Revue Ouverte
d'Intelligence
Artificielle

Robert Demolombe

Confiance dans l'information transmise par une séquence d'agents

Volume 5, nº 1 (2024), p. 131-151.

https://doi.org/10.5802/roia.67

La Revue Ouverte d'Intelligence Artificielle est membre du
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 2967-9672

Revue Ouverte d'Intelligence Artificielle
Volume 5, n° 1, 2024, 131-151

Confiance dans l'information transmise par une séquence d'agents

Robert Demolombe[a]

[a] Chercheur Indépendant
E-mail : robert.demolombe@orange.fr.

Résumé. — Quand un agent reçoit une information d'un autre agent, qui lui-même l'a reçue d'une séquence d'agents, la question que se pose le récepteur est : « est-ce que je peux avoir confiance dans la validité de cette information? ». La réponse à cette question dépend de la confiance qu'il a dans ces agents. Plus précisément nous considérons ici le cas où il ne connaît que l'agent qui lui a transmis directement l'information, le cas où il connaît tous les agents de la séquence, et finalement le cas où il ne connaît qu'un agent qui lui sert de référent pour l'informer sur les autres agents.

Pour modéliser les raisonnements que peut faire le récepteur nous présentons tout d'abord des définitions des différents types de propriétés des agents sur lesquelles porte la confiance : validité, sincérité, compétence, complétude, coopérativité et vigilance. Ces notions sont formalisées en logique modale (logique doxastique et logique de l'action) et on montre, d'abord sur des exemples, puis dans le cas général, les raisonnements que peut faire l'agent récepteur, sous certaines hypothèses, pour conclure sur l'information qu'il a reçue.

Mots-clés. — Communication entre agents, confiance, logiques modales.

1. Introduction

Quand un agent x transmet une information à un agent r (le récepteur), et que x dit à r par quelle séquence d'agents il a reçu cette information, il est utile pour r d'avoir des informations sur ces agents de la séquence pour savoir dans quelle mesure il peut avoir confiance dans l'information qu'il a reçue.

Le but de cet article est d'analyser dans ces situations par quels types de raisonnements r peut arriver à telle ou telle conclusion concernant l'information émise par le premier agent de la séquence.

Exemple. — Maria a dit à Robert que (John lui a dit que (Le président Peter a dit à John : « l'obligation de confinement est terminée »)).

Soit, avec des notations un peu plus précises : x = Maria, a dit à r = Robert, que (y = John, a dit à x que (z = Le président Peter, (a dit à y que ϕ))), où ϕ représente l'information « l'obligation de confinement est terminée ».

Si on accepte certaines hypothèses sur la confiance de r dans x, r peut en déduire que :

$$y \text{ a dit à } x \text{ que } (z \text{ a dit à } y \text{ que } \phi).$$

Si on accepte des hypothèses complémentaires sur la confiance dans les agents, r peut en déduire que :

$$z \text{ a dit à } y \text{ que } \phi.$$

Et si on accepte encore d'autres hypothèses sur la confiance dans les agents, r peut en déduire que : ϕ.

Il y a aussi des situations où r **n'a pas reçu** cette information de x, et où il peut en déduire :

$$x \text{ n'a pas dit à } r \text{ que } (y \text{ a dit à } x \text{ que } (z \text{ a dit à } y \text{ que } \phi)).$$

Et si on accepte d'autres hypothèses sur la confiance des agents, r peut en déduire de façon similaire : $\neg\phi$.

Ce type de situation peut être illustré avec l'exemple ci-dessus quand r n'a pas reçu de x une information exprimant le fait que « l'obligation de confinement est terminée », et quand r peut en déduire qu'il est faux que « l'obligation de confinement est terminée ».

La confiance des agents les uns par rapport aux autres peut avoir des structures différentes. Dans cet article on considère les trois types de scénarios suivants.

SCÉNARIOS DE TYPE A. — *Le récepteur r ne connaît que l'agent x qui lui a transmis l'information, et c'est sa confiance en x qui lui permet de conclure sur ϕ. Dans la séquence chaque agent a confiance dans ce que lui a transmis le précédent, et il transmet ses informations sur la confiance au suivant, jusqu'à x qui les transmet à r. La confiance de r dans x lui permet alors de conclure sur l'information transmise par le premier agent. Nous avons déjà analysé ce type de scénario dans [9] (voir section 5).*

SCÉNARIOS DE TYPE B. — *Le récepteur r connaît tous les agents de la séquence et il a confiance dans les informations qu'ils transmettent les uns aux autres. L'agent x ne transmet à r que la séquence des transmissions, depuis le premier qui a transmis ϕ. L'agent r peut alors conclure, à l'aide de la confiance qu'il a dans chaque agent, sur l'information transmise par le premier agent.*

SCÉNARIOS DE TYPE C. — *Ces scénarios sont similaires aux scénarios de type B. Mais r ne connaît pas les agents de la séquence, et les informations sur la confiance qu'il a dans ces agents lui sont transmises par un référent, rf, en qui r a confiance.*

Dans les sections suivantes seront présentés, tout d'abord, le système formel qui permet d'exprimer les informations et les règles pour raisonner sur ces informations, ensuite, une définition générale des différents types de confiance qu'un agent peut avoir dans un autre agent. C'est dans ce contexte qu'ensuite seront définis les trois types de scénarios et leurs propriétés, et les dernières sections donneront une comparaison aux autres travaux dans le même domaine avant de présenter la conclusion.

2. Système formel

Les différents scénarios sont formalisés dans le langage L de la logique définie ci-dessous.

Langage

Définition du langage L.

$$\phi \ ::= p \mid \neg\phi \mid \phi \vee \phi \mid \phi \wedge \phi \mid \mathrm{Bel}_i\, \phi \mid \mathrm{Inf}_{i,j}\, \phi$$

où p est une formule atomique, ϕ est une formule de L, et i et j dénotent des agents.

La signification intuitive de $\mathrm{Bel}_i\, \phi$ est : l'agent i croit l'information représentée par la formule ϕ.

La signification intuitive de $\mathrm{Inf}_{i,j}\, \phi$ est : l'agent i a transmis à l'agent j une information représentée par la formule ϕ.

Axiomatique

L'axiomatique de la logique exprimée dans le langage L est définie de la façon suivante :

- axiomatique du Calcul des Propositions (CP).
- pour les modalités $\mathrm{Bel}_i\, \phi$: logique (KD). Soit :

$$\mathrm{Bel}_i(\phi) \wedge \mathrm{Bel}_i(\phi \to \psi) \to \mathrm{Bel}_i(\psi) \tag{K}$$

$$\mathrm{Bel}_i(\phi) \to \neg\, \mathrm{Bel}_i(\neg\phi) \tag{D}$$

- pour $\mathrm{Inf}_{j,i}\, \phi$: Si $\vdash \phi \leftrightarrow \psi$ alors $\vdash \mathrm{Inf}_{j,i}\, \phi \leftrightarrow \mathrm{Inf}_{j,i}\, \psi$
- pour les relations entre $\mathrm{Inf}_{j,i}\, \phi$ et $\mathrm{Bel}_i\, \phi$[(1)] :

$$\mathrm{Inf}_{j,i}\, \phi \to \mathrm{Bel}_i(\mathrm{Inf}_{j,i}\, \phi) \tag{IB1}$$

Intuitivement (IB1) signifie qu'une information transmise est effectivement reçue par le destinataire.

Pour ce système axiomatique, même si nous n'avons pas jugé pertinent de l'expliciter, nous pourrions considérer par exemple une sémantique fondée sur les structures de Kripke *e.g.* $(\mathbb{W}, \mathrm{Inf}_{i,j},\ B_j)$ where $i, j \in \mathbb{N}$, où $\mathbb{W}$ est un ensemble non vide de mondes possibles, et $\mathrm{Inf}_{i,j}$ et B_j sont des ensembles de relations binaires associées pour chaque agent i et j, pour représenter la sémantique de la modalité $\mathrm{Inf}_{i,j}$ et B_j. Les

---

[(1)]Si on accepte

$$\neg\, \mathrm{Inf}_{j,i}\, \phi \to \mathrm{Bel}_i(\neg\, \mathrm{Inf}_{j,i}\, \phi), \tag{IB2}$$

on peut déduire que i croit $\neg\, \mathrm{Inf}_{j,i}\, \phi$ pour toutes formules ϕ du langage, à l'exception des formules ψ pour lesquelles on a $\mathrm{Inf}_{j,i}\, \psi$. Donc l'ensemble des formules $\mathrm{Bel}_i(\neg\, \mathrm{Inf}_{j,i}\, \phi)$ n'est pas défini parce qu'il n'est pas explicitement restreint.

Le rejet de (IB2) n'est pas contradictoire avec le fait d'accepter, dans une situation donnée, telle, ou telle, hypothèse particulière de la forme : $\neg\, \mathrm{Inf}_{j,i}\, p \to \mathrm{Bel}_i(\neg\, \mathrm{Inf}_{j,i}\, p)$, où p est une formule donnée.

propriétés de correction sont évidentes à construire en contraignant ces relations binaires (*e.g.* axiome 4 est équivalent à la transitivité, etc.) et prouver par contraposition. Puis en appliquant la méthode de preuve à la Henkin, par construction d'un modèle canonique et en prouvant le lemme de vérité, il est facile de prouver la complétude d'un tel système.

Cependant, il nous a paru plus intuitif de définir le sens de ces opérateurs à travers les raisonnements qu'ils permettent de faire, comme on l'a vu dans les scénarios présentés dans l'introduction.

3. Définitions de la confiance

Il y a de nombreuses définitions de la confiance (voir [4, 12, 15, 16]), ici nous nous basons sur la définition qui est introduite dans [6, 7, 18, 19].

La confiance d'un agent i dans un autre agent j au sujet de certaines propriétés est définie comme le fait que l'agent i croit que l'agent j satisfait ces propriétés. Formellement la confiance est représentée d'une façon générale sous la forme : $\mathrm{Bel}_i(\mathrm{proprm}_j)$.(2)

Dans le contexte de la communication d'informations les propriétés que l'on considère sont : la validité, la sincérité et la compétence, et les propriétés duales, respectivement : la complétude, la coopérativité et la vigilance. Ces propriétés sont représentées formellement de la façon suivante.

Validité

j informe i de façon valide au sujet de ϕ ssi (si j transmet l'information ϕ à i, alors ϕ est vraie). Formellement on a :

$$\mathrm{Inf}_{j,i}\,\phi \rightarrow \phi$$

Sincérité

j est sincère lorsqu'il informe i au sujet de ϕ ssi (si j transmet l'information ϕ à i, alors j croit ϕ). Formellement on a :

$$\mathrm{Inf}_{j,i}\,\phi \rightarrow \mathrm{Bel}_j(\phi)$$

Compétence

j est compétent au sujet de ϕ ssi (si j croit ϕ, alors ϕ est vraie). Formellement on a :

$$\mathrm{Bel}_j(\phi) \rightarrow \phi$$

(2) Il est important de ne pas confondre les hypothèses qui expriment, dans telle situation particulière, la confiance de tel agent dans tel autre agent, avec des schémas d'axiome qui seraient vrais dans toutes les situations pour tous les agents.

COMPLÉTUDE

j informe i de façon complète au sujet de ϕ ssi (si ϕ est vraie, alors j transmet l'information ϕ à i). Formellement on a :

$$\phi \rightarrow \mathrm{Inf}_{j,i}\,\phi$$

COOPÉRATIVITÉ

j est coopératif au sujet de ϕ ssi (si j croit que ϕ est vraie, alors j transmet l'information ϕ à i). Formellement on a :

$$\mathrm{Bel}_j(\phi) \rightarrow \mathrm{Inf}_{j,i}\,\phi$$

Ce qui peut s'exprimer de façon équivalente sous la forme :

$$\neg\,\mathrm{Inf}_{j,i}\,\phi \rightarrow \neg\,\mathrm{Bel}_j(\phi)$$

VIGILANCE

j est vigilant au sujet de ϕ ssi (si ϕ est vraie, alors j croit ϕ). Formellement on a :

$$\phi \rightarrow \mathrm{Bel}_j(\phi)$$

On notera que ces propriétés ne sont pas indépendantes. La conjonction de la sincérité et de la compétence implique la validité, et la conjonction de la vigilance et de la coopérativité implique la complétude.

4. ANALYSE DES SCÉNARIOS

4.1. SCÉNARIOS A

Par la suite on notera r l'agent terminal qui reçoit l'information et les autres agents seront dénotés par des entiers. L'agent qui a transmis l'information à r est l'agent 1, qui lui-même a reçu l'information de l'agent 2, qui l'a reçue de l'agent 3, et ainsi de suite jusqu'au premier agent qui a transmis l'information et qui est désigné par n. Donc l'information transmise par 1 exprime tout ce que les agents précédents ont transmis les uns aux autres.

Pour $i \leqslant n$ on note ϕ_i l'information que l'agent i a transmis à l'agent $i-1$.

L'exemple de l'introduction peut être représenté schématiquement de la façon suivante :

$$\mathrm{Agent}_3 \longrightarrow \phi_3 \longrightarrow \mathrm{Agent}_2 \longrightarrow \phi_2 \longrightarrow \mathrm{Agent}_1 \longrightarrow \phi_1 \longrightarrow \mathrm{Agent}_r$$

où ϕ_3 dénote la proposition : « l'obligation de confinement est terminée ».

ϕ_2 signifie que l'agent 3 a dit à l'agent 2 que « l'obligation de confinement est terminée », soit formellement : $\phi_2 = \mathrm{Inf}_{3,2}(\phi_3)$

ϕ_1 signifie que l'agent 2 a dit à l'agent 1 ϕ_2, soit formellement : $\phi_1 = \mathrm{Inf}_{2,1}(\phi_2)$, et d'après la définition de ϕ_2 : $\phi_1 = \mathrm{Inf}_{2,1}(\mathrm{Inf}_{3,2}(\phi_3))$.

Et le fait que l'agent 1 a transmis à l'agent r ϕ_1 est représenté par : $\mathrm{Inf}_{1,r}(\phi_1)$, soit $\mathrm{Inf}_{1,r}(\phi_1) = \mathrm{Inf}_{1,r}(\mathrm{Inf}_{2,1}(\mathrm{Inf}_{3,2}(\phi_3)))$. Ce qui signifie intuitivement que 1 a dit à r que 2 a dit à 1 que 3 a dit à 2 que « l'obligation de confinement est terminée ».

Dans le cas général on peut représenter de la façon suivante le fait que l'agent i a transmis à l'agent $i-1$ l'information ϕ_i, où ϕ_i concerne l'information ϕ_{i+1} que l'agent $i+1$ a transmis à i.

$$\mathrm{Agent}_{i+1} \longrightarrow \phi_{i+1} \longrightarrow \mathrm{Agent}_i \longrightarrow \phi_i \longrightarrow \mathrm{Agent}_{i-1}$$

Scénario A 1

Dans ce scénario chaque agent transmet au suivant ce que le précédent lui a transmis.

Notation. — ϕ_i : information qu'a transmise l'agent i à l'agent suivant $i-1$. On a :

$$\phi_i \stackrel{\mathrm{def}}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1}).$$

Dans l'exemple les agents sont : 3, 2, 1 et r.

Notations. — $\phi_1 \stackrel{\mathrm{def}}{=} \mathrm{Inf}_{2,1}(\phi_2)$ et $\phi_2 \stackrel{\mathrm{def}}{=} \mathrm{Inf}_{3,2}(\phi_3)$.

Le scénario est exprimé par l'hypothèse : 1 a transmis à r l'information ϕ_1, soit :

$$\mathrm{Inf}_{1,r}(\phi_1) \tag{h}$$

On en déduit :

(1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h) et (IB1)
(2) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\mathrm{Inf}_{2,1}(\mathrm{Inf}_{3,2}(\phi_3))))$ d'après (1) et les définitions de ϕ_1 et ϕ_2.

La conclusion (2) signifie que r croit que 1 lui a dit que 2 lui a dit que 3 lui a dit ϕ_3. On peut noter que dans ce scénario r connaît la succession des transmissions d'informations, mais il ne peut pas conclure sur le fait que ϕ_3 est vraie ou fausse.

Scénario A 2

Dans ce scénario chaque agent i transmet au suivant l'information ϕ_{i+1} que le précédent $i+1$ lui a transmise, et le fait que i a confiance dans la validité de $i+1$ pour ϕ_{i+1}.

On a donc :

$$\phi_i \stackrel{\mathrm{def}}{=} \mathrm{Bel}_i(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}))$$

Ce que croit i au sujet de l'agent $i+1$ est noté :

$$\psi_{i+1} \stackrel{\mathrm{def}}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1})$$

D'où : $\phi_i \stackrel{\mathrm{def}}{=} \mathrm{Bel}_i(\psi_{i+1})$

Soit, en particulier :

$$\phi_1 \stackrel{\text{def}}{=} \text{Bel}_1(\text{Inf}_{2,1}(\phi_2) \wedge (\text{Inf}_{2,1}(\phi_2) \rightarrow \phi_2))$$
$$\psi_2 \stackrel{\text{def}}{=} \text{Inf}_{2,1}(\phi_2) \wedge (\text{Inf}_{2,1}(\phi_2) \rightarrow \phi_2)$$
$$\phi_2 \stackrel{\text{def}}{=} \text{Bel}_2(\text{Inf}_{3,2}(\phi_3) \wedge (\text{Inf}_{3,2}(\phi_3) \rightarrow \phi_3))$$
$$\psi_3 \stackrel{\text{def}}{=} \text{Inf}_{3,2}(\phi_3) \wedge (\text{Inf}_{3,2}(\phi_3) \rightarrow \phi_3)$$

HYPOTHÈSES. — 1 *a transmis à r l'information* ϕ_1 *et r a confiance dans la validité de* 1 *pour* ϕ_1*, soit :*

(h1) $\text{Inf}_{1,r}(\phi_1)$
(h2) $\text{Bel}_r(\text{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$

On a en déduit :

(1) $\text{Bel}_r(\text{Inf}_{1,r}(\phi_1))$ d'après (h1) et (IB1)
(2) $\text{Bel}_r(\phi_1)$ d'après (1), (h2) et (K)
(3) $\text{Bel}_r(\text{Bel}_1(\psi_2))$ d'après (2) et la définition de ϕ_1
(4) [3] $\vdash \psi_2 \rightarrow \phi_2$ d'après (CP)
(5) $\text{Bel}_r(\text{Bel}_1(\phi_2))$ d'après (3), (4) et (KD)
(6) $\text{Bel}_r(\text{Bel}_1(\text{Bel}_2(\psi_3)))$ d'après (5) et la définition de ϕ_2
(7) $\vdash \psi_3 \rightarrow \phi_3$ d'après (CP)
(8) $\text{Bel}_r(\text{Bel}_1(\text{Bel}_2(\phi_3)))$ d'après (6), (7) et (KD)

La conclusion (8) signifie que r croit que 1 croit que 2 croit ϕ_3. Dans ce scénario r est informé sur ce que croit chaque agent dans la séquence mais il ne peut pas conclure sur le fait ϕ_3 est vraie ou fausse.

4.1.1. *Scénario A explicite*

Dans ce scénario chaque agent i transmet au suivant l'information ϕ_{i+1} que le précédent $i+1$ lui a transmis, le fait que i a confiance dans la validité de $i+1$ pour ϕ_{i+1}, et le fait que $i+1$ est compétent au sujet de ψ_{i+2}. Ce qui est formellement représenté par :

$$\phi_i \stackrel{\text{def}}{=} \text{Bel}_i(\text{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\text{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\text{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2}))$$

Ce que croit i au sujet de l'agent $i+1$ est noté :

$$\psi_{i+1} \stackrel{\text{def}}{=} \text{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\text{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\text{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2})$$

D'où : $\phi_i \stackrel{\text{def}}{=} \text{Bel}_i(\psi_{i+1})$

[3] Comme indiqué dans Chellas [5] le symbole $\vdash \phi$ signifie que ϕ est un théorème dans la logique qui a été choisie.

On a en particulier :

$$\phi_1 \stackrel{\text{def}}{=} \mathrm{Bel}_1(\mathrm{Inf}_{2,1}(\phi_2) \wedge (\mathrm{Inf}_{2,1}(\phi_2) \rightarrow \phi_2) \wedge (\mathrm{Bel}_2(\psi_3) \rightarrow \psi_3))$$

$$\psi_2 \stackrel{\text{def}}{=} \mathrm{Inf}_{2,1}(\phi_2) \wedge (\mathrm{Inf}_{2,1}(\phi_2) \rightarrow \phi_2) \wedge (\mathrm{Bel}_2(\psi_3) \rightarrow \psi_3)$$

$$\phi_1 \stackrel{\text{def}}{=} \mathrm{Bel}_1(\psi_2)$$

$$\phi_2 \stackrel{\text{def}}{=} \mathrm{Bel}_2(\mathrm{Inf}_{3,2}(\phi_3) \wedge (\mathrm{Inf}_{3,2}(\phi_3) \rightarrow \phi_3))$$

$$\psi_3 \stackrel{\text{def}}{=} \mathrm{Inf}_{3,2}(\phi_3) \wedge (\mathrm{Inf}_{3,2}(\phi_3) \rightarrow \phi_3)$$

$$\phi_2 \stackrel{\text{def}}{=} \mathrm{Bel}_2(\psi_3)$$

Remarque. — La définition de ψ_3 est différente de celle de ψ_2 parce que l'agent 3 n'a pas de prédécesseur.

HYPOTHÈSES.
(h1) $(\mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$
(h3) $\mathrm{Bel}_r(\mathrm{Bel}_1(\psi_2) \rightarrow \psi_2)$

Les hypothèses (h2) et (h3) expriment que r a confiance dans la validité de 1 pour ϕ_1 et r a confiance dans la compétence de 1 pour ψ_2.

On a :
(1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h1) et (IB1)
(2) $\mathrm{Bel}_r(\mathrm{Bel}_1(\psi_2)))$ d'après (1), (h2), (KD) et la définition de ϕ_1
(3) $\mathrm{Bel}_r(\psi_2)$ d'après (2), (h3) et (KD)
(4) $\vdash \psi_2 \rightarrow \phi_2$ d'après (CP)
(5) $\mathrm{Bel}_r(\phi_2)$ d'après (3), (4) et (KD)
(6) $\mathrm{Bel}_r(\mathrm{Bel}_2(\psi_3)))$ d'après (5) et la définition de ϕ_2
(7) $\vdash \psi_2 \rightarrow (\mathrm{Bel}_2(\psi_3) \rightarrow \psi_3)$ d'après (CP)
(8) $\mathrm{Bel}_r(\mathrm{Bel}_2(\psi_3) \rightarrow \psi_3)$ d'après (3), (7) et (KD)
(9) $\mathrm{Bel}_r(\psi_3)$ d'après (6), (8) et (KD)
(10) $\vdash \psi_3 \rightarrow \phi_3$ d'après (CP)
(11) $\mathrm{Bel}_r(\phi_3)$ d'après (9), (10) et (KD)

L'information ϕ_1 transmise par 1 exprime la totalité de la succession des informations que les agents se sont transmis les uns aux autres à partir du premier. Le raisonnement de r consiste à décomposer étape par étape l'historique de cette succession. Ainsi (3) exprime ce que r croit au sujet des croyances transmises par 2, et (9) exprime ce que r croit au sujet des croyances transmises par 3. Enfin, (11) montre que r croit que ϕ_3 est vraie.

Cas général.

Les définitions et les raisonnements que nous avons vu sur l'exemple sont exprimés dans le cas général par le théorème suivant.

THÉORÈME 4.1. — *Soit, pour* $i \in [1, n-1]$ *les définitions :*

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2}))$$

$$\psi_{i+1} \stackrel{def}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2})$$

Soit, $\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\psi_{i+1})$,

et pour $i = n$ *:*

$$\psi_n \stackrel{def}{=} \mathrm{Inf}_{n,n-1}(\phi_n) \wedge (\mathrm{Inf}_{n,n-1}(\phi_n) \rightarrow \phi_n)$$

Si on a :

(h1) $\mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$
(h3) $\mathrm{Bel}_r(\mathrm{Bel}_1(\psi_2) \rightarrow \psi_2)$

Alors on a : $\mathrm{Bel}_r(\phi_n)$.

4.1.2. *Scénario A implicite.*

Dans ce scénario le récepteur ne reçoit pas d'information et il croit que si l'information qui l'intéresse était vraie, l'agent 1 lui aurait transmis cette information. Ce qui lui permet de conclure, sous certaines hypothèses, que cette information est fausse.

Plus précisément, à partir de certaines hypothèses, et à partir de ce que n'a pas transmis 1, pour chaque agent $i+1$, r peut déduire :

- $i+1$ n'a pas transmis au suivant i l'information ϕ_{i+1}, et
- $i+1$ est complet pour ce qu'il croit (i.e. ϕ_{i+1}), et
- $i+1$ est vigilant pour ce que croit $i+2$ (*i.e.* ϕ_{i+2}).

Notation. — $\phi_i \stackrel{\text{def}}{=} \mathrm{Bel}_i(\psi_{i+1})$.

$$\neg\psi_{i+1} \stackrel{\text{def}}{=} \neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}) \wedge (\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2})$$

On a en particulier :

$$\begin{aligned}
\phi_1 &\stackrel{\text{def}}{=} \mathrm{Bel}_1(\psi_2)\\
\neg\psi_2 &\stackrel{\text{def}}{=} \neg\,\mathrm{Inf}_{2,1}(\phi_2) \wedge (\neg\,\mathrm{Inf}_{2,1}(\phi_2) \rightarrow \neg\phi_2) \wedge (\neg\,\mathrm{Bel}_2(\psi_3) \rightarrow \neg\psi_3)\\
\phi_2 &\stackrel{\text{def}}{=} \mathrm{Bel}_2(\psi_3)\\
\neg\psi_3 &\stackrel{\text{def}}{=} \neg\,\mathrm{Inf}_{3,2}(\phi_3) \wedge (\neg\,\mathrm{Inf}_{3,2}(\phi_3) \rightarrow \neg\phi_3)
\end{aligned}$$

Remarque. — La définition de $\neg\psi_3$ est différente de celle de $\neg\psi_2$ parce que l'agent 3 n'a pas de prédécesseur.

HYPOTHÈSES.

(h1) $\neg \mathrm{Inf}_{1,r}(\phi_1)$
(k1) $\neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1)$
(h3) $\mathrm{Bel}_r(\neg \mathrm{Bel}_1(\psi_2) \rightarrow \neg\psi_2)$

L'hypothèse (k1) peut être interprétée comme une propriété duale de (IB1) pour le cas particulier où 1 n'informe pas r au sujet de ϕ_1. L'hypothèse (h2) exprime la confiance de r dans la coopérativité de 1 pour ϕ_1 et l'hypothèse (h3) exprime la confiance de r dans la vigilance de 1 pour ψ_2.

On a :

(1) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1))$ d'après (h1) et (k1)
(2) $\mathrm{Bel}_r(\neg \mathrm{Bel}_1(\psi_2)))$ d'après (1), (h2), (KD) et la définition de ϕ_1
(3) $\mathrm{Bel}_r(\neg\psi_2)$ d'après (2), (h3) et (KD)
(4) $\vdash \neg\psi_2 \rightarrow \neg\phi_2$ d'après (CP)
(5) $\mathrm{Bel}_r(\neg\phi_2)$ d'après (3), (4) et (KD)
(6) $\mathrm{Bel}_r(\neg \mathrm{Bel}_2(\psi_3)))$ d'après (5) et la définition de ϕ_2
(7) $\vdash \neg\psi_2 \rightarrow (\neg \mathrm{Bel}_2(\psi_3) \rightarrow \neg\psi_3)$ d'après (CP)
(8) $\mathrm{Bel}_r(\neg \mathrm{Bel}_2(\psi_3) \rightarrow \neg\psi_3)$ d'après (3), (7) et (KD)
(9) $\mathrm{Bel}_r(\neg\psi_3)$ d'après (6), (8) et (KD)
(10) $\vdash \neg\psi_3 \rightarrow \neg\phi_3$ d'après (CP)
(11) $\mathrm{Bel}_r(\neg\phi_3)$ d'après (9), (10) et (KD)

Commentaire. — Le raisonnement de r est similaire au raisonnement présenté dans le scénario A explicite, à la différence que ϕ_1 exprime ce que croit l'agent 1 au sujet de ce qui n'a pas été transmis. Ainsi on voit que l'étape (3) dans la démonstration exprime ce que r croit au sujet de 2, et cette croyance dit : que 1 n'a pas reçu l'information ϕ_2 venant de 2, que 1 est complet vis à vis de 2 au sujet de ϕ_2, et que 2 est vigilant au sujet de ψ_3.

Cas général

THÉORÈME 4.2. — *Soit, pour $i \in [1, n-1]$ les définitions :*

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\neg \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\neg \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}) \wedge (\neg \mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2}))$$

$$\neg\psi_{i+1} \stackrel{def}{=} \neg \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\neg \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}) \wedge (\neg \mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2})$$

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\psi_{i+1}),$$

et pour $i = n$:

$$\neg\psi_n \stackrel{def}{=} \neg \mathrm{Inf}_{n,n-1}(\phi_n) \wedge (\neg \mathrm{Inf}_{n,n-1}(\phi_n) \rightarrow \neg\phi_n)$$

Si on a :
(h1) $\neg \mathrm{Inf}_{1,r}(\phi_1)$
(k1) $\neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1)$
(h3) $\mathrm{Bel}_r(\neg \mathrm{Bel}_1(\psi_2) \rightarrow \neg\psi_2)$

Alors on a : $\mathrm{Bel}_r(\neg\phi_n)$.

4.2. Scénarios B

Dans ces scénarios on suppose que r connaît chacun des agents de la séquence et qu'il a confiance dans leur validité pour l'information qu'ils transmettent.

4.2.1. *Scénario B explicite*

Notation. — $\phi_i \stackrel{\text{def}}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1})$

Les hypothèses sur la confiance de r dans la validité de chaque agent de la séquence sont de la forme :

$$\mathrm{Bel}_r(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \tag{h_i}$$

On a en particulier :

$$\phi_1 \stackrel{\text{def}}{=} \mathrm{Inf}_{2,1}(\phi_2)$$
$$\phi_2 \stackrel{\text{def}}{=} \mathrm{Inf}_{3,2}(\phi_3)$$

Hypothèses.
(h) $\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$
(h2) $\mathrm{Bel}_r(\mathrm{Inf}_{2,1}(\phi_2) \rightarrow \phi_2)$
(h3) $\mathrm{Bel}_r(\mathrm{Inf}_{3,2}(\phi_3) \rightarrow \phi_3)$

On a :
(1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h) et (IB1)
(2) $\mathrm{Bel}_r(\phi_1)$ d'après (1), (h2) et (KD)
(3) $\mathrm{Bel}_r(\mathrm{Inf}_{2,1}(\phi_2))$ d'après (2) et la définition de ϕ_1
(4) $\mathrm{Bel}_r(\phi_2)$ d'après (3), (h2) et (KD)
(5) $\mathrm{Bel}_r(\mathrm{Inf}_{3,2}(\phi_3))$ d'après (4) et la définition de ϕ_2
(6) $\mathrm{Bel}_r(\phi_3)$ d'après (5), (h3) et (KD)

Théorème 4.3. — *Si on a :*
(h) $\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$.
Pour $i \in [1, n-1]$ *(h_i)* $\mathrm{Bel}_r(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1})$.

Alors on a : $\mathrm{Bel}_r(\phi_n)$*.*

4.2.2. *Scénario B implicite*

Dans ce scénario la confiance de r porte sur la complétude des agents.

On a les définitions :

$$\phi_i \stackrel{\text{def}}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1})$$

Les hypothèses sur la confiance de r dans la complétude de chaque agent de la séquence sont de la forme :

$$\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \to \neg\phi_{i+1}) \qquad (h_i)$$

Soit en particulier :

$$\phi_1 \stackrel{\text{def}}{=} \mathrm{Inf}_{2,1}(\phi_2)$$
$$\phi_2 \stackrel{\text{def}}{=} \mathrm{Inf}_{3,2}(\phi_3)$$

HYPOTHÈSES.

(h) $\neg\,\mathrm{Inf}\ 1, r(\phi_1)$
(k) $\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \neg\phi_1)$
(h2) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{2,1}(\phi_2) \to \neg\phi_2)$
(h3) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{3,2}(\phi_3) \to \neg\phi_3)$

On a :

(1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h) et (k)
(2) $\mathrm{Bel}_r(\neg\phi_1)$ d'après (1), (h1) et (KD)
(3) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{2,1}(\phi_2))$ d'après (2) et la définition de ϕ_1
(4) $\mathrm{Bel}_r(\neg\phi_2)$ d'après(3), (h2) et (KD)
(5) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{3,2}(\phi_3))$ d'après (4) et la définition de ϕ_2
(6) $\mathrm{Bel}_r(\neg\phi_3)$ d'après(5), (h3) et (KD)

THÉORÈME 4.4. — *Si on a :*

(h) $\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(k) $\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \neg\phi_1)$
Pour $i \in [1, n-1]$ *(h_i)* $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \to \neg\phi_{i+1})$

Alors on a : $\mathrm{Bel}_r(\neg\phi_n)$*.*

4.3. SCÉNARIOS C

On suppose que r ne connaît pas les agents de la séquence, mais r connaît un agent rf qui lui sert de référent pour l'informer sur les agents de la séquence, et r a confiance dans l'information que lui transmet le référent rf.

4.3.1. *Scénario C explicite*

On a les définitions :

$$k_0 \stackrel{\text{def}}{=} \text{Inf}_{1,r}(\phi_1) \rightarrow \phi_1$$

Pour $i \in [1, n]$: $k_{i+1} \stackrel{\text{def}}{=} \text{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}$

$$k \stackrel{\text{def}}{=} \bigwedge_{i \in [0,n]} k_i$$

Intuitivement k exprime que tous les agents de la séquence sont valides pour l'information qu'ils transmettent au suivant.

Soit en particulier :

$$k_0 \stackrel{\text{def}}{=} \text{Inf}_{1,r}(\phi_1) \rightarrow \phi_1$$
$$k_1 \stackrel{\text{def}}{=} \text{Inf}_{2,1}(\phi_2) \rightarrow \phi_2$$
$$k_2 \stackrel{\text{def}}{=} \text{Inf}_{3,2}(\phi_3) \rightarrow \phi_3$$

HYPOTHÈSES.

(h′) $\text{Inf}_{rf,r}(\text{Bel}_{rf}(k))$
(h′1) $\text{Bel}_r(\text{Inf}_{rf,r}(\text{Bel}_{rf}(k)) \rightarrow \text{Bel}_{rf}(k))$
(h′2) $\text{Bel}_r(\text{Bel}_{rf}(k) \rightarrow k)$

L'hypothèse (h′1) exprime que r a confiance dans la validité du référent *rf* au sujet de ce que le référent croit sur la validité des agents de la séquence à propos de ce qu'ils transmettent au suivant. L'hypothèse (h′2) exprime que r a confiance dans la compétence de *rf* au sujet de ce qu'il croit sur la validité des agents de la séquence.

On a :

(1) $\text{Bel}_r(\text{Inf}_{rf,r}(\text{Bel}_{rf}(k)))$ d'après (h′) et IB1
(2) $\text{Bel}_r(\text{Bel}_{rf}(k))$ d'après (1), (h′1) et (KD)
(3) $\text{Bel}_r(k)$ d'après (2),n (h′2) et (KD)
(h1) $\text{Bel}_r(\text{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$ d'après (3), (KD) et la définition de k_1
(h2) $\text{Bel}_r(\text{Inf}_{2,1}(\phi_2) \rightarrow \phi_2)$ d'après (3) (KD) et la définition de k_2
(h3) $\text{Bel}_r(\text{Inf}_{3,2}(\phi_3) \rightarrow \phi_3)$ d'après (3) (KD) et la définition de k_3

Les conclusions (h1), (h2) et (h3) montrent que le référent *rf* a transmis à r les hypothèses qu'accepte r dans le scénario B explicite pour déduire $\text{Bel}_r(\phi_3)$.

4.3.2. *Scénario C implicite*

On a les définitions :

$$k_0 \stackrel{\text{def}}{=} \neg\,\text{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1$$

Pour $i \in [1,n]$: $k_{i+1} \stackrel{\text{def}}{=} \neg \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}$

$$k \stackrel{\text{def}}{=} \bigwedge_{i\in[0,n]} k_i$$

Soit en particulier :

$$k_0 \stackrel{\text{def}}{=} \neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1$$
$$k_1 \stackrel{\text{def}}{=} \neg \mathrm{Inf}_{2,1}(\phi_2) \rightarrow \neg\phi_2$$
$$k_2 \stackrel{\text{def}}{=} \neg \mathrm{Inf}_{3,2}(\phi_3) \rightarrow \neg\phi_3$$

HYPOTHÈSES.

(h′) $\mathrm{Inf}_{rf,r}(\mathrm{Bel}_{rf}(k))$
(h′1) $\mathrm{Bel}_r(\mathrm{Inf}_{rf,r}(\mathrm{Bel}_{rf}(k)) \rightarrow \mathrm{Bel}_{rf}(k))$
(h′2) $\mathrm{Bel}_r(\mathrm{Bel}_{rf}(k) \rightarrow k)$

On a :

(1) $\mathrm{Bel}_r(\mathrm{Inf}_{rf,r}(\mathrm{Bel}_{rf}(k)))$ d'après (h′) et IB1
(2) $\mathrm{Bel}_r(\mathrm{Bel}_{rf}(k))$ d'après (1), (h′1) et (KD)
(3) $\mathrm{Bel}_r(k)$ d'après (2), (h′2) et (KD)
(h1) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1)$ d'après (3), (KD) et la définition de k_1
(h2) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{2,1}(\phi_2) \rightarrow \neg\phi_2)$ d'après (3), (KD) et la définition de k_2
(h3) $\mathrm{Bel}_r(\neg \mathrm{Inf}_{3,2}(\phi_3) \rightarrow \neg\phi_3)$ d'après (3), (KD) et la définition de k_3

Les conclusions (h1), (h2) et (h3) sont les hypothèses utilisées par r dans le scénarios B implicite pour déduire $\mathrm{Bel}_r(\neg\phi_3)$.

5. COMPARAISONS

Dans [17] Leturc et Bonnet définissent un opérateur de confiance basé uniquement dans la confiance dans la sincérité. Cet opérateur ne fait pas intervenir la notion d'action de communication. On peut dire intuitivement qu'il exprime le fait que dans une **situation donnée** où une action de communication a été réalisée un agent i a confiance dans un agent j, tandis que la définition de la confiance que nous avons présentée s'applique **avant** que l'action de communication soit réalisée.

Drawel *et al.* dans [11] proposent un formalisme qui permet de raisonner sur la confiance d'un agent ou d'un groupe d'agents dans un autre agent. Il permet, en particulier, d'analyser la propagation de la confiance dans le temps. Cependant la notion de confiance n'explicite pas les propriétés sur lesquelles porte la confiance, telle que la sincérité ou la compétence, ni les actions de communication entre agents comme c'est le cas dans ce que nous avons présenté.

Sakama *et al.* dans [24] analysent et comparent différentes notions liées à des actes de communication qui ont pour effet que le récepteur croit, ou ne croit pas, qu'une proposition est vraie. On peut, pour simplifier, dire que dans certains cas un agent informe un autre agent, et dans d'autres cas trompe un autre agent. La démarche présentée dans cet article a des points communs avec notre approche : elle exprime les notions de façon précise dans des logiques modales dont les opérateurs essentiels sont : la croyance, l'intention et l'acte de communication. Une différence très importante est que les auteurs analysent certaines propriétés des agents, mais ne considère pas le fait qu'un agent croit qu'un autre agent a telle ou telle propriété, ce qui montre que la notion de confiance elle-même n'intervient pas.

Primiero dans [21] propose des définitions de la confiance fondée sur la consistance de l'information transmise par l'agent émetteur avec les informations dont dispose le récepteur. La confiance dans cet article est considérée comme une notion globale et ne fait pas intervenir des propriétés telles que la compétence ou la sincérité, ou les propriétés duales que nous avons présentées.

Il faut noter que les travaux présentés dans [11, 17, 21, 24] ne font pas intervenir de séquence d'actes de communication.

Dans [9] les différentes formes de confiance que nous avons définies sont les mêmes que celles qui sont présentées ici. Les théorèmes 4 et 5 de cet article présentent des formes de transitivité similaires à celles présentées ici dans les scénarios de type A, mais dans une forme plus simple. Les théorèmes 5 et 6 présentent une autre forme de transitivité dans laquelle la confiance dans chaque agent de la séquence est déduite par r de la confiance que r a dans les référents de chaque agent de la séquence. Ce type de situation est différent des scénarios de type C, car dans les scénarios de type C r a un seul référent pour tous les agents de la séquence. Par ailleurs les scénarios de type B et C ne sont pas envisagés dans [9].

Les travaux sur la confiance dans les sciences cognitives [4, 12] et en philosophie [2, 15, 16] montrent qu'il y a de nombreuses définitions de la confiance. Nous allons examiner des travaux dans lesquels les définitions adoptées pour formaliser la transitivité ont certains points communs avec ce que nous avons présenté.

Dans [14] Huang et Fox proposent une notion de transitivité basée sur une définition différente de la confiance. Deux sortes de confiance sont présentées. La première exprimée sous la forme : trust $-b(d, e, x, k)$ signifie que l'agent d a confiance dans le fait que la croyance de l'agent e dans l'information représentée par x est valide dans le contexte k, où x représente une formule quelconque du Calcul des Propositions. La seconde exprimée sous la forme : trust $-p(d, e, x, k)$ signifie que l'agent d a confiance dans le fait que la situation où x est vraie a été causée par une action réalisée par l'agent e dans le contexte k.

Dans ce formalisme la propriété de transitivité de la confiance (voir Théorème 8 dans [14]) exprime que si (l'agent d a confiance dans toutes les informations x que croit c et l'agent c a confiance dans l'agent e pour la capacité de causer n'importe quelle proposition x), alors (d a confiance dans e pour la capacité de causer n'importe quelle proposition x).

On peut noter que cette propriété de transitivité est basée sur des hypothèses très fortes qui paraissent peu vraisemblables dans une situation réelle. En effet, il serait étonnant qu'un agent ait une confiance de type trust$-b(d,e,x,k)$ et de type trust$-p(d,e,x,k)$ pour n'importe quelle information x du Calcul des Propositions, quel que soit le thème sur lequel porte cette information.

Par ailleurs, dans cette étude, il n'est pas fait référence à une séquence d'actions d'information, mais à des relations de confiance entre agents.

Il en est de même dans les nombreux travaux sur la transitivité auxquels nous avons eu accès qui, par ailleurs, n'explicitent pas la définition de la confiance. La confiance est simplement représentée par une valeur numérique (voir [3, 13, 20, 22]). Par exemple, dans [23] la confiance entre deux agents est représentée par une relation entre deux agents à laquelle est associée une probabilité, et la transitivité s'exprime par un calcul de propagation de ces probabilités en suivant ces relations.

Le travail présenté dans [3] a des points communs avec ce que nous avons présenté. Il s'appuie sur une notion de confiance qui est basée sur la Subjective Logic définie par Jøsang qui combine des probabilités et une logique épistémique dans laquelle les informations ne sont représentées que par des propositions atomiques.

En fait dans ces travaux la confiance est implicitement supposée transitive et le modèle consiste à propager des paramètres numériques en suivant ces relations.

6. Conclusion

La logique modale permet de donner des définitions précises et détaillées de la confiance. On peut choisir d'autres définitions que celles que nous avons présentées, mais si on accepte ces définitions on peut déduire de façon rigoureuse les conséquences d'hypothèses qui décrivent des situations complexes.

Plus précisément nous avons considéré des situations où un agent reçoit des informations qui ont été transmises par une séquence d'agents. Plusieurs scénarios ont été envisagés selon que le dernier récepteur à tel ou tel type de confiance dans le dernier agent qui lui a transmis l'information, ou quand il a certains types de confiance dans chacun des agents qui sont intervenus dans la séquence, ou dans un référent de confiance qui l'informe sur les agents de la séquence.

On peut raffiner ces définitions de la confiance en considérant des analyses plus fines où on n'a pas simplement les cas où un agent fait confiance ou ne fait pas confiance, mais où un agent peut avoir différents degrés de confiance comme définis dans [1, 8, 10]. Dans ces définitions on considère des relations d'ordre qualitatives entre les degrés de confiance, à la différence des définitions quantitatives que nous avons mentionnées plus haut. Un prolongement des travaux présentés dans cet article consisterait à les reprendre entièrement sur la base de ces degrés de confiance qualitatifs.

Annexe

Théorème 4.1. — *Soit, pour $i \in [1, n-1]$ les définitions :*

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2}))$$

$$\psi_{i+1} \stackrel{def}{=} \mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1}) \wedge (\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2})$$

Soit, $\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\psi_{i+1})$, *et pour* $i = n$ *:*

$$\psi_n \stackrel{def}{=} \mathrm{Inf}_{n,n-1}(\phi_n) \wedge (\mathrm{Inf}_{n,n-1}(\phi_n) \rightarrow \phi_n)$$

Si on a :

(h1) $\mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$
(h3) $\mathrm{Bel}_r(\mathrm{Bel}_1(\psi_2) \rightarrow \psi_2)$

Alors on a : $\mathrm{Bel}_r(\phi_n)$.

Démonstration. — On a :

(1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h1) et (IB1)
(2) $\mathrm{Bel}_r(\mathrm{Bel}_1(\psi_2)))$ d'après (1), (h2), (KD) et définition de ϕ_1
(3) $\mathrm{Bel}_r(\psi_2)$ d'après (2), (h3) et (KD)

Preuve par induction sur $i+1$. — Si (1) $\mathrm{Bel}_r(\psi_{i+1})$ alors

(2) $\vdash \psi_{i+1} \rightarrow \phi_{i+1}$ d'après définition de ψ_{i+1} et (CP)
(3) $\mathrm{Bel}_r(\phi_{i+1})$ d'après (1), (2) et (KD)
(4) $\mathrm{Bel}_r(\mathrm{Bel}_{i+1}(\psi_{i+2}))$ d'après définition de ϕ_{i+1}
(5) $\vdash \psi_{i+1} \rightarrow (\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2})$ d'après définition de ψ_{i+1} et (CP)
(6) $\mathrm{Bel}_r(\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \psi_{i+2})$ d'après (1), (5) et (KD)
(7) $\mathrm{Bel}_r(\psi_{i+2})$ d'après (4), (6) et (KD)

Donc $\mathrm{Bel}_r(\psi_{i+1})$ implique par induction $\mathrm{Bel}_r(\psi_{i+2})$, et on en déduit : $\mathrm{Bel}_r(\psi_2)$ implique (8) $\mathrm{Bel}_r(\psi_n)$.

On a :

(9) $\vdash \psi_n \rightarrow \phi_n$ d'après définition de ψ_n

Donc $\mathrm{Bel}_r(\phi_n)$ d'après (8), (9) et (KD) □

Théorème 4.2. — *Soit, pour $i \in [1, n-1]$ les définitions :*

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}) \\ \wedge (\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2}))$$

$$\neg\psi_{i+1} \stackrel{def}{=} \neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \wedge (\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \neg\phi_{i+1}) \wedge (\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2})$$

$$\phi_i \stackrel{def}{=} \mathrm{Bel}_i(\psi_{i+1}),$$

et pour i = n :

$$\neg\psi_n \stackrel{def}{=} \neg\,\mathrm{Inf}_{n,n-1}(\phi_n) \wedge (\neg\,\mathrm{Inf}_{n,n-1}(\phi_n) \rightarrow \neg\phi_n)$$

Si on a :

(h1) $\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(k1) $\neg\,\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(h2) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \neg\phi_1)$
(h3) $\mathrm{Bel}_r(\neg\,\mathrm{Bel}_1(\psi_2) \rightarrow \neg\psi_2)$

Alors on a : $\mathrm{Bel}_r(\neg\phi_n)$.

Démonstration. — On a :

(1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h1) et (k1)
(2) $\mathrm{Bel}_r(\neg\,\mathrm{Bel}_1(\psi_2)))$ d'après (1), (h2), (KD) et définition de ϕ_1
(3) $\mathrm{Bel}_r(\neg\psi_2)$ d'après (2), (h3) et (KD)

Preuve par induction sur i + 1. — Si (1) $\mathrm{Bel}_r(\neg\psi_{i+1})$ alors

(2) $\vdash \neg\psi_{i+1} \rightarrow \neg\phi_{i+1}$ d'après définition de $\neg\psi_{i+1}$ et (CP)
(3) $\mathrm{Bel}_r(\neg\phi_{i+1})$ d'après (1), (2) et (KD)
(4) $\mathrm{Bel}_r(\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}))$ d'après définition de ϕ_{i+1}
(5) $\vdash \neg\psi_{i+1} \rightarrow (\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2})$ d'après définition de $\neg\psi_{i+1}$ et (CP)
(6) $\mathrm{Bel}_r(\neg\,\mathrm{Bel}_{i+1}(\psi_{i+2}) \rightarrow \neg\psi_{i+2})$ d'après (1), (5) et (KD)
(7) $\mathrm{Bel}_r(\neg\psi_{i+2})$ d'après (4), (6) et (KD)

Donc $\mathrm{Bel}_r(\neg\psi_{i+1})$ implique $\mathrm{Bel}_r(\neg\psi_{i+2})$, et on en déduit par induction : $\mathrm{Bel}_r(\neg\psi_2)$ implique (8) $\mathrm{Bel}_r(\neg\psi_n)$.

On a :

(9) $\vdash \neg\psi_n \rightarrow \neg\phi_n$ d'après définition de $\neg\psi_n$

Donc $\mathrm{Bel}_r(\neg\phi_n)$ d'après (8), (9) et (KD) □

THÉORÈME 4.3. — *Si on a :*

(h) $\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$
Pour $i \in [1, n-1]$ *(h_i)* $\mathrm{Bel}_r(\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \rightarrow \phi_{i+1})$

Alors on a : $\mathrm{Bel}_r(\phi_n)$.

Démonstration. — On a :

(1) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h) et (IB1)
(2) $\mathrm{Bel}_r(\mathrm{Inf}_{1,r}(\phi_1) \rightarrow \phi_1)$ d'après (h1)
(3) $\mathrm{Bel}_r(\phi_1)$ d'après (1), (2) et (KD)

Preuve par induction sur i. — Si on a (1) $\mathrm{Bel}_r(\phi_{i-1})$ alors on a :
(2) $\mathrm{Bel}_r(\mathrm{Inf}_{i,i-1}(\phi_i))$ d'après la définition de ϕ_{i-1}
(3) $\mathrm{Bel}_r(\mathrm{Inf}_{i,i-1}(\phi_i) \to \phi_i)$ d'après (h_{i-1})
(4) $\mathrm{Bel}_r(\phi_i)$ d'après (1), (2) et (KD)

Donc on a : $\mathrm{Bel}_r(\phi_n)$. □

THÉORÈME 4.4. — *Si on a :*
(h) $\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(k) $\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1)$
(h1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \neg\phi_1)$
Pour $i \in [1, n-1]$ (h_i) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{i+1,i}(\phi_{i+1}) \to \neg\phi_{i+1})$

Alors on a : $\mathrm{Bel}_r(\neg\phi_n)$*.*

Démonstration. — On a :
(1) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1))$ d'après (h) et (k)
(2) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{1,r}(\phi_1) \to \neg\phi_1)$ d'après (h1)
(3) $\mathrm{Bel}_r(\neg\phi_1)$ d'après (1), (2) et (KD)

Preuve par induction sur i. — Si on a (1) $\mathrm{Bel}_r(\neg\phi_{i-1})$ alors on a :
(2) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{i,i-1}(\phi_i))$ d'après la définition de ϕ_{i-1}
(3) $\mathrm{Bel}_r(\neg\,\mathrm{Inf}_{i,i-1}(\phi_i) \to \neg\phi_i)$ d'après (h_{i-1})
(4) $\mathrm{Bel}_r(\neg\phi_i)$ d'après (1), (2) et (KD)

Donc on a : $\mathrm{Bel}_r(\neg\phi_n)$. □

REMERCIEMENTS

Nous remercions Luis Fariñas del Cerro pour tous les commentaires très pertinents qu'il a fait sur notre travail. Nos remerciements vont également aux trois rapporteurs dont les commentaires ont beaucoup contribué à la qualité de l'article.

BIBLIOGRAPHIE

[1] L. AMGOUD & R. DEMOLOMBE, « An argumentation-based approach for reasoning about trust in information sources », *Argument & Computation* **5** (2014), nº 2-3, p. 191-215.

[2] M. BACHARACH & D. GAMBETTA, « Trust as Type Detection », in *Trust and Deception in Virtual Societies* (C. Castelfranchi & Y.-H. Tan, éds.), Springer Netherlands, 2001, p. 1-26.

[3] T. BHUIYAN, A. JOSANG & Y. XU, « An Analysis of Trust Transitivity Taking Base Rate into Account », in *2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, 2009, p. 34-39.

[4] C. CASTELFRANCHI & R. FALCONE, *Trust Theory : A Socio-Cognitive and Computational Model*, Wiley, 2010.

[5] B. F. CHELLAS, *Modal Logic : An introduction*, Cambridge University Press, 1988.

[6] R. DEMOLOMBE, « To Trust Information Sources : A Proposal for a Modal Logical Framework », in *Trust and Deception in Virtual Societies* (C. Castelfranchi & Y.-H. Tan, éds.), Springer Netherlands, 2001, p. 111-124.

[7] ———, « Reasoning About Trust : A Formal Logical Framework », in *Trust Management* (C. Jensen, S. Poslad & T. Dimitrakos, éds.), Springer, 2004, p. 291-303.

[8] ———, « Graded Trust », in *Proceedings of the Trust in Agent Societies Workshop at AAMAS 2009* (R. Falcone, S. Barber, J. Sabater-Mir & M. Singh, éds.), 2009.

[9] ———, « Transitivity and Propagation of Trust in Information Sources : An Analysis in Modal Logic », in *Computational Logic in Multi-Agent Systems, LNAI 6814* (J. Leite, P. Torroni, T. Ågotnes, G. Boella & L. van der Torre, éds.), Springer, 2011, p. 13-28.

[10] R. Demolombe & C.-j. Liau, « A Logic of Graded Trust and Belief Fusion », in *Proc. of the 4th Workshop on Deception, Fraud and Trust in Agent Societies* (C. Castelfranci & R. Falcone, éds.), 2001.

[11] N. Drawel, J. Bentahar, A. Laarej & G. Rjoub, « newblock Formal verification of group and propagated trust in multi-agent systems », *em Auton. Agents Multi Agent Syst.* **36** (2022), nº 1, article no. 19.

[12] R. Falcone & C. Castelfranchi, « Social Trust : A Cognitive Approach », in *Trust and Deception in Virtual Societies* (C. Castelfranchi & Y.-H. Tan, éds.), Springer Netherlands, 2001, p. 55-90.

[13] C.-W. Hang, Y. Wang & M. P. Singh, « Operators for propagating trust and their evaluation in social networks », in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '09, vol. 2, International Foundation for Autonomous Agents and Multiagent Systems, 2009, p. 1025–1032.

[14] J. Huang & M. S. Fox, « An ontology of trust : formal semantics and transitivity », in *International Conference on Evolutionary Computation*, 2006.

[15] A. J. I. Jones, « On the concept of trust », *Decision Support Systems* **33** (2002), nº 3, p. 225-232.

[16] A. J. I. Jones & B. S. Firozabadi, « On the Characterisation of a Trusting Agent. Aspects of a Formal Approach », in *Trust and Deception in Virtual Societies* (C. Castelfranchi & Y.-H. Tan, éds.), Springer Netherlands, 2001, p. 157-168.

[17] C. Leturc & G. Bonnet, « A Normal Modal Logic for Trust in the Sincerity », in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, 2018, p. 175-183.

[18] E. Lorini & R. Demolombe, « Trust and Norms in the Context of Computer Security : A Logical Formalization », in *Deontic Logic in Computer Science* (R. van der Meyden & L. van der Torre, éds.), Springer, 2008, p. 50-64.

[19] ———, « From Trust in Information Sources to Trust in Communication Systems : An Analysis in Modal Logic », in *Knowledge Representation for Agents and Multi-Agent Systems KRAMAS 2008* (J.-J. C. Meyer & J. Broersen, éds.), Lecture Notes in Computer Science, vol. 5605, Springer, 2009, p. 81-98.

[20] N. Osman, C. Sierra & J. Sabater-Mir, « Propagation of Opinions in Structural Graphs », in *Proceedings of the 2010 Conference on ECAI 2010 : 19th European Conference on Artificial Intelligence*, IOS Press, 2010, p. 595–600.

[21] G. Primiero, « A logic of negative trust », *Journal of Applied Non-Classical Logics* **30** (2020), nº 3, p. 193-222.

[22] D. Quercia, S. Hailes & L. Capra, « Lightweight Distributed Trust Propagation », in *2007 7th IEEE International Conference on Data Mining (ICDM 2007)*, IEEE Computer Society, 2007, p. 282-291.

[23] O. Richters & T. P. Peixoto, « Trust transitivity in social networks », Technical report, Darmstadt Technical University, 2010.

[24] C. Sakama, M. Caminada & A. Herzig, « A formal account of dishonesty », *Logic Journal of the IGPL* **23** (2014), nº 2, p. 259-294.

ABSTRACT. — When an agent receives information from another agent, who in turn has received it from a sequence of agents, the question the receiver asks himself is: "Can I trust the validity of this information?". The answer to this question depends on the trust he has in these agents. More precisely, we consider here the case where he knows only the agent who directly transmitted the information to him, the case where he knows all the agents in the sequence, and finally the case where he knows only one agent who serves as a referent to inform him about the other agents.

To model the reasoning that the receiver can make, we first present definitions of the different types of agent properties on which trust is based: validity, sincerity, competence, completeness, cooperativeness and vigilance. These notions are formalized in modal logic (doxastic logic and action logic) and we show, first on examples, then in the general case, the reasoning that the receiving agent can make, under certain hypotheses, to conclude on the information he has received.

KEYWORDS. — Agent communication, Trust, Modal logics.

Manuscrit reçu le 16 janvier 2023, accepté le 29 mars 2024.