HAL open science

La confiance comme posture de crédulité

Jean-Pascal Palus, Adrien Revault d'Allonnes, Nicolas Jouandeau

To cite this version:

Jean-Pascal Palus, Adrien Revault d'Allonnes, Nicolas Jouandeau. La confiance comme posture de crédulité. Rencontres Francophones sur la Logique Floue et ses Applications (LFA) 2022, Oct 2022, Toulouse, France. hal-03832463

HAL Id: hal-03832463

https://hal.science/hal-03832463v1

Submitted on 27 Oct 2022

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

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

La confiance comme posture de crédulité

Trust as a credulity stance

J.P. Palus, A. Revault d'Allonnes & N. Jouandeau
LIASD, PASTIS, Université Paris 8

{jpp,ara,n}@up8.edu

Résumé :
La confiance est un concept étudié par de nombreuses disciplines et dont il existe un grand nombre de définitions parfois incompatibles entre elles. Dans ce papier nous essaierons tout d'abord de montrer qu'il existe des points communs entre ces différentes définitions, puis nous proposerons un formalisme modal permettant de représenter ce qui nous apparaît comme une définition minimale de la confiance entre deux agents.

Mots-clés :
Confiance, défiance, méfiance, logique modale

Abstract:
Trust is a concept studied by many disciplines and for which there exist numerous, sometimes incompatible definitions. In this paper we will first try to show that there are commonalities between these different definitions, then we will propose a modal formalism of what appears to us a minimal definition of trust between two agents.

Keywords:
Trust, distrust, mistrust, modal logic

1 Introduction

La confiance est un concept qui joue un rôle dans de nombreuses disciplines scientifiques. L'une des définitions les plus citées de la confiance stipule qu'elle "est un état psychologique comprenant l'intention d'accepter la vulnérabilité basée sur des attentes positives des intentions ou des comportements d'autrui" [25]. En d'autres mots, c'est un état mental dans lequel se trouve un agent A qui, lorsque confronté à un autre agent B sur le point d'accomplir une action, décide de croire que B accomplira l'action d'une manière qui maximisera les bénéfices pour A.

Le terme *agent* tel qu'utilisé ci-après décrit tout agent réel ou virtuel modélisant le comportement humain. Si la généralisation de ce concept à tous les agents intelligents est un travail intéressant il n'est pas l'objet de ce papier qui se veut proposer un modèle minimal de la confiance que peut exprimer un être humain envers un autre être humain.

La confiance est à la fois un état mental et une position épistémique que l'on pourrait qualifier de second ordre car faisant intervenir un ensemble de croyances ainsi qu'un élément volontaire qui la distingue de l'espérance.

L'agent accordant sa confiance sera qualifié de *trustor* et celui à qui la confiance est accordée, ou pas, de *trustee*.

Les nombreuses approches de la confiance et des concepts connexes que sont la méfiance et la défiance (qui désignent comme nous le verrons non pas la simple absence de confiance mais des positions épistémiques négatives et neutres quant à l'accomplissement de l'action par le *trustee*), tentent de répondre aux besoins et perspectives spécifiques des disciplines qui en sont à l'origine.

Dans la section suivante nous présenterons un bref état de l'art de ces disciplines dans leur étude du concept de confiance. Nous proposerons ensuite une définition minimale, issue des points communs pouvant être établis entre ces différentes approches. Pour finir nous proposerons un formalisme modal permettant de représenter formellement cette définition.

2 Confiance interdisciplinaire

La définition précédemment utilisée est celle couramment utilisée en philosophie [31] et en science de l'éducation [29]. La psychologie, de son côté, tend à se concentrer sur les relations interpersonnelles intimes et voit dans

la confiance la croyance que le *trustee* ne va pas chercher à faire du mal au *trustor* [24]. L'économie la voit souvent comme la construction sociale poussant l'homme à généralement ne pas avoir de comportements égoïstes par évaluation d'une balance bénéfices/risques [6]. Pour la sociologie, la confiance marque la volonté du *trustor* à être vulnérable à la possibilité que le *trustee* puisse lui causer du tort, tout en croyant que ce dernier ne fera pas une telle chose [27]. Les sciences politiques [11], les sciences du management [26] et parfois les neurosciences [12] se concentrent sur la facette organisationnelle de la confiance et la définissent généralement comme le jugement collectif d'un groupe d'agents sur la bonne volonté et le respect de règles établies par un autre groupe d'agents.

Bien d'autres ont tenté d'en proposer des modèles universels, bien que parfois incompatibles, alors que certains au contraire doutent qu'un tel modèle puisse être inventé [14] ou que l'usage familier du concept est trop vague pour être défini de manière exhaustive [17].

3 Caractéristiques communes entre ces différentes approches

Certains proposent de confiner la définition de la confiance à un domaine particulier, comme l'information journalistique [5] ou la hiérarchie militaire [20], ou encore de se restreindre à ne définir formellement qu'une partie du concept comme la confiance en la sincérité [13]. D'autres se sont penchés sur les caractéristiques communes qui pouvaient être extraites des approches propres aux disciplines qui ont étudiées ce concept.

Rousseau et al. [25] reconnaissent plusieurs sous-concepts qui forment des composantes de la confiance.

Ils distinguent d'abord la part utilitariste de la confiance, qui est un mécanisme qui se mettrait en place quand le *trustor* dispose d'un moyen de pression sur le *trustee* et qui identifierait la confiance comme un substitut à un mécanisme de contrôle, une "attitude positive sur les motivations" d'un agent. Ils identifient ensuite la partie rationnelle de la confiance qui repose sur l'acquisition préalable d'information de la part du *trustor* à propos du *trustee* ; des informations comme sa compétence, sa crédibilité ou ses qualifications. Enfin, ils proposent le concept de confiance relationnelle qui décrit à la fois la confiance établie par une absence de trahison sur la durée par le *trustee* mais également la part émotionnelle des relations interpersonnelles. La confiance relationnelle est celle qui s'établirait entre deux agents affectivement proches ou entre une institution et ses employés. On constate que, pour Rousseau et al., les *trustor* et *trustee* peuvent être autant des personnes physiques que morales.

Pour Pursiainen et Forsberg [22], la confiance est une forme d'espérance et implique donc une croyance sur les intentions du *trustee* et la possibilité pour le *trustor* de s'exposer à sa trahison. La confiance implique également l'absence de contrôle sur les actions du *trustee*. De manière connexe ils identifient la méfiance comme l'absence de confiance et la défiance comme l'opposé de la confiance : une attitude dans laquelle le *trustor* s'attendrait à une issue négative et où la trahison serait remplacée par de la surprise.

Ils font également la distinction entre la part rationnelle de la confiance et sa part plus émotionnelle.

McKnight et Chervany [17], tout en reconnaissant que les définitions de la confiance sont multiples, proposent une analyse typologique d'un grand nombre de ces définitions dans le but d'en identifier les composants communs et la manière dont ceux-ci s'articulent pour former un modèle plus général de la confiance ou de montrer de quelle manière les différents types de confiances proposés sont liés entre eux.

4 Les composants de la confiance

De cette analyse lexicale sont extraites seize caractéristiques que nous regroupons en quatre catégories. La compétence du *trustee*, sa fiabilité, sa proximité émotionnelle avec le *trustor* et son honnêteté. Cette catégorisation ne se veut pas représenter de manière exhaustive les définitions de la confiance que l'on peut trou-

ver à travers les disciplines qui ont tenté de la formaliser, si une telle chose est possible, mais propose de rassembler les caractéristiques communes à la majorité d'entre elles.
Ces caractéristiques sont des positions doxastiques que l'agent *trustor* possède sur l'agent *trustee*. Lorsqu'on parle de caractéristique ici il n'est donc pas question de mesurer ontologiquement la valeur de vérité d'un point de vue extérieur, mais bien de décrire la croyance que le *trustor* porte sur le *trustee*.

4.1 Compétence

La *compétence* représente la croyance que le *trustor* porte sur la capacité du *trustee* à accomplir une action, sur la base de son expertise, de son professionnalisme, son expérience, son intelligence, son pouvoir ou de son niveau d'éducation. Hendriks et al. [8] y ajoutent également la propension à la coopération.
L'évaluation de la compétence du *trustee* par le *trustor* peut se faire sur la base des interactions passées mais aussi sur un apriori inhérent à la nature de l'action sur laquelle se porte la confiance. On peut estimer qu'un médecin sera de facto plus compétent qu'un boulanger lorsque le sujet de l'action est de nature médicale, et ce sans avoir besoin de connaître la réelle compétence du médecin en question. Une action dont la complexité semblera triviale au *trustor* demandera un investissement épistémique moins important lui autorisant à accorder sa confiance plus facilement.

4.2 Fiabilité

La fiabilité est la croyance du *trustor* en sa capacité à prédire la résolution positive de l'action par le *trustee* sur la base de la perception de ses actions passées [17].

4.3 Bienveillance

La bienveillance est la croyance portant sur les motivation du *trustee*, ce que la tradition aristotélicienne nomme la "bonne volonté" [23]. Un agent bienveillant n'est pas opportuniste, il est motivé à agir dans l'intérêt du *trustor* et se soucie de ne pas lui causer de tort [17].
Elle comprend également l'évaluation de la moralité, de l'éthique et de la capacité d'attention du *trustor* [8] mais représente aussi la proximité émotionnelle ou sympathie qui peut exister entre les deux agents. En ce sens, elle peut être qualifiée de composante épistémologique externaliste de la confiance, faisant intervenir des processus mentaux pas nécessairement sous le contrôle introspectif du *trustor* [31].
C'est une composante émotionnelle qui traduit à la fois la motivation du *trustor* à projeter sur le *trustee* des caractéristiques morales mais également la composante de la confiance que l'on accorde à quelqu'un que l'on connaît intimement et à qui on fait confiance malgré l'éventuel manque d'éléments plus objectifs sur ses motivations.

4.4 Honnêteté

Bien qu'intrinsèquement liée à la fiabilité [17, 16], l'honnêteté est un concept à part entière. Elle désigne la croyance du *trustor* en la capacité du *trustee* à être de bonne foi, à dire la vérité et à tenir ses promesses [2].
Elle se distingue de la fiabilité dans le sens où celle-ci n'est pas directement liée à la croyance du *trustor* en les motivations du *trustee* mais à sa capacité à produire de manière constante les mêmes résultats face à des situations similaires.

4.5 Confiance thérapeutique

Nous avons vu que la confiance, pour se distinguer de l'espérance, se voit souvent ajouter la nécessité pour le *trustor* de s'exposer non pas à une simple déception mais à un sentiment de trahison [22].
Il existe cependant un comportement qui est souvent identifié comme faisant partie de la confiance tout en ne s'exposant pas aux mêmes risques de déception et qui est qualifié de *confiance thérapeutique* [18].
Carter [3] la définit simplement comme étant la confiance que l'on accorde à quelqu'un lorsqu'il n'y a pas de raison particulière de le faire, hormis la volonté d'établir une relation interpersonnelle pouvant provoquer des comportements de confiance.

Il utilise l'analogie suivante : supposons que nous devons nous absenter pour le weekend et confier notre maison à quelqu'un pour s'occuper de nos animaux de compagnie et arroser nos plantes. Nous avons le choix de demander à un ami à qui on a déjà confié la maison par le passé et dont on sait qu'il est à la fois honnête et fiable. Nous avons également le choix de confier la maison à un jeune membre de notre famille, qui n'a jamais eu ce genre de responsabilité, mais dont on n'a aucune raison particulière de se méfier.

Plusieurs raisons rationnelles pourraient motiver de faire le choix de la deuxième personne. Faire ce choix de manière rationnelle dans le but d'inciter le comportement qu'on attendrait d'une personne de confiance [10], la rendre digne de confiance en lui accordant préemptivement notre confiance est l'attitude qualifiée de confiance thérapeutique.

Une autre vision de la confiance thérapeutique est souvent proposée, c'est la confiance thérapeutique qualifiée d'*active* [3]. À la différence de la définition précédente qu'on qualifiera de *passive*, la confiance thérapeutique active voit le *trustor* faire abstraction de ses positions épistémiques quand à la fiabilité, l'honnêteté, la bienveillance ou la fiabilité du *trustee* pour lui accorder sa confiance.

Pettit [19] propose l'argument que dans le cas où le *trustor* accorde sa confiance à quelqu'un dans l'espoir que le fait de se savoir digne de la confiance de quelqu'un d'autre rendra le *trustee* réellement digne de confiance, il donne à la fois au *trustee* une raison d'être digne de confiance et donne au *trustor* une justification pour sa croyance en la valeur du *trustee*. Cette justification épistémique est une composante nécessaire de la confiance thérapeutique qui la différencie de la naïveté.

Dans les deux cas le *trustor*, en cas de déception de la part du *trustee*, ne sera pas exposé aux mêmes sentiments de trahison que dans le cas de la confiance classique, c'est pourquoi la confiance thérapeutique est souvent qualifiée d'*impure* [9] bien que considérée comme étant un composant de la confiance.

5 Notre définition de la confiance

La confiance est un concept qui est souvent confondu avec celui de crédibilité, coopération ou foi dans les études qui ne portent pas directement sur sa formalisation [1]. Dans le langage courant la confiance désigne aussi, souvent, le sentiment de sécurité ou l'optimisme que l'on peut ressentir envers une personne, ne se limitant pas à une action précise mais faisant intervenir un contexte plus général et comprenant une facette émotionnelle importante.

Nous définissons le fait pour un *trustor* d'accorder sa confiance à un *trustee* comme l'expression de sa volonté de prendre un risque basé sur la croyance que le *trustee* fera preuve d'un comportement maximisant les intérêts du *trustor* dans une situation précise.

Pour être distincte de l'espérance elle nécessite la présence d'un risque de trahison du *trustor* par le *trustee*. Ce risque doit pouvoir induire un sentiment plus fort que de la déception dans le cas où la confiance aurait été mal placée.

Elle doit être circonscrite à une situation donnée, c'est à dire donnée pour l'accomplissement d'une action précise et non accordée de manière générale à un agent. Cette situation est généralement de la forme : un agent *trustor* accorde sa confiance à un agent *trustee* pour accomplir une action définie.

Elle est composée d'un ensemble de croyances faisant intervenir l'opinion du *trustor* sur la compétence, la fiabilité, la bienveillance et l'honnêteté du *trustee*.

La confiance comprend également une dimension active, thérapeutique, ayant la possibilité de passer outre l'ensemble de croyances potentiellement négatives et permettant d'être octroyée par un acte volontaire et rationnel de l'agent *trustor*.

La confiance n'est pas un mécanisme de contrôle mais un substitut à ce dernier [25]. Elle ne peut réellement exister si le *trustee* est constamment sous le regard du *trustor* ou si (la confiance reposant sur un risque pour le *trustor*) le coût de la trahison par le *trustee* est trop peu élevé, comme dans le cas où des mécanismes

de pressions ou de punitions existeraient pour forcer le *trustee* à accomplir l'action [28].

5.1 Défiance

Bien que la défiance nécessite une absence de confiance, elle en est un concept distinct [31]. Tout comme la confiance n'est pas juste de l'espérance ou de la dépendance, la défiance n'est pas simplement de la crainte ou de la désespérance.

La défiance n'est pas la simple absence de confiance puisqu'il est possible de ne pas faire confiance tout en ne faisant pas non plus défiance [7]. Bien que les deux concepts ne représentent pas respectivement l'absence l'un de l'autre, ils sont malgré tout mutuellement exclusifs, un *trustor* ne peut pas en même temps faire confiance et défiance à un *trustee* sur la même action [30].

De même, être conscient pour un *trustee* d'être sujet de défiance de la part du *trustor* provoque un sentiment plus fort que simplement se savoir non sujet à la confiance, accentuant de par le fait la nécessité à distinguer les deux concepts [4]. Ne comportant pas l'élément thérapeutique que peut apporter la confiance volontaire, la défiance est plus facile à accorder que la confiance.

Tout comme pour la confiance, la défiance est décomposable en caractéristiques qui peuvent être associées à des positions doxastiques. En miroir de la confiance nous identifions donc la compétence, la fiabilité, la bienveillance et l'honnêteté, illustrant la croyance du *trustor* en la capacité du *trustee* à ne pas accomplir l'action. Les valeurs de vérité de ces caractéristiques sont les opposées des caractéristiques de la confiance accordée par le même *trustor* au même *trustee* sur la même action car elles sont liées et mutuellement exclusives. La croyance en la compétence du *trustee* pour accomplir l'action est celle qui entre aussi en jeu dans l'évaluation de sa capacité à ne pas l'accomplir, il en est de même pour les autres composantes.

De la même manière que pour la confiance, la défiance comporte un élément conscient et volontaire [7] capable de se substituer aux positions doxastiques et qui traduit une volonté de faire défiance. La nature de cette volonté peut être multiple, pouvant découler d'une inimitié personnelle ou être le fruit de stéréotypes sociaux négatifs envers le *trustee* [21].

5.2 Méfiance

La méfiance est définie comme l'absence de confiance qui serait assortie ou non d'une absence de défiance. À ce titre elle n'est pas un concept séparé de la confiance mais simplement son absence.

C'est une position d'agnosticisme quant à la valeur de confiance que l'on doit accorder à un agent et est la position par défaut lorsque le *trustor* n'a aucune information sur le *trustee* [30]. Elle est généralement la conséquence d'une absence de confiance thérapeutique ou volontaire.

Elle recouvre les deux usages qui sont fait du mot dans le langage courant qui est défini à la fois comme l'absence de confiance et comme la suspicion négative envers quelqu'un.

6 La confiance comme modalité

Nous avons proposé une définition de la confiance se voulant syncrétique des approches la définissant comme une relation épistémique entre deux agents, *trustor* et *trustee*, au regard d'une action dont l'accomplissement maximiserait les bénéfices pour le *trustor*, ce que Leturc et Bonnet [13] nomment la confiance *dispositionnelle*.

Nous exprimons la confiance comme un opérateur modal $T_\alpha\phi$ où α est le *trustor* et ϕ la proposition enrichie, sujet de la confiance de α. ϕ est définie ici comme un triplet $\langle\beta, \varphi, k\rangle$ où β est l'agent *trustee*, φ l'action allant être accomplie par β et k le contexte dans lequel cette action est censée s'accomplir.

$T_\alpha\phi$ se comprend donc de la manière suivante : "α fait confiance à β pour accomplir l'action φ dans le contexte k". Le contexte k peut être une contrainte de lieu, de temps ou tout autre limitation situationnelle imposée par le *trustor* sur l'accomplissement de l'action φ par le *trustee*.

6.1 La distribution de la confiance

Si certains, comme Liau [15], considèrent que l'agent *trustor* peut faire preuve d'irrationalité et que dans la situation où $T_\alpha\phi$ et où T tient pour vrai que $\phi \Rightarrow \psi$ on ne peut en déduire $T_\alpha\psi$.
Comme nous ne considérons que des agents rationnels nous pouvons faire cette inférence et acceptons l'axiome de distribution (1) :

$$\vdash T_\alpha\phi \wedge T_\alpha(\phi \Rightarrow \psi) \Rightarrow T_\alpha\psi \qquad (1)$$

6.2 La confiance non contradictoire

Nous considérons que l'agent *trustor* est rationnel. À cet effet nous estimons qu'il ne peut croire deux propositions opposées. L'opposé de $T_\alpha\phi$ dans lequel ϕ serait le triplet $\langle\beta, \varphi, k\rangle$ est $T_\alpha\neg\phi$ où $\neg\phi$ serait le triplet $\langle\beta, \neg\varphi, k\rangle$ et nous acceptons l'axiome de non contradiction (2) :

$$\vdash T_\alpha\phi \Rightarrow \neg T_\alpha\neg\phi \qquad (2)$$

Notons que ne pas faire confiance à deux propositions opposées ne signifie pas ne pas faire confiance à deux propositions distinctes mais dont les actions seraient la contradiction l'une de l'autre. Ainsi $\vdash T_\alpha\phi \wedge T_\alpha\psi$ où ϕ serait le triplet $\langle\beta, \varphi, k\rangle$ et où ψ serait le triplet $\langle\gamma, \neg\varphi, k\rangle$ est une formule valide si β est différent de γ.

6.3 La confiance introspective

Nous considérons qu'un agent est conscient de la confiance qu'il accorde à un autre agent et par extension de la confiance qu'il peut s'accorder à lui même introspectivement. Accorder sa confiance est une action qui elle-même peut être l'objet de la confiance, $T_\alpha T_\alpha\phi$ est donc une formule valide qui peut être réécrite $T_\alpha\psi$ et se comprend de la manière suivante : "α fait confiance à α pour accorder sa confiance à β pour accomplir l'action φ dans le contexte k".
Cette introspection n'apportant pas d'information supplémentaire sur l'agent *trustee* β mais étant une confiance pouvant être qualifiée de second ordre (c'est à dire se portant sur l'ensemble de croyances que le *trustor* peut avoir sur lui-même) nous pouvons considérer qu'elle équivaut à faire confiance à la proposition augmentée initiale. Nous acceptons donc l'axiome de compacité (3) :

$$\vdash T_\alpha T_\alpha\phi \Rightarrow T_\alpha\phi \qquad (3)$$

7 La défiance comme modalité

Nous avons précédemment défini la méfiance comme un concept indépendant bien que lié à celui de la confiance. À ce titre nous pouvons déclarer une nouvelle modalité représentant la relation épistémique entre deux agents, *trustor* et *trustee*, au regard d'une action dont le non accomplissement minimiserait les bénéfices pour le *trustor*.
En miroir de la confiance, nous exprimons la défiance comme un opérateur modal $D_\alpha\phi$ où α est l'agent *trustor* et ϕ la proposition enrichie, sujet de la défiance de α.
De la même manière ϕ est définie comme un triplet $\langle\beta, \varphi, k\rangle$ où β est l'agent *trustee*, φ l'action allant ne pas être accomplie par β et k le contexte dans lequel cette action est censée ne pas s'accomplir.
$D_\alpha\phi$ se comprend donc de la manière suivante : "α pense que β ne va pas accomplir l'action φ dans le contexte k".
La défiance n'est pas l'opposé de la confiance et $\neg T_\alpha\phi$ est une condition nécessaire pour $D_\alpha\phi$ mais n'en est pas une condition suffisante. Si de premier abord $T_\alpha\neg\phi$ peut également passer pour de la défiance, elle n'en est pas, faisant intervenir un sentiment de déception différent en cas de non accomplissement de l'action. Dans le langage courant $\neg T_\alpha\phi$ traduirait le sens de la phrase "je ne fais pas confiance à Bob pour réussir à faire cela" tandis que $T_\alpha\neg\phi$ pourrait être "j'ai confiance en Bob... Il va échouer à faire ceci". Si les deux situations utilisent le même mot "confiance" en français (ce qui n'est pas forcément le cas dans toutes les langues, l'anglais par exemple disposant des mots *distrust* et *mistrust* pour désigner les concepts antagonistes de la confiance) ces deux phrases peuvent décrire deux situations différentes, la seconde portant une espérance en l'échec de Bob supérieure à la première.

7.1 La distribution de la défiance

Nous considérons ici également *trustor* (que nous pouvons également nommer, par analogie, *distrustor*) comme un agent rationnel. A ce titre nous pouvons accepter sans difficulté l'axiome (4).

$$\vdash D_\alpha\phi \wedge D_\alpha(\phi \Rightarrow \psi) \Rightarrow D_\alpha\psi \qquad (4)$$

7.2 La défiance non contradictoire

Nous pouvons accepter également que deux propositions opposées ne peuvent cohabiter dans l'esprit d'un agent rationnel et proposer l'axiome (5).

$$\vdash D_\alpha\phi \Rightarrow \neg D_\alpha\neg\phi \qquad (5)$$

7.3 La défiance introspective

Le comportement de l'introspection est cependant différent dans le cas de la défiance. Un agent trustor rationnel qui tiendrait $D_\alpha D_\alpha\phi$ pour vrai serait en fait en train d'exprimer sa croyance en l'absence de défiance envers ϕ. De la même manière, accepter $D_\alpha T_\alpha\phi$ n'exprimerait que l'absence de croyance en sa confiance en ϕ. Dans le cadre de l'introspection nous ne pouvons donc qu'ajouter des clauses de confiance à sa propre défiance, grâce à l'axiome (6) :

$$\vdash D_\alpha\phi \Rightarrow T_\alpha D_\alpha\phi \qquad (6)$$

8 Le cas de la méfiance

Nous avons vu que la méfiance était définie comme l'absence de confiance assortie ou non d'une absence de défiance et à ce titre être simplement exprimée comme $\neg T_\alpha\phi$ se comprenant de la manière suivante : "α se méfie de β pour accomplir l'action φ dans le contexte k".

9 Représentation multi-modale des composantes de la confiance

Nous avons vu que la confiance pouvait être réduite en un petit nombre de positions doxastiques. La confiance qu'un agent *trustee* accomplira une action dans un contexte particulier, au bénéfice d'un agent *trustor* est liée à l'évaluation par ce dernier de la compétence, la fiabilité, la bienveillance et l'honnêteté du *trustee*. Nous avons vu également que la confiance possédait une part volontaire, bien que rationnelle, pouvant outrepasser les autres composantes.

Ces caractéristiques peuvent être exprimées sous forme de modalités de la manière suivante :

$$\vdash T_\alpha\phi \Rightarrow [C_\alpha\phi \wedge (H_\alpha\phi \vee B_\alpha\phi) \wedge R_\alpha\phi] \vee V_\alpha^+\phi \qquad (7)$$

Où $C_\alpha\phi$ représente la croyance que α porte sur la compétence de *trustee* à accomplir l'action dans un contexte donné, $H_\alpha\phi$ la croyance en son honnêteté, $B_\alpha\phi$ la croyance en sa bienveillance et $R_\alpha\phi$ en sa fiabilité. $V_\alpha^+\phi$ est la part active qui peut positivement influencer la confiance portée en la proposition ϕ.

Comme nous l'avons vu, les composantes de la confiance ne sont pas totalement indépendantes les unes des autres et une forte corrélation existe entre l'honnêteté et la bienveillance [17, 8]. Cette corrélation se traduit par la disjonction entre les deux modalités qui permet d'exprimer la plus faible influence que pourrait avoir l'absence de l'une sur la présence de l'autre.

9.1 Composantes de la défiance

De la même manière que pour la confiance, nous pouvons représenter les composantes de la défiance sous la forme de plusieurs modalités, comme suit :

$$\vdash D_\alpha\phi \Rightarrow [\neg C_\alpha\phi \vee \neg H_\alpha\phi \vee \neg B_\alpha\phi \vee \neg R_\alpha\phi] \vee V_\alpha^-\phi \qquad (8)$$

Les composantes $C_\alpha\phi$, $H_\alpha\phi$, $B_\alpha\phi$ et $R_\alpha\phi$ sont les mêmes qui entrent en jeu dans la représentation des composantes de la confiance (7). En effet, il est raisonnable de penser que puisque quand $T_\alpha\phi$ est vrai, $D_\alpha\phi$ est faux et réciproquement alors les composantes de compétence, d'honnêteté, de bienveillance et de fiabilité qui influent positivement sur la confiance sont les mêmes positions doxastiques qui influent négativement sur la défiance pour un même triplet $\langle\beta, \varphi, k\rangle$.

9.2 Rationalité de la confiance

Puisque nous représentons la confiance et la défiance impliquant des agents rationnels ceux-ci ne peuvent faire défiance en une proposition s'ils font confiance à cette même proposition. Soit les propriétés suivantes :

$$\vdash T_a\phi \Rightarrow \neg D_a\phi$$

$$\vdash D_a\phi \Rightarrow \neg T_a\phi$$

Propriétés desquelles nous pouvons déduire, en introduisant les composantes respectives de chaque modalité :

$$\vdash [(\, C_\alpha\phi \,\wedge\, (\, H_\alpha\phi \,\vee\, B_\alpha\phi \,) \,\wedge\, R_\alpha\phi \,) \,\vee\, V^+_\alpha\phi]$$
$$\Rightarrow \neg[(\, \neg C_\alpha\phi \,\vee\, \neg H_\alpha\phi \vee \neg B_\alpha\phi \vee \neg R_\alpha\phi \,) \,\vee\, V^-_\alpha\phi]$$

10 Conclusion et perspectives

Nous avons dans ce papier proposé un formalisme multi-modal permettant de représenter une version minimale de la confiance et de la défiance que peut porter un agent sur l'accomplissement d'une action par un autre agent. Plusieurs de ces modalités pouvant être représentées sous la forme de positions doxastiques il pourrait être intéressant de se pencher sur les interactions qu'elles pourraient avoir avec les logiques couramment utilisées pour représenter la croyance et la connaissance telle que la logique doxastique KD45 ou la logique épistémique S5.

Dans ce papier nous avons abordé la confiance qu'un agent humain peut exprimer envers un autre agent de même nature, il serait intéressant de tenter de généraliser cette approche à la confiance que peut porter un agent humain à un agent intelligent pouvant prendre des décisions et agir en fonction de ces dernières que l'IA embarquée d'une voiture autonome ou la modération automatique d'un service web.

De la même manière il sera intéressant de décrire plus avant les propriétés des composantes de la confiance, en particulier en y apportant de la gradualité.

Références

[1] Z. M. Aljazzaf, M. Perry, and M. A. Capretz. Online trust : Definition and principles. In *2010 Fifth Int. Multi-conference on Computing in the Global Information Technology*, pages 163–168. IEEE, 2010.

[2] P. Bromiley and L. Cummings. Transaction costs in organizations with trust. *Research on Negotiations in Organizations*, 5 :19–247, 1995.

[3] J. A. Carter. Therapeutic trust. *Philosophical Psychology*, pages 1–24, 2022.

[4] J. Domenicucci and R. Holton. Trust as a two-place relation. *The philosophy of trust*, pages 149–160, 2017.

[5] K. M. Engelke, V. Hase, and F. Wintterlin. On measuring trust and distrust in journalism : Reflection of the status quo and suggestions for the road ahead. *Journal of Trust Research*, 9(1) :66–86, 2019.

[6] A. M. Evans and J. I. Krueger. The psychology (and economics) of trust. *Social and Personality Psychology Compass*, 3(6) :1003–1017, 2009.

[7] K. Hawley. Trust, distrust and commitment. *Noûs*, 48(1) :1–20, 2014.

[8] F. Hendriks, D. Kienhues, and R. Bromme. Measuring laypeople's trust in experts in a digital age : The muenster epistemic trustworthiness inventory (meti). *PloS one*, 10(10).

[9] P. Hieronymi. The reasons of trust. *Australasian Journal of Philosophy*, 86(2) :213–236, 2008.

[10] H. J. N. Horsburgh. The ethics of trust. *The Philosophical Quarterly*, 10(41) :343–354, 1960.

[11] D. Karmis and F. Rocher. *Trust, distrust, and mistrust in multinational democracies : Comparative perspectives*.

[12] J. Kugler and P. J. Zak. Trust, cooperation, and conflict : Neuropolitics and international relations. In *Advancing interdisciplinary approaches to international relations*, pages 83–114. Springer, 2017.

[13] C. Leturc and G. Bonnet. A normal modal logic for trust in the sincerity. In *AAMAS*, pages 175–183, 2018.

[14] J. D. Lewis and A. J. Weigert. Social atomism, holism, and trust. *The sociological quarterly*, 26(4) :455–471, 1985.

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

[16] R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3) :709–734, 1995.

[17] D. H. McKnight and N. L. Chervany. What is trust ? a conceptual analysis and an interdisciplinary model. 2000.

[18] P. J. Nickel. Trust and obligation-ascription. *Ethical theory and moral practice*, 10(3) :309–319, 2007.

[19] P. Pettit. The cunning of trust. *Philosophy & Public Affairs*, 24(3) :202–225, 1995.

[20] D. Platt, S. A. Haque, N. S. Wong, W. Leibzon, X. Xu, R. Minxha, and L. Liu. A simple agent-based model of the development of trust in hierarchical organizations.

[21] N. N. Potter. Interpersonal trust. In *The Routledge handbook of trust and philosophy*, pages 243–255. Routledge, 2020.

[22] C. Pursiainen and T. Forsberg. Relations of trust and mistrust. In *The Psychology of Foreign Policy*, pages 299–336. Springer, 2021.

[23] C. Rapp. Aristotle's rhetoric. *The Stanford Encyclopedia of Philosophy*, 2002.

[24] J. K. Rempel, J. G. Holmes, and M. P. Zanna. Trust in close relationships. *Journal of personality and social psychology*, 49(1) :95, 1985.

[25] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer. Not so different after all : A cross-discipline view of trust. *Academy of management review*, 23(3) :393–404, 1998.

[26] J. Rutter. From the sociology of trust towards a sociology of 'e-trust'. *Int. journal of new product development & innovation management*, 2(4) :371–385, 2001.

[27] O. Schilke, M. Reimann, and K. S. Cook. Trust in social relations. *Annual Review of Sociology*, 47 :239–259, 2021.

[28] S. B. Sitkin and N. L. Roth. Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization science*, 4(3) :367–392, 1993.

[29] M. Tschannen-Moran and W. K. Hoy. A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of educational research*, 70(4) :547–593, 2000.

[30] E. Ullmann-Margalit. *Trust, Distrust, and In Between.* Russell Sage Found., 2004.

[31] E. N. Zalta, U. Nodelman, C. Allen, and J. Perry. Stanford encyclopedia of philosophy, 1995.