phi, \psi \vee \theta)$ □

Ceci nous permet de prouver quelques déductions logiques naturelles comme, par exemple, lorsqu'un agent i a confiance dans un agent j pour ϕ et que cet agent i a aussi confiance dans l'agent j pour que, dans le contexte ϕ, l'agent j veille à ψ alors l'agent i a confiance en l'agent j pour ψ. Nous pouvons déduire aussi que si l'agent i a confiance en j qu'il veille à $\psi \Rightarrow \theta$ et si i a une confiance dispositionnelle en j pour que, dans le contexte ϕ, j veille à ψ et i a aussi une confiance dispositionnelle en j pour que, dans le contexte θ, j veille à ϑ alors i a aussi une confiance dispositionnelle en j pour que, dans le contexte ϕ, j veille à ϑ. Enfin, si i a confiance en j pour ϕ et qu'il a aussi une confiance dispositionnelle en j pour que, dans le contexte $\phi \wedge \psi$, j veille à θ alors i a aussi une confiance dispositionnelle envers j pour que, dans le contexte ψ, j veille à θ.

Proposition 5.3. *Soit $i, j \in \mathcal{N}$,*
i) $(T_{i,j}\phi \wedge Ts_{i,j}(\phi,\psi)) \Rightarrow T_{i,j}\psi$
ii) $(T_{i,j}[stit]_j(\psi \Rightarrow \theta) \wedge Ts_{i,j}(\phi,\psi) \wedge Ts_{i,j}(\theta,\vartheta)) \Rightarrow Ts_{i,j}(\phi,\vartheta)$
iii) $T_{i,j}\phi \wedge Ts_{i,j}(\phi \wedge \psi,\theta) \Rightarrow Ts_{i,j}(\psi,\theta)$

Démonstration. Soit $i, j \in \mathcal{N}$,

i) Supposons $(T_{i,j}\phi \wedge Ts_{i,j}(\phi,\psi))$. Par définition, $T_{i,j}\phi \wedge T_{i,j}(\phi \Rightarrow [stit]_j\psi)$. Par application de K, nous déduisons $T_{i,j}\phi \wedge (T_{i,j}\phi \Rightarrow T_{i,j}[stit]_j\psi)$. Par modus ponens, $T_{i,j}[stit]_j\psi$. Or, par nécessitation sur $T_{[stit]}$ et K, $T_{i,j}[stit]_j\psi \Rightarrow T_{i,j}\psi$. Ainsi, par modus ponens, nous prouvons $T_{i,j}\psi$. Par synthèse, nous prouvons alors $(T_{i,j}\phi \wedge Ts_{i,j}(\phi,\psi)) \Rightarrow T_{i,j}\psi$.

ii) Supposons $T_{i,j}[stit]_j(\psi \Rightarrow \theta) \wedge Ts_{i,j}(\phi,\psi) \wedge Ts_{i,j}(\theta,\vartheta)$, c'est-à-dire $T_{i,j}([stit]_j(\psi \Rightarrow \theta) \wedge (\phi \Rightarrow [stit]_j\psi) \wedge (\theta \Rightarrow [stit]_j\vartheta))$. En appliquant l'axiome K sur $[stit]$, nous avons $T_{i,j}(([stit]_j\psi \Rightarrow [stit]_j\theta) \wedge (\phi \Rightarrow [stit]_j\psi) \wedge (\theta \Rightarrow [stit]_j\vartheta))$. Or, $([stit]_j\psi \Rightarrow [stit]_j\theta) \Rightarrow (\phi \Rightarrow ([stit]_j\psi \Rightarrow [stit]_j\theta))$ est une tautologie et donc, par modus ponens, nous déduisons $\phi \Rightarrow ([stit]_j\psi \Rightarrow [stit]_j\theta)$. De plus, $(\phi \Rightarrow ([stit]_j\psi \Rightarrow [stit]_j\theta)) \Rightarrow ((\phi \Rightarrow [stit]_j\psi) \Rightarrow (\phi \Rightarrow [stit]_j\theta))$ est aussi une tautologie. Ainsi, par application de la nécessitation, K, puis modus ponens, nous prouvons $T_{i,j}(\phi \Rightarrow [stit]_j\theta)$.

À ce point, nous avons donc $T_{i,j}(\phi \Rightarrow [stit]_j\theta) \wedge T_{i,j}(\theta \Rightarrow [stit]_j\vartheta)$. Or, l'axiome T sur $[stit]$ indique que $[stit]_j\theta \Rightarrow \theta$ est un théorème. Par un raisonnement analogue, nous avons $T_{i,j}(\phi \Rightarrow \theta) \wedge T_{i,j}(\theta \Rightarrow [stit]_j\vartheta)$ et donc $T_{i,j}(\phi \Rightarrow [stit]_j\vartheta)$, c'est-à-dire, $Ts_{i,j}(\phi,\vartheta)$. Par synthèse, nous concluons $(T_{i,j}[stit]_j(\psi \Rightarrow \theta) \wedge Ts_{i,j}(\phi,\psi) \wedge Ts_{i,j}(\theta,\vartheta)) \Rightarrow Ts_{i,j}(\phi,\vartheta)$.

iii) Supposons $T_{i,j}\phi \wedge Ts_{i,j}(\phi \wedge \psi,\theta)$. Par définition, $T_{i,j}\phi \wedge T_{i,j}(\phi \wedge \psi \Rightarrow [stit]_j\theta)$, c'est-à-dire $T_{i,j}(\phi \wedge (\phi \wedge \psi \Rightarrow [stit]_j\theta))$. En substituant dans la tautologie $\alpha \wedge (\alpha \wedge \beta \Rightarrow \gamma) \Rightarrow (\beta \Rightarrow \gamma)$ [3], nous déduisons $T_{i,j}(\psi \Rightarrow [stit]_j\theta))$. Par synthèse, nous prouvons alors $T_{i,j}\phi \wedge Ts_{i,j}(\phi \wedge \psi,\theta) \Rightarrow Ts_{i,j}(\psi,\theta)$. □

5.3 Similarités avec la théorie de Singh

Nous pouvons faire un lien direct entre notre axiomatique et celle proposée par Singh [18]. Ces derniers utilisent vingt-et-un axiomes, notés de T_1 à T_{21}. Nous ne considérons pas T_1, T_4, T_5 et T_{19} qui s'appuient sur le fait que la confiance est une modalité qui s'active ou non selon le contexte. Nous ne considérons pas non plus les axiomes T_{10} à T_{13} et T_{20} et T_{21} qui font appel à des modalités ou des prédicats supplémentaires comme l'engagement, le scepticisme ou l'espérance. Nos propriétés algébriques et déductions logiques nous permettent de retrouver les autres axiomes : T_2, T_3, T_6, T_7, T_8, T_9. Nous traitons les axiomes T_{14} à T_{18} relatifs à la confiance collective dans la section suivante.

3. Rappellons que $\alpha \wedge (\alpha \wedge \beta \Rightarrow \gamma) \equiv \alpha \wedge (\neg\alpha \vee \neg\beta \vee \gamma) \equiv \alpha \wedge \neg\alpha \vee \alpha \wedge (\neg\beta \vee \gamma) \equiv \alpha \wedge (\beta \Rightarrow \gamma)$. Et $\alpha \wedge (\beta \Rightarrow \gamma) \Rightarrow (\beta \Rightarrow \gamma)$.

Notre Proposition 5.3.iii est similaire à l'axiome T_2 $(T_{i,j}(\phi \wedge \phi',\psi) \wedge \phi \Rightarrow T_{i,j}(\phi',\psi))$ qui indique que si un agent i a confiance envers j pour ψ dans le contexte $\phi \wedge \phi'$ et que ϕ est vrai alors l'agent i a confiance en j pour ψ dans le contexte ϕ' .

Notre Proposition 5.2.i permet de déduire l'axiome T_3 $(T_{i,j}(\phi,\psi \wedge \psi') \wedge \neg\psi \Rightarrow T_{i,j}(\phi,\psi))$ qui indique que si un agent i a confiance envers j pour $\psi \wedge \psi'$ dans le contexte ϕ et que $\neg\psi$ est vrai alors l'agent i a confiance envers j pour ψ dans le contexte ϕ. En effet, en appliquant l'augmentation d'hypothèse R_1 sur notre Proposition 5.2.i, nous déduisons la même propriété. En effet, si $Ts_{i,j}(\phi,\psi \wedge \psi') \Rightarrow Ts_{i,j}(\phi,\psi)$ alors, par R_1, nous avons $Ts_{i,j}(\phi,\psi) \Rightarrow (Ts_{i,j}(\phi,\psi \wedge \psi') \wedge \neg\psi \Rightarrow Ts_{i,j}(\phi,\psi))$. Donc, si $Ts_{i,j}(\phi,\psi \wedge \psi')$ est prouvé alors, par modus ponens, nous prouvons $Ts_{i,j}(\phi,\psi \wedge \psi') \wedge \neg\psi \Rightarrow Ts_{i,j}(\phi,\psi)$.

L'axiome T_6 (De $T_{i,j}(\phi,\psi)$, $\phi' \vdash \phi$ et $\phi \nvdash \psi$, l'agent infère $T_{i,j}(\phi',\psi)$) est naturel par syllogisgme dans notre formalisme. En effet, il exprime que si un agent i a confiance envers j pour ψ dans un contexte ϕ et que $\phi' \Rightarrow \phi$ est un théorème alors i a confiance envers j pour ψ dans le contexte ϕ'.

Nos Propositions 5.2.iii et 5.2.i sont similaires aux axiomes T_7 $(T_{i,j}(\phi,\psi) \wedge T_{i,j}(\phi',\psi) \Rightarrow T_{i,j}(\phi \vee \phi',\psi))$ et T_8 $(T_{i,j}(\phi,\psi) \wedge T_{i,j}(\phi,\psi') \Rightarrow T_{i,j}(\phi,\psi \wedge \psi'))$ indiquant respectivement (T_7) que si un agent i a confiance envers j pour ψ dans un contexte ϕ et si i a confiance envers j pour ψ dans un contexte ϕ' alors l'agent i a confiance envers j pour ψ dans le contexte $\phi \vee \phi'$, et (T_8) que si un agent i a confiance envers j pour ψ dans le contexte ϕ et qu'il a confiance envers j pour ψ' dans le contexte ϕ alors l'agent i a confiance envers j pour $\psi \wedge \psi'$ dans le contexte ϕ.

Enfin, notre Proposition 5.3.ii est similaire à l'axiome T_9 (De $T_{i,j}(\phi,\psi)$, $\psi \vdash \phi'$ et $T_{i,j}(\phi',\psi')$, l'agent infère $T(\phi,\psi')$) indiquant que si un agent i a confiance en j pour ψ dans le contexte ϕ, qu'il a aussi confiance en j pour ψ' dans le contexte ϕ' et que $\psi \Rightarrow \phi'$ alors l'agent i a confiance en j pour ψ' dans le contexte ϕ.

6 Vers une confiance collective

Jusqu'à présent, nous nous sommes intéressés à la confiance qu'il pouvait y avoir entre deux agents. Dans cette section, nous étendons la notion de confiance à des groupes d'agents afin d'exprimer de la *confiance collective* avec deux notions : la *confiance commune* et la *confiance par transitivité*. Remarquons que, si nous nous restreignons dans ce travail à ces deux notions, nous sommes bien conscients de l'intérêt d'autres aspects de la confiance collective, comme la *confiance réciproque* ou encore la *confiance mutuelle*.

6.1 Confiance commune

Afin de définir la *confiance commune*, nous nous appuyons sur la définition de Smith *et al.* [19] : un groupe d'agents I a une confiance commune envers un autre groupe d'agents J si, et seulement si, tous les agents de I ont confiance en tous les agents de J.

$$\forall I, J \subseteq \mathcal{N} : Tc_{I,J}\phi \triangleq \bigwedge_{(i,j)\in I\times J} T_{i,j}\phi$$

Il s'agit d'un consensus en ce sens que tous les agents du groupe I doivent avoir confiance dans tous les agents du groupe J à propos du même énoncé. De plus, nous considérons une notion duale au prédicat $Tc_{I,J}$, notée $Tc^*_{I,J}$, comme étant le prédicat suivant.

$$\forall I, J \subseteq \mathcal{N} : Tc^*_{I,J}\phi \triangleq \bigvee_{(i,j)\in I\times J} T_{i,j}\phi$$

Ce prédicat exprime qu'il existe au moins un agent de I faisant confiance à un agent de J. En effet, si aucun agent de I ne fait confiance aux agents de J pour ϕ alors nous pouvons écrire $\neg Tc^*_{I,J}\phi$.

Remarquons qu'il existe d'autres définitions dans la littérature. Par exemple, Herzig *et al.* [10] considèrent un prédicat de *réputation* indiquant qu'il y a consensus si une *majorité* d'agents de I a une confiance dispositionnelle dans les agents de J. Pour raison de simplicité, nous n'introduisons pas de notion de majorité et ne considérons donc pas cette notion de réputation.

6.2 Confiance par transitivité

Notre notion de confiance considère de manière abstraite chaque aspect de la confiance, seule l'application de notre logique va spécifier l'aspect considéré. Lorsque nous considérons, par exemple, l'aspect de la confiance comme une croyance en la sincérité et la fiabilité d'un agent à produire un énoncé, nous pouvons caractériser certaines propriétés spécifiques à cet aspect comme la *transitivité*. En effet, il semble naturel de dire que si un agent i a confiance en la confiance qu'accorde un agent j à propos d'un troisième agent k alors i a confiance en k si et seulement si i pense que j est sincère et fiable. Nous présentons alors dans cette section une confiance collective qui pourrait être construite sur cette notion. Celle-ci repose sur l'idée que si deux agents i et j ont mutuellement confiance dans la confiance que l'autre accorde à un agent k pour ϕ alors les deux agents font confiance à l'agent k pour ϕ.

$$Ta_{\{i,j\},\{k\}}\phi \triangleq (T_{j,i}T_{i,k}\phi \Rightarrow T_{j,k}\phi) \wedge (T_{i,j}T_{j,k}\phi \Rightarrow T_{i,k}\phi)$$

Nous pouvons par ailleurs nous demander pourquoi nous ne considérons pas la transitivité de la confiance $T_{i,j}T_{j,k}\phi \Rightarrow T_{i,k}\phi$ comme un principe de raisonnement naturel à la confiance ? En effet, cette notion de transitivité s'applique très bien à certains aspects de la confiance comme la croyance en la fiabilité d'un agent. En revanche ceci ne peut pas s'appliquer pour d'autres aspect de la confiance comme la confiance en l'honnêteté d'un discours. Considérons par exemple qu'un agent i a confiance en l'honnêteté d'un agent j à propos de la confiance que j accorde à un agent k. Supposons que i sait que k n'est pas fiable et qu'il sait que j ne le sait pas. Dans ces conditions, i peut faire confiance à j quand à l'honnêteté de sa déclaration mais il ne peut pas faire confiance à k par transitivité. C'est pour cette raison que nous ne pouvons pas considérer comme principe généralla transitivité de la confiance. Toutefois, il est possible d'utiliser ce principe de manière locale lorsque les agents s'associent avec d'autres agents et pensent que ces agents sont justement fiables. C'est l'idée fondatrice de notre *confiance par transitivité*. Ainsi, nous généralisons cette confiance par transitivité à un groupe d'agents $I \subseteq \mathcal{N}$ envers un autre groupe $K \subseteq \mathcal{N}$.

$$Ta_{I,K}\phi \triangleq \bigwedge_{(i,j)\in I\times I} \wedge_{k\in K}(T_{i,j}T_{j,k}\phi \Rightarrow T_{i,k}\phi)$$

De la même manière, nous exprimons la *non confiance par transitivité*. Si un agent i a confiance en j dans le fait que ce dernier n'accorde pas sa confiance à k pour ϕ alors l'agent i ne fait pas confiance à k pour ϕ. Pour tout $I, K \subseteq \mathcal{N}$:

$$Tb_{I,K}\phi \triangleq \bigwedge_{(i,j)\in I\times I} \wedge_{k\in K}(T_{i,j}\neg T_{j,k}\phi \Rightarrow \neg T_{i,k}\phi)$$

6.3 Propriétés usuelles

La confiance commune a les propriétés usuelles.

Proposition 6.1. *Pour tout $I, J, K \subseteq \mathcal{N}$,*

$$i) \quad Tc_{I,J}\phi \wedge Tc_{I,J}\psi \equiv Tc_{I,J}(\phi \wedge \psi)$$

$$ii) \quad (Tc_{I,J}\phi \vee Tc_{I,J}\psi) \Rightarrow Tc_{I,J}(\phi \vee \psi)$$

$$iii) \quad Tc_{I,J}\phi \Rightarrow \neg Tc^*_{I,K}\neg\phi$$

Démonstration. Pour tout $I, J, K \subseteq \mathcal{N}$,

i) Supposons $Tc_{I,J}\phi \wedge Tc_{I,J}\psi$. Or pour tout $i, j \in I \times J$, $T_{i,j}\phi \wedge T_{i,j}\psi \equiv T_{i,j}(\phi \wedge \psi)$. Donc :

$$\begin{aligned} Tc_{I,J}\phi \wedge Tc_{I,J}\psi &\equiv \bigwedge_{(i,j)\in I\times J} T_{i,j}\phi \wedge \bigwedge_{(i,j)\in I\times J} T_{i,j}\psi \\ &\equiv \bigwedge_{(i,j)\in I\times J} (T_{i,j}\phi \wedge T_{i,j}\psi) \\ &\equiv \bigwedge_{(i,j)\in I\times J} T_{i,j}(\phi \wedge \psi) \\ &\equiv Tc_{I,J}(\phi \wedge \psi) \end{aligned}$$

ii) Supposons $Tc_{I,J}\phi \vee Tc_{I,J}\psi$. Donc :

$$\bigwedge_{(i,j)\in I\times J} T_{i,j}\phi \vee \bigwedge_{(i,j)\in I\times J} T_{i,j}\psi$$

Or pour tout $(k,l) \in I \times J$, nous avons :

$$(T_{k,l}\phi \vee T_{k,l}\psi) \Rightarrow T_{k,l}(\phi \vee \psi)$$

Par soucis de lisibilité, notons Δ la formule $(i,j) \in (I \times J) \setminus \{(k,l)\}$, Δ' la formule $(i,j) \in (I\times J)$ et Δ'' la formule $(k,l) \in (I \times J)$. Ainsi :

$$\bigwedge_{\Delta} T_{i,j}\phi \wedge T_{k,l}\phi \vee \bigwedge_{\Delta} T_{i,j}\psi \wedge T_{k,l}\psi$$

En appliquant (R_8) :

$$\begin{aligned}(\bigwedge_{\Delta} T_{i,j}\phi \vee (\bigwedge_{\Delta} T_{i,j}\psi \wedge T_{k,l}\psi)) \wedge \quad (T_{k,l}\phi \vee \bigwedge_{\Delta} T_{i,j}\psi)\\ \wedge \quad (T_{k,l}\phi \vee T_{k,l}\psi) \Rightarrow T_{k,l}(\phi \vee \psi)\end{aligned}$$

Donc :

$$\begin{aligned}\bigwedge_{\Delta''}[(\bigwedge_{\Delta} T_{i,j}\phi \wedge T_{k,l}\phi \vee \bigwedge_{\Delta} T_{i,j}\psi \wedge T_{k,l}\psi) \Rightarrow T_{k,l}(\phi \vee \psi)]\\ \equiv [(\bigwedge_{\Delta'} T_{i,j}\phi \vee \bigwedge_{\Delta'} T_{i,j}\psi)] \Rightarrow \bigwedge_{\Delta''} T_{k,l}(\phi \vee \psi)\end{aligned}$$

Par conséquent, par application de la synthèse :

$$(Tc_{I,J}\phi \vee Tc_{I,J}\psi) \Rightarrow Tc_{I,J}(\phi \vee \psi)$$

iii) Supposons $Tc_{I,J}\phi$. Donc :

$$\bigwedge_{(i,j)\in I\times J} T_{i,j}\phi$$

Or pour tout $(i,j) \in I \times J$, nous avons $T_{i,j}\phi \Rightarrow \neg T_{i,j}\neg\phi$. En appliquant deux fois (R_7) :

$$\bigwedge_{(i,j)\in I\times J} (T_{i,j}\phi \Rightarrow \neg T_{i,j}\neg\phi)$$

$$\bigwedge_{(i,j)\in I\times J} (T_{i,j}\phi \wedge (T_{i,j}\phi \Rightarrow \neg T_{i,j}\neg\phi))$$

Par modus ponens, nous déduisons :

$$\bigwedge_{(i,j)\in I\times J} \neg T_{i,j}\neg\phi$$

Or, par le théorème de De Morgan :

$$\bigwedge_{(i,j)\in I\times J} \neg T_{i,j}\neg\phi \equiv \neg \bigvee_{(i,j)\in I\times J} T_{i,j}\neg\phi \equiv \neg Tc^*_{I,J}\neg\phi$$

Par synthèse, nous prouvons :

$$Tc_{I,J}\phi \Rightarrow \neg Tc^*_{I,K}\neg\phi$$

□

6.4 Liens entre confiances collectives

La transitivité de la confiance au sein d'un groupe $I \cup J$ se retrouve à l'échelle du groupe lorsqu'il y a une confiance commune envers un groupe K.

Proposition 6.2. *Pour tout $I, J, K \subseteq \mathcal{N}$,*

$$i) \quad Ta_{I\cup J,K}\phi \wedge Tc_{I,J}Tc_{J,K}\phi \Rightarrow Tc_{I,K}\phi$$

$$ii) \quad Tb_{I\cup J,K}\phi \wedge Tc_{I,J}\neg T^*_{J,K}\phi \Rightarrow \neg Tc^*_{I,K}\phi$$

Démonstration. *i*) Supposons $Ta_{I\cup J,K}\phi \wedge Tc_{I,J}Tc_{J,K}\phi$. Par (R_8) :

$$\begin{aligned}&\bigwedge_{(i,j)\in I\times J} T_{i,j}(\bigwedge_{(j',k)\in J\times K} T_{j',k}\phi)\\ \equiv &\bigwedge_{(i,j)\in I\times J} T_{i,j}(\bigwedge_{(j',k)\in J\setminus\{j\}\times K} T_{j',k}\phi \bigwedge_{k\in K} T_{j,k}\phi)\end{aligned}$$

Nous en déduisons [4] :

$$\begin{aligned}&\bigwedge_{(i,j)\in I\times J} T_{i,j}(\bigwedge_{(j',k)\in J\setminus\{j\}\times K} T_{j',k}\phi \bigwedge_{k\in K} T_{j,k}\phi)\\ &\Rightarrow (\bigwedge_{(i,j)\in I\times J} T_{i,j}(\bigwedge_{k\in K} T_{j,k}\phi))\end{aligned}$$

Par modus ponens et par la proposition 5.1.i :

$$(\bigwedge_{(i,j)\in I\times J} T_{i,j}(\bigwedge_{k\in K} T_{j,k}\phi)) \equiv (\bigwedge_{(i,j)\in I\times J} \bigwedge_{k\in K} T_{i,j}(T_{j,k}\phi))$$

Or, nous avons aussi par (R_8) les propriétés de transitivité pour tous les agents $(i,j) \in I \times J$ selon l'hypothèse $Ta_{I\cup J,K}\phi$. Par conséquent, pour tout $(i,j) \in I \times J$ et $k \in K$: $T_{i,j}T_{j,k}\phi \Rightarrow T_{i,k}\phi$. Par (R_7), nous avons :

$$(\bigwedge_{(i,j)\in I\times J} \bigwedge_{k\in K}(T_{i,j}T_{j,k}\phi \Rightarrow T_{i,k}\phi))$$

Puis par modus ponens, nous déduisons :

$$(\bigwedge_{(i,j)\in I\times J} \bigwedge_{k\in K} T_{i,k}\phi)) \equiv Tc_{I,K}\phi$$

Par synthèse, nous prouvons alors :

$$Ta_{I\cup J,K}\phi \wedge Tc_{I,J}T_{J,K}\phi \Rightarrow Tc_{I,K}\phi$$

ii) Supposons $Tb_{I\cup J,K}\phi \wedge Tc_{I,J}\neg T^*_{J,K}\phi$. Par (R_8) :

$$Tc_{I,J}\neg T^*_{J,K}\phi \equiv \bigwedge_{(i,j)\in I\times J} \neg \bigvee_{(j',k)\in J\times K} T^*_{j',k}\phi$$

4. Dans le détail, nous appliquons la nécessitation pour tous les $T_{i,j}$ puis K, R_7 puis R_8. Pour des soucis de lisibilité, nous nous permettons ce raccourci trivial.

Donc,

$$\bigwedge_{(i,j)\in I\times J}\bigwedge_{(j',k)\in J\times K}\neg T_{j',k}\phi$$
$$\equiv \bigwedge_{(i,j)\in I\times J}(\bigwedge_{(j',k)\in J\setminus\{j\}\times K}\neg T_{j',k}\phi \wedge \neg T_{j,k}\phi)$$
$$\equiv \bigwedge_{(i,j)\in I\times J}(\bigwedge_{(j',k)\in J\setminus\{j\}\times K}\neg T_{j',k}\phi) \wedge \bigwedge_{k\in K}\neg T_{j,k}\phi)$$

Par (R_8) et par la propriété 5.1.i, nous déduisons :

$$\bigwedge_{(i,j)\in I\times J}\bigwedge_{k\in K}\neg T_{j,k}\phi$$

Or, notre hypothèse, nous permet aussi de déduire $Tb_{I\cup J,K}\phi$ par (R_8). Ainsi, pour tout $(i,j) \in I\times I \wedge k \in K, T_{i,j}$:

$$\neg T_{j,k}\phi \Rightarrow \neg T_{i,k}\phi$$

Donc par (R_7) :

$$\bigwedge_{(i,j)\in I\times J}\bigwedge_{k\in K}(T_{i,j}\neg T_{j,k}\phi \wedge (T_{i,j}\neg T_{j,k}\phi \Rightarrow \neg T_{i,k}\phi))$$

Puis par modus ponens, nous déduisons :

$$\bigwedge_{(i,j)\in I\times J}\bigwedge_{k\in K}\neg T_{i,k}\phi$$

Par simplification et par De Morgan, nous avons :

$$\bigwedge_{(i,k)\in I\times K}\neg T_{i,k}\phi \equiv \neg\bigvee_{(i,k)\in I\times K}T_{i,k}\phi$$

Enfin, par synthèse, nous prouvons :

$$Tb_{I\cup J,K}\phi \wedge Tc_{I,J}\neg T^*_{J,K}\phi \Rightarrow \neg Tc^*_{I,K}\phi$$

□

6.5 Analogies avec Singh

Comme pour la confiance individuelle, nous pouvons faire des liens entre notre modèle et des propriétés axiomatisées par Singh. Appliquons par exemple la notion de confiance commune à la la confiance dispositionnelle. Considérons le prédicat :

$$\forall I, J \subseteq \mathcal{N} : Td_{I,J}(\phi,\psi) \triangleq \bigwedge_{(i,j)\in I\times J} Ts_{i,j}(\phi,\psi)$$

Nous retrouvons alors l'axiome T_{15} de Singh ($T_{i,j}(\phi,\psi) \Rightarrow T_{i,\{j,k\}}(\phi,\psi)$) qui stipule que si un agent i fait confiance à j pour ψ dans un contexte ϕ, alors i fait confiance au groupe $\{j,k\}$ pour la même chose. Pour déduire une telle proposition dans notre modèle, nous avons simplement besoin de prouver qu'il y a une confiance commune sur $Ts_{i,j}(\phi,\psi)$ et $Ts_{i,k}(\phi,\psi)$. De la même manière, nous caractérisons avec le prédicat $Td_{I,J}(\phi,\psi)$ l'axiome T_{16} ($T_{i,j}(\phi,\psi) \Rightarrow T_{\{i,k\},j}(\phi,\psi)$) stipulant que si un agent i fait confiance à j pour ψ dans un contexte ϕ alors le groupe d'agents $\{i,k\}$ fait confiance au groupe j pour la même chose.

En revanche, nous ne pouvons pas déduire l'axiome T_{14} ($T_{i,j}(\phi,\psi) \wedge T_{j,i}(\psi,\phi) \Rightarrow T_{i,\{i,j\}}(\top,\phi\wedge\psi)$) qui dit que si un agent i fait confiance à j pour ψ dans un contexte ϕ et si un agent j fait confiance à i pour ϕ dans le contexte ψ alors i fait confiance au groupe $\{i,j\}$ pour $\phi\wedge\psi$, et ce quelque soit le contexte. En effet, si nous essayons de démontrer une caractérisation de cet axiome (par exemple avec une contraposition), nous obtenons $\neg Td_{\{i\},\{i,j\}}(\top,\phi\wedge\psi) \equiv \neg(Ts_{i,j}(\top,\phi\wedge\psi)\wedge Ts_{i,i}(\top,\phi\wedge\psi))$. Le constat est le même pour les axiomes T_{17} ($T_{i,j}(\phi,\psi)\wedge T_{i,k}(\psi,\psi') \Rightarrow T_{i,\{j,k\}}(\phi,\psi\wedge\psi')$) et T_{18} ($T_{i,j}(\phi,\psi), T_{j,k}(\phi',\psi'), \psi' \vdash \psi, \phi \vdash \phi'$, l'agent infère $T_{i,\{j,k\}}(\phi,\psi')$). Pour caractériser ces axiomes, il nous faudrait considérer des propriétés supplémentaires sur notre notion de confiance. Par exemple pour T_{14}, il s'agirait de considérer $Ts_{i,j}(\phi,\psi)\wedge Ts_{j,i}(\psi,\phi) \Rightarrow Td_{\{i\},\{i,j\}}(\top,\phi\wedge\psi)$.

7 Conclusion et ouvertures

Nous avons présenté dans cet article une logique modale pour exprimer la confiance qu'un agent a envers le discours d'un autre agent. Il s'agit essentiellement d'un système KD associé à un formalisme STIT. Ceci nous permet de retrouver les propriétés de la confiance dispositionnelle formalisées par Herzig *et al.*, les propriétés doxastiques de la confiance formalisées par Dundua et Uridia ainsi que de caractériser les principaux axiomes de la confiance postulés par Singh. Ce formalisme présente bien entendu des limites. Tout d'abord, nous n'avons pas présenté la sémantique et les propriétés de cohérence et complétude de notre formalisme. Toutefois, en considérant les travaux de Van Benthem [20] et ceux de Dundua et Uridia [7], notre système est intuitivement complet et cohérent si nous utilisons une sémantique de Kripke. Enfin, nous n'avons pas évoqué les aspects de complexité d'un tel modèle.

En termes de perspectives, il serait intéressant d'introduire des prédicats supplémentaires pour caractériser certains axiomes de Singh concernant la confiance collective. Nous pourrions aussi définir un prédicat de *confiance mutuelle* représenté par $T_{i,j}\phi \wedge T_{j,i}\phi \wedge T_{i,j}T_{j,i}\phi \wedge T_{j,i}T_{i,j}\phi$ ou de *confiance réciproque* par $T_{i,j}\phi \wedge T_{j,i}\phi$. De plus, nous pensons qu'il serait intéressant d'associer notre formalisme avec un système de normes sociales. En effet, la norme sociale pourrait exprimer l'axiomatique choisie par un agent pour faire confiance. Ceci apparaît en filigrane lorsque nous définissons par exemple la confiance par transitivité. Notre objectif est alors de pouvoir raisonner sur ce que certains agents malintentionnés pourraient faire s'ils considéraient ces normes.

Références

[1] Nuel Belnap and Michael Perloff. Seeing to it that : a canonical form for agentives. *Theoria*, 54(3) :175–199, 1988.

[2] Christiano Castelfranchi and Rino Falcone. *Trust theory : A socio-cognitive and computational model*. John Wiley & Sons, 2010.

[3] Brian F Chellas. Time and modality in the logic of agency. *Studia Logica*, 51(3-4) :485–517, 1992.

[4] Bruce Christianson and William Harbison. Why isn't trust transitive ? In *Security protocols*, pages 171–176, 1997.

[5] Mehdi Dastani, Andreas Herzig, Joris Hulstijn, and Leendert Van Der Torre. Inferring trust. In *5th CLIMA*, pages 144–160. Springer, 2004.

[6] Robert Demolombe. Reasoning about trust : A formal logical framework. In *2nd iTrust*, pages 291–303, 2004.

[7] Besik Dundua and Levan Uridia. Trust and belief, interrelation. In *3rd WAT*, 2010.

[8] Rino Falcone, Giovanni Pezzulo, and Cristiano Castelfranchi. A fuzzy approach to a belief-based trust computation. In *5th Workshop on Deception, Fraud and Trust in Agent Societies*, pages 73–86, 2002.

[9] Tuan-Fang Fan and Churn-Jung Liau. Reasoning about justified belief based on the fusion of evidence. In *15th JELIA*, pages 240–255. Springer, 2016.

[10] Andreas Herzig, Emiliano Lorini, Jomi Fred Hübner, and Laurent Vercouter. A logic of trust and reputation. *Logic Journal of the IGPL*, 18(1) :214–244, 2010.

[11] Audun Josang and Roslan Ismail. The beta reputation system. In *15th Bled Electronic Commerce Conference*, pages 2502–2511, 2002.

[12] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *12th WWW*, pages 640–651. ACM, 2003.

[13] Vibhor Kant and Kamal K Bharadwaj. Fuzzy computational models of trust and distrust for enhanced recommendations. *Int. J. of Intelligent Systems*, 28(4) :332–365, 2013.

[14] Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems - a modal logic formulation. *Artificial Intelligence*, 149(1) :31–60, 2003.

[15] Emiliano Lorini, Guifei Jiang, and Laurent Perrussel. Trust-based belief change. In *21st ECAI*, pages 549–554. IOS Press, 2014.

[16] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking : bringing order to the web. 1999.

[17] Yefeng Ruan and Arjan Durresi. A survey of trust management systems for online social communities – trust modeling, trust inference and attacks. *Knowledge-Based Systems*, 106 :150–163, 2016.

[18] Munindar P. Singh. Trust as dependence : A logical approach. In *10th AAMAS*, pages 863–870, 2011.

[19] Clara Smith, Agustín Ambrossio, Leandro Mendoza, and Antonino Rotolo. Combinations of normal and non-normal modal logics for modeling collective trust in normative mas. In *4th AICOL*, pages 189–203, 2011.

[20] Johan Van Benthem. Correspondence theory. In *Handbook of philosophical logic*, pages 167–247. Springer, 1984.

[21] Jin-Long Wang and Shih-Ping Huang. Fuzzy logic based reputation system for mobile ad hoc networks. In *11th KES*, pages 1315–1322, 2007.