Dynamique des relations de confiance dans une équipe d'agents virtuels

L. Callebert[a] lucile.callebert@hds.utc.fr
D. Lourdeaux[a] domitile.lourdeaux@hds.utc.fr
JP. Barthès[a] barthes@utc.fr

[a]Sorbonne universités, Université de Technologie de Compiègne,
CNRS, Heudiasyc UMR 7253

Résumé

De nombreuses études en psychologie sociale montrent que les relations de confiance interindividuelles influencent les dynamiques de groupe et la performance d'une équipe. Pour l'émergence de comportements collectifs cohérents dans un groupe d'agents, nous proposons un modèle d'agent cognitif dont le processus décisionnel individuel prend en compte ses relations de confiance, dont la modélisation et la dynamique est inspirée du modèle proposé par [Mayer et al., 1995].

Mots-clés : *système multi-agents, psychologie-sociale, modèle cognitif de la confiance*

Abstract

Studies in social psychology showed that trust relationships among people influence both group dynamics and team performance. For coherent collective behaviors to emerge among a group of agents, we propose a cognitive agent model whose decisional process takes into account the trust relationships. Trust relationships and their dynamics are defined following the model of organizational trust proposed by [Mayer et al., 1995].

Keywords: *multi-agents system, social psychology, cognitive model of trust*

1 Introduction

Les environnements sociotechniques sont aujourd'hui de plus en plus complexes, et le recours à la simulation peut être un bon outil de formation pour les personnes qui s'y trouvent confrontées. Dans le cadre de formation technique à la gestion de situations complexes ou critiques, des facteurs humains entrent en compte : les personnes sont amenées à faire face à des situations stressantes, à gérer leurs relations aux autres, etc. En effet, dans ces environnements complexes, les personnes ne travaillent pas seules mais sont bien souvent interdépendantes et le pouvoir de décision est partagé. Les compétences mises en jeu ne sont alors plus seulement techniques mais également sociales, et incluent des capacités spécifiquement liées au travail en équipe telles que le leadership ou la gestion des communications. L'un des paramètres clés influençant la mise en oeuvre de ces capacités est celui de la confiance : un climat de confiance dans une équipe permet notamment une répartition plus souple des tâches, un échange plus libre d'informations lors des communications, un meilleur investissement de chacun des membres de l'équipe [Jones and George, 1998]. Dans ce cadre de formation à des compétences non techniques, nous cherchons à modéliser le processus de prise de décision d'agents virtuels évoluant dans des environnements sociotechniques complexes en prenant en compte ce facteur de la confiance.

Pour peupler environnements ces evironnement sociotechniques complexes d'agents dont les comportements rendent compte de l'activité humaine telle qu'observée en environnement réel, nous proposons dans cet article une formalisation du modèle et de la dynamique de la confiance de [Mayer et al., 1995]. Nous expliquons ensuite comment la confiance est utilisée dans le processus décisionnel individuel des agents pour l'émergence d'une activité collective. Enfin nous présentons un exemple simple de la dynamique de la confiance avant de conclure.

2 Travaux connexes

2.1 Confiance et ACA

En informatique affective, la notion de confiance a principalement été abordée dans le domaine des agents conversationnels animés (ACA), en référence à la relation de confiance qu'un ACA doit établir avec l'utilisateur. Partant du principe qu'établir une relation de confiance agent-utilisateur va améliorer la qualité de l'interaction et l'appréciation de l'agent

par l'utilisateur, les travaux de recherche sont centrés sur les facteurs qui favorisent cette confiance. Par exemple [Antos et al., 2011] et [de Melo et al., 2013] montrent qu'un utilisateur fait plus confiance et s'engage plus facilement dans une relation de coopération avec un agent exprimant des émotions qu'avec un agent n'en exprimant pas.

Dans l'optique de l'amélioration de la relation agent-utilisateur,[Bickmore and Cassell, 2001], [Bickmore and Cassell, 2005] ont développé un modèle de dialogue social visant créer une relation de confiance avec l'utilisateur. Se basant sur des travaux en psychologie sociale, ils identifient plusieurs dimensions importantes dans une relation liant deux personnes : la *solidarité*, qui désigne le degré de similarité entre les dispositions comportementales de deux personnes ; la *familiarité*, qui désigne le degré d'échange d'information ; et la textitdimension affective, qui désigne le degré d'appréciation. Ces dimensions sont utilisées pour la définition de stratégies permettant à l'agent d'améliorer sa relation avec l'utilisateur. A l'issue de l'interaction avec l'agent, les sujets remplissent un questionnaire pour estimer le niveau de confiance qu'ils ont en l'agent : ce niveau de confiance reportée est utilisé comme critère d'évaluation. Dans ces travaux, aucune opérationnalisation de la relation de confiance n'est faite.

[Sansonnet and Bouchet, 2011], dans leurs travaux sur les ACA dotés de comportements rationnels et psychologiques, se servent du concept de la confiance : chaque agent a une confiance en soit, qui est dynamique et représente l'assurance cognitive de l'agent. Pour modéliser la relation à l'autre de l'agent, les auteurs se servent d'un vecteur à trois dimensions : la $dominance_{i,j}$, qui représente la puissance de i par rapport à j, $cooperation_i$ qui correspond à la tendance de i à se montrer coopératif et $trust_{i,j}$ qui caractérise la confiance de i et j. Toutes ces valeurs sont mises à jour au cours de l'interaction avec l'utilisateur, et sont utilisées pour influencer le processus de décision rationnelle de l'agent. Dans ces travaux comme dans les précédents, la confiance fait référence à un concept global se rapportant au climat qui règne entre l'utilisateur et l'agent. Cette notion vague englobe plusieurs questions : l'utilisateur pense-t-il que les informations données par l'agents sont fiables ? L'utilisateur pense-t-il que l'agent va agir dans son intérêt, a l'intention de l'aider ? Ce manque de différenciation limite les possibilités de mise en place de stratégies visant à l'amélioration de la confiance.

2.2 Confiance et négociation

La notion de confiance a également été étudiée dans les travaux sur la négociation : l'hypothèse est communément faite que le succès de la négociation dépend de la capacité du négociateur à mettre l'autre parti en confiance. Dans leurs travaux pour l'entrainement à la négociation, [Traum et al., 2005] modélisent la confiance selon les axes suivants : la *familiarité*, la *solidarité*, et la textitcrédibilité, représentant le degré de partage des croyances. La confiance est opérationnalisée comme une combinaison de ces trois dimensions sous la forme d'une valeur appartenant à l'intervalle $[0; 1]$ mise à jour à chaque interaction, et fait référence à la situation globale. La confiance est ensuite utilisée par l'agent virtuel pour décider de croire ou non les informations données par son interlocuteur et pour décider d'une stratégie de négociation. Dans des travaux ultérieurs, [Traum et al., 2008a], [Traum et al., 2008b] étendent leur modèle pour permettre une négociation multipartis : la confiance n'est plus globale mais devient spécifique à un agent.

2.3 Confiance et simulation sociale

Des travaux se sont également intéressés à la confiance dans le domaine de la simulation sociale : la relation de confiance est utilisée pour modéliser des agents aux comportements plus crédibles. La confiance est opérationnalisée par [Marsella et al., 2004] : une relation de confiance (ou méfiance) lie chaque paire d'agents. Cette relation est mise à jour à chaque interaction entre les agents et sert en partie à déterminer la valeur accordée par un agent aux messages reçus de la part d'un autre. D'une façon similaire, [Silverman et al., 2011] utilisent la confiance comme l'une des composantes des relations entre agents. Elle est opérationnalisée comme la *familiarité*, influencée par les actions des agents et utilisée pour déterminer la quantité d'information à donner lors d'un échange entre agents. Dans ces modèles, le seul 'type' de confiance utilisé fait référence à la fiabilité des informations échangées. Un autre manque peut être évoqué : aucune distinction n'est faite en fonction du contexte du dialogue. Or dans un environnement complexe, les domaines de compétences mis en jeu sont divers et les compé-

tences de chaque agent sont limitées à certains domaines. L'information donnée par un agent n'a donc pas la même valeur selon son niveau de compétence dans le domaine considéré.

Cette notion de contextualisation de la confiance a été étudiée par [Castelfranchi and Falcone, 1998], dont les travaux s'inscrivent à la fois dans les domaines des sciences cognitives et de l'informatique. [Castelfranchi and Falcone, 2001] proposent un modèle computationnel de la confiance pour l'étude des phénomènes sociaux. Un agent cognitif i attribue une valeur de confiance à j pour la réalisation d'un but identifié g en considérant principalement les dimensions suivantes :
– la capacité de j à réaliser g.
– la nécessité pour i que g soit réalisé par j.
– l'engagement de j par rapport à g.
– la motivation de j à aider i pour réaliser g.
La notion centrale de ces travaux est la délégation : un niveau de confiance suffisant de i à j permet à l'agent i de déléguer, de manière explicite ou implicite, une partie de l'activité à l'agent j. De cette notion de délégation émerge le concept d'activité collective. Dans cette approche, la confiance est entièrement définie par rapport à un but identifié et est donc contextualisée. La relation de confiance liée à l'appréciation de j pour i est également considérée ici : de part le paramètre de motivation. Cependant cette motivation est dirigée uniquement vers i ; or dans un contexte de travail collectif, la motivation de j peut être dirigée vers son équipe de travail et pas forcément spécifiquement vers l'un des membres de l'équipe, avec qui il peut par ailleurs ne pas entretenir d'affinités particulières.

Nous proposons une formalisation et une dynamique de la confiance basées sur le modèle de [Mayer et al., 1995] prenant en compte la double contextualisation de la confiance : par rapport à une personne particulière et dans un contexte identifié.

3 Modèle de la confiance

Le modèle de la confiance proposé par [Mayer et al., 1995] est issu d'études en psychologie sociale et organisationnelle. Les auteurs ont étudié différents modèles pour en construire un s'appliquant à tous les niveaux organisationnels et dont les dimensions englobent celles proposées dans la littérature. La relation de confiance de A à B est alors caractérisée par des éléments propres à celui qui fait confiance, mais aussi à sa représentation de l'autre. D'une part la *disposition* de A à faire confiance en règle générale correspond à la part irrationnelle de la confiance et est fortement liée à la personnalité de A ; elle va être déterminante dans le cas où A ne dispose pas d'informations sur B. D'autre part les éléments liés à la représentation de l'autre sont les suivants :
– la *bienveillance* : A doit croire que B veut son bien et ne va pas agir uniquement dans son propre intérêt. La bienveillance est liée à une relation particulière d'attachement entre A et B et est proche de la notion d'affinité ou d'appréciation.
– l'*intégrité* : A et B doivent partager des valeurs communes, et A doit penser que B va agir en accord avec ces valeurs. Dans le monde du travail, il est préférable qu'un employé partage les valeurs de son entreprise.
– les *capacités*. Cette notion est centrale puisqu'elle permet d'une part de découpler la notion de confiance de celle d'affinité (*e.g.* A peut apprécier B sans pour autant lui faire confiance) et d'autre part de contextualiser la confiance (*e.g.* A peut faire confiance à B pour faire la cuisine mais pas pour faire le ménage).
La confiance constitute une relation asymétrique entre deux personnes et est contextualisée : elle n'existe que par rapport à une personne identifiée sur une tâche particulière. Le modèle de [Mayer et al., 1995] est un modèle cognitif de la confiance : une relation de confiance implique des croyances issues d'une évaluation cognitive de la situation. La figure 1 illustre le modèle proposé par [Mayer et al., 1995].

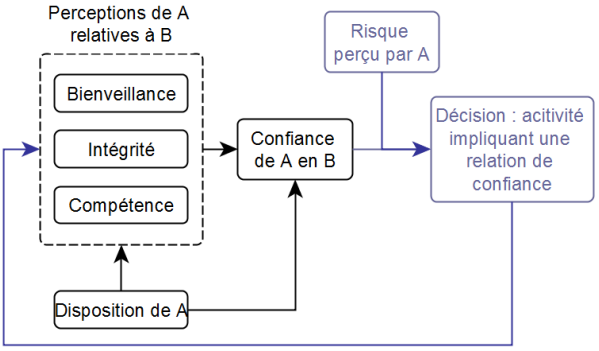

FIGURE 1 – Modèle de la confiance proposé par [Mayer et al., 1995]

Suivant le modèle de [Mayer et al., 1995], nous proposons de modéliser cette confiance $trust$ d'un agent i à un agent j pour une tâche k à un instant t comme une combinaison des éléments

de disposition, intégrité, bienveillance et capacité :

$$trust_{i,j,k} = \Psi(d_i, v_{b_{i,j}} b_{i,j}, v_{i_{i,j}} i_{i,j}, v_{c_{i,j,k}} c_{i,j,k})$$
$$\text{avec } c_{i,j,k} = \left\langle c_{1_{i,j}}, c_{2_{i,j}}, ..., c_{n_{i,j}} \right\rangle$$

où

- $d_i(t) \in D : [-1; 1]$ représente la disposition de i à faire confiance. Une valeur positive réprésente quelqu'un de confiant par nature tandis qu'une valeur négative caractérise une personne de naturel méfiant.
- $b_{i,j} \in B : [-1; 1]$ représente la confiance de i en la bienveillance de j à son égard. Une valeur positive indique que i pense que j lui veut du bien alors qu'une valeur négative indique que i croit que j veut lui nuire. La valeur 0 représente une relation neutre.
- $i_{i,j} \in I : [-1; 1]$ représente la confiance de i en l'intégrité de j. De même que pour la bienveillance, une valeur positive (*resp.* négative) indique des intentions positives (*resp.* négatives) de la part de j à l'égard de l'équipe.
- $c_{i,j,k}(t) \in C : [0; 1]$ représente la confiance de i en les capacités de j sur l'ensemble des compétences nécessaires à la tâche k. Une valeur de 1 indique que j maîtrise parfaitement les compétences alors que 0 indique qu'il ne les maitrise pas du tout.
- $v_X \in [0; 1]$ représente une valeur de croyance associée à chacune des dimensions liée à la représentation de l'autre (*i.e.* $b_{i,j}$, $i_{i,j}$ et $c_{i,j,k}$). En effet, ces valeurs représentent une croyance : l'agent A *pense* qu'il peut avoir confiance en la bienveillance de B. L'agent A peut avoir un degré de certitude par rapport à ce qu'il pense plus ou moins élevé, par exemple selon la source de l'information : l'information peut être issue d'une observation directe ou être rapportée verbalement. v_X est égale à 1 si l'agent est certain, alors que v_X vaut 0 si l'agent ne sait pas du tout.

4 Dynamique de la confiance

La confiance entre deux personnes est évolutive : les niveaux de confiance entre deux personnes déterminent si elles s'engagent dans une collaboration impliquant une relation de confiance, et l'évaluation des événements et interactions liés à cette collaboration vont réflexivement modifier les niveaux de confiance entre ces personnes. La figure 2 illutre ce processus. Par ailleurs, d'après [Mayer et al., 1995], [Karsenty, 2010], en l'absence d'évenements dans l'environnement, la confiance reste stable. Il n'y a donc pas de dynamique temporelle interne de la confiance. La dynamique de la confiance est dont liée aux événements qui se produisent dans l'environnement.

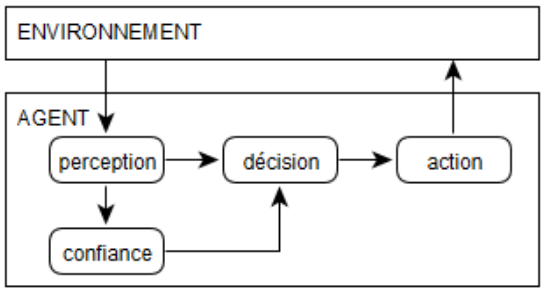

FIGURE 2 – Cycle d'un agent

La disposition à la confiance étant liée à la personnalité et donc stable, les composantes de la confiance qui sont dynamiques sont celles liées à la représentation de l'autre.

4.1 Dynamique des capacités

Les croyances de l'agent A sur les capacités de B évoluent dynamiquement lorsque A obtient, de manière directe ou indirecte, des informations relatives à un comportement de B :

- A obtient des informations de manière directe sur un comportement de B lorsqu'il observe B effectuer ce comportement dans l'environnement. A oeut alors évaluer la qualité d'exécution de la tâche en fonction du résultat et/ou du temps d'exécution. La réussite (*resp.* l'échec) d'une tâche k par j entraîne une augmentation (*resp.* une diminution) de la confiance de i en les compétences nécessaires à la réalisation de k. A est alors *sûr* de la compétence de B : $v_{c_{i,j,k}}$ est égale à 1.
- A obtient des informations de manière indirecte sur un comportement de B lors d'un dialogue avec un autre agent (pas obligatoirement B). La valeur accordée à l'information est alors pondérée par la confiance que A a en la source de l'information : $v_{c_{i,j,k}} \sim trust_{i,source}$.

C étant l'ensemble de définition des capacités, nous proposons la fonction :

$\mu : C \to C$
$\mu(c_{i,j,k}, comportement_j)$: représente la variation de la confiance de i en les compétences de j nécessaires à la réalisation de la tâche k.

4.2 Dynamique de la bienveillance et l'intégrité

Dans une équipe d'agents ayant des intentions communes, l'agent A va modifier ses croyances sur la bienveillance et l'intégrité de B en fonction du comportement de B. Le comportement de B est évalué par A en référence aux intentions propres de A et aux intentions communes à l'équipe. Le calcul de la confiance étant un processus cognitif, nous nous basons sur les théories de l'évaluation cognitive ordinnairement utilisées pour le calcul des émotions. L'évaluation d'un événement se fait selon plusieurs dimensions dont nous retenons la congruence motivationnelle et l'attribution de la responsabilité. Un comportement de B ayant un impact positif sur l'équipe (*resp.* sur A) sera évalué positivement par A, ce qui augmente la confiance de A en la l'intégité (*resp.* la bienveillance) de B. Cette évaluation est à modérer en fonction de l'attribution de la responsabilité à B dans le comportement qu'il exhibe : B avait-il conscience des conséquences de son comportement ? B était-il libre ou contraint de son comportement ?

B étant l'ensemble de définition de la bienveillance, I celui de l'intégrité et INT celui des intentions, nous proposons les fonctions suivantes :

$\Phi 1 : B \times INT \rightarrow B$
$\Phi 1(b_{i,j}, intentions_i, comportement_j)$: représente la variation de la confiance de i en la bienveillance de j liée à la perception et l'évaluation cognitive par i d'un comportement de j.

$\Phi 2 : I \times INT \rightarrow I$
$\Phi 2(i_{i,j}, intentions_{team}, comportement_j)$: représente la variation de la confiance de i en l'intégrité de j liée à la perception et l'évaluation cognitive par i d'un comportement de j.

De la même manière que pour les capacités, les valeurs de croyance associées à la confiance de A en la bienveillance et l'intégrité de B est mise à jour selon la manière dont l'information est obtenue (*e.g.* par observation directe ou rapportée aucours d'un dialogue) et selon la source de l'information.

5 Prise de décision

Nous avons présenté dans la partie précédente la dynamique de la confiance : des processus cognitifs permettent à l'agent d'évaluer la situation et de mettre à jour ses niveaux de confiance à chaque interaction. Les niveaux de confiance sont ensuite utilisés par l'agent pour prendre une décision, comme le montre la figure 2.

5.1 Prise de décision et plans partagés

Nous nous intéressons aux équipes d'agents évoluant dans des domaines complexes : les agents sont amenés à collaborer et il est nécessaire que chaque agent prenne en compte les autres dans sa décision individuelle. Lorsque les rôles de chacun des agents sont attribués *à priori* et les procédures de travail pré établies, comme dans [Rickel and Johnson, 1999], la prise en compte des autres est limitée : certaines des actions doivent être synchronisées et chaque agent doit savoir qui joue quel rôle, mais aucune décision n'est à prendre quant à la construction d'un plan commun ou aux tâches dont chacun est responsable. D'autres travaux se sont intéressés à la construction de tels plans partagés permettant de supporter une activité collective : [Pollack, 1986] et [Grosz and Sidner, 1988] ont été parmis les premiers à formaliser cette notion dans le contexte d'un dialogue agent-utilisateur. [Sidner, 1994] a introduit le concept de *négociation collaborative* pour désigner le processus par lequel deux agents A et B négocient la construction d'un plan partagé. [Chu-Carroll and Carberry, 1994], [Chu-Carroll and Carberry, 2000] introduisent la notion de *préférence* par rapport à une action pour la construction d'un plan partagé agent-utilisateur : l'agent prend en compte les préférences de l'utilisateur avant de lui faire des propositions de plans.

Tous ces travaux se placent dans le contexte d'une collaboration entre agents, mais deux principales limites sont à relever :
– les agents sont par définition collaboratifs et n'ont pas la possibilité de refuser la collaboration. Or en réalité les comportements humains ne sont pas toujours motivés par de bonnes intentions et la collaboration est rarement la seule option ;
– la notion de compétence n'est pas prise en compte : les agents sont conosidérés comme compétents par défaut sur l'ensemble des tâches composant le plan partagé.

5.2 La confiance pour la construction de plans partagés

[Mayer et al., 1995] distinguent dans leur modèle la confiance en elle-même et l'engagement

dans une activité (collaborative ou non) impliquant une relation de confiance. En effet, il peut y avoir collaboration sans confiance, par exemple lorsque la collaboration est contrainte. L'inverse est vrai également : une relation de confiance n'implique pas forcément que les agents s'engagent dans une activité collaborative : par exemple s'ils n'en ont pas l'utilité, ou parce que les risques liés à la collaboration sont trop élevés. Par ailleurs, une fois que les agents ont décidé de collaborer, la confiance intervient à chaque étape de construction d'un plan partagé : il faut prendre en compte les compétences de chacun avant de proposer un plan et une répartition des tâches.

L'engagement dans une activité collaborative et la construction d'un plan partagé nécessite donc, pour l'agent, de considérer :

– d'une part ses niveaux de confiance en les agents concernés. Le contexte de la collaboration intervient en premier lieu à ce moment puisque la confiance en les capacités est contextuelle et dépend des compétences mises en jeu ;
– d'autre part une évaluation des risques liés en particulier à cette collaboration. Les risques sont indépendants de la confiance et liés à la situation : la prise en compte du contexte est encore une fois indispensable.

C'est la combinaison de ces deux paramètres qui va permettre à l'agent de prendre sa décision. D'après [Mayer et al., 1995], une relation de confiance implique l'*envie* de prendre le risque de collaborer avec la personne, et l'évaluation des risques permet d'engager un comportement de confiance (*i.e. trusting behavior*, ici une activité collaborative) en toute connaissance de cause. L'évaluation des risques distingue les comportements de 'confiance aveugle' de ceux de 'confiance consciente' (distinction entre *confidance* et *trust* dans [Mayer et al., 1995]).

6 Illustration : collaboration pour aménager un bureau

Pour illustrer la dynamique des relations sociales, nous avons déroulé un exemple sur un scénario simple impliquant deux agents : A et B faisant partie d'une nouvelle équipe de travail et ne se connaissant pas veulent aménager le bureau qu'ils partagent. A et B vont donc collaborer pour mener à bien cette tâche, et établir des relations de confiance. Dans cet exemple, nous nous plaçons du point de vue de l'agent A.

La colonne de droite du tableau 1 décrit les événements qui se produisent dans l'environnement. La colonne de gauche présente les évolutions des niveaux de confiance de A en B suite à l'évaluation de la situation par A. A étant de nature confiante, $disposition_A > 0$. A $t0$, l'agent A ne connait pas l'agent B et n'a donc pas d'informations sur sa bienveillance, son intégrité ou ses compétences. La décision de A de se lancer dans une activité collaborative avec B est donc basée sur $disposition_A$. A $t1$, A et collaborent pour monter le bureau ($T1$). A met à jour ses niveaux de confiance : B participe à l'activité donc B est compétent pour $T1$, et $competence_{A,B,T1}$ augmente. B n'est pas contraint de réaliser $T1$ et permet d'atteindre un but de l'équipe : les intentions des B sont compatibles avec celles de l'équipe. L'intégrité de B est donc évaluée psitivement par A : $integrite_{A,B}$ augmente. L'action de B n'est âs pertinente du point de vue des intentions personnelles de A, A n'a donc toujours pas d'information sur $bienveillance_{A,B}$. A $t2$, B casse un pied du bureau : B n'était donc pas compétent pour $T1$ et $competence_{A,B,T1}$ chute. A étant de nature confiante évalue cependant l'action de B comme involontaire. $integrite_{A,B}$ reste stable. A ne dispose toujours pas d'information sur $bienveillance_{A,B}$. A termine de monter le bureau seul à $t3$ et n'a pas d'interaction avec B : les niveaux de confiance restent stables en l'abscence d'interaction entre les agents. A $t4$ B prépare un café pour A. B est donc compétent sur $T2$: $competence_{A,B,T2}$. De nature confiante, A interprète cette événement comme une bonne intention de à son égard (A aime le café donc action compatible avec les intentions de A) : $bienveillance_{A,B}$ augemente. Cette action n'est pas pertinente du point de vue de l'équipe, $integrite_{A,B}$ reste stable.

FIGURE 3 – Exemple de dynamique de la confiance

La figure 3 présente les courbes d'évolution des niveaux de confiance de A en B aucours de ce scénario. Les décisions futures de A seront prises par rapport à ces niveaux de confiance. Par soucis de simplicité, ce scénario à été conçu

TABLE 1 – Exemple de scénario illustrant la dynamique de la confiance

Situation initiale	**Confiance initiale de A à B**
A et B ne se connaissent pas mais savent qu'ils ont un but commun. A est de nature confiante.	$disposition_A > 0$ Pas d'information sur $competence_{A,B,x}$? $integrite_{A,B}$? $bienveillance_{A,B}$?
Evénement	**Evaluation par l'agent A : impact sur la relation de confiance entre A et B**
A et B collaborent pour monter le bureau (T1)	$competence_{A,B,T1} \nearrow$: B est compétent. $integrite_{A,B} \nearrow$ et $bienveillance_{A,B}$? : Congruence motivationnelle : réalisation d'un but commun. Responsabilité : B aide l'équipe et B n'est pas contraint.
B casse un pied du bureau	$competence_{A,B,T1} \searrow$: Action T1 échouée : B n'est pas compétent. $integrite_{A,B} stable$ et $bienveillance_{A,B}$? : Congruence motivationnelle : contre la réalisation d'un but commun. Responsabilité : action involontaire.
A monte le bureau seul	$competence_{A,B,T1}$, $integrite_{A,B}$ et $bienveillance_{A,B} stables$: Pas d'interactions entre A et B.
B prépare un café pour A	$competence_{A,B,T2} \nearrow$: Action réussie, B est compétent. $integrite_{A,B} stable$ et $bienveillance_{A,B} \nearrow$: Congruence motivationnelle : A aime le café. Responsabilité : B aide A et n'est pas contraint.

pour présenter une situation à deux agents où les événements sont observés par A directement dans l'environnement : nous n'avons pas représenté ici la valeur de certitude que l'agent A accordre à chacun de ces niveaux de confiance. L'agent A est sûr que l'agent B est compétent pour prépaper le café puisqu'il le voit faire. Cependant, dans le cas d'un travail en équipe, beaucoup d'informations sont échangées entre les agents par le dialogue. Dans ce cas la valeur de certitude prend toute son importance : si B *propose* à A de lui préparer un café, A *pense* que B est compétent pour préparer un café, mais n'en est pas *sûr* temps que B n'a pas effectivement préparer le café. L'importance accordée par A à $competenceA, B, T2$ dans sa prise de décision (*e.g.* A accepte-t-il la proposition de B ?) sera moindre.

Cet exemple simple illustre également l'importance de la contextualisation : la confiance de A en B ne fait pas référence à une valeur globale pouvant s'appliquer de la même manière dans toutes les situations. A distingue son niveau de confiance en B selon les compétences mises en jeu ; et l'échec de B sur $T1$ n'empêchera pas A de collaborer à nouveau avec B sur des tâches mettant en jeu d'autres compétences. Cette distinction est extrêment importante dans un environnement de travail complexe où les compétence mises en jeu sont diverses et variées. Par ailleurs, un niveau de confiance faible sur une compétence précise n'est pas forcément un point négatif pour la collaboration : c'est une source d'information qui permet d'adapter les comportements (*e.g.* A peut surveiller l'agent sur les tâches dont il a la charge mais pour lesquelles ses compétences sont faibles) et ainsi limiter les erreur [Karsenty, 2010], [Amalberti, 2001].

7 Conclusion

Pour la modélisation d'agents sociaux évoluant dans des domaines complexes et supportant une activité collaborative, nous proposons de prendre en compte le facteur de la confiance. Nous avons présenté dans cet article un modèle des relations de confiance et de leur dynamique. Les relations de confiance entre un agent A et un agent B sont constituées de trois dimensions : la confiance de A en la bienveillance de B, en l'intégrité de B et en les compétences de B. La prise en compte du contexte, indispensable dans le cadre des domaines complexes, est faite à deux niveaux : la confiance est une relation spécifiquement dirigée d'un agent à un autre, et est re-

lative à une compétence identifée.

La relation de confiance évolue au fur et à mesure des observations par i des comportements des autres agents : la dynamique de la confiance est basée sur l'évaluation par i des comportements de j. La confiance est ensuite utilisée par les agents pour prendre une décision relative à une activité collaborative : tout d'abord la décision de s'engager dans une collaboration est basée sur la confiance, puis la confiance est également utilisée pour établir un plan partagé.

Références

[Amalberti, 2001] Amalberti, R. (2001). La maÃ®trise des situations dynamiques. *Psychologie franÃ§aise*, 46(2) :107–118.

[Antos et al., 2011] Antos, D., De Melo, C., Gratch, J., and Grosz, B. J. (2011). The Influence of Emotion Expression on Perceptions of Trustworthiness in Negotiation. In *AAAI*.

[Bickmore and Cassell, 2001] Bickmore, T. and Cassell, J. (2001). Relational agents : a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 396–403. ACM.

[Bickmore and Cassell, 2005] Bickmore, T. and Cassell, J. (2005). Social Dialongue with Embodied Conversational Agents. In *Advances in natural multimodal dialogue systems*, pages 23–54. Springer.

[Castelfranchi and Falcone, 1998] Castelfranchi, C. and Falcone, R. (1998). Principles of trust for MAS : Cognitive anatomy, social importance, and quantification. In *Multi Agent Systems, 1998. Proceedings. International Conference on*, pages 72–79. IEEE.

[Castelfranchi and Falcone, 2001] Castelfranchi, C. and Falcone, R. (2001). Social Trust : A Cognitive Aproach. *J. Pitt. London : Wiley*.

[Chu-Carroll and Carberry, 1994] Chu-Carroll, J. and Carberry, S. (1994). A plan-based model for response generation in collaborative task-oriented dialogues. *arXiv preprint cmp-lg/9405011*.

[Chu-Carroll and Carberry, 2000] Chu-Carroll, J. and Carberry, S. (2000). Conflict resolution in collaborative planning dialogs. *International Journal of Human-Computer Studies*, 53(6) :969–1015.

[de Melo et al., 2013] de Melo, C., Carnevale, P., and Gratch, J. (2013). People's biased decisions to trust and cooperate with agents that express emotions. In *Proc. AAMAS*.

[Grosz and Sidner, 1988] Grosz, B. J. and Sidner, C. L. (1988). Plans for discourse. Technical Report BBN-6728, BBN LABS INC CAMBRIDGE MA.

[Jones and George, 1998] Jones, G. R. and George, J. M. (1998). The experience and evolution of trust : Implications for cooperation and teamwork. *Academy of management review*, 23(3) :531–546.

[Karsenty, 2010] Karsenty, L. (2010). Comment faire confiance dans les situations Ã risque ?

[Marsella et al., 2004] Marsella, S. C., Pynadath, D. V., and Read, S. J. (2004). PsychSim : Agent-based modeling of social interactions and influence. In *Proceedings of the international conference on cognitive modeling*, volume 36, pages 243–248. Citeseer.

[Mayer et al., 1995] Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3) :709.

[Pollack, 1986] Pollack, M. E. (1986). A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics.

[Rickel and Johnson, 1999] Rickel, J. and Johnson, W. L. (1999). Virtual humans for team training in virtual reality. In *Proceedings of the ninth international conference on artificial intelligence in education*, volume 578, page 585. Citeseer.

[Sansonnet and Bouchet, 2011] Sansonnet, J.-P. and Bouchet, F. (2011). Personnification d'entitÃ©s par Agents Conversationnels.

[Sidner, 1994] Sidner, C. L. (1994). An artificial discourse language for collaborative negotiation. In *AAAI*, volume 94, pages 814–819.

[Silverman et al., 2011] Silverman, B. G., Pietrocola, D., Nye, B., Weyer, N., Osin, O., Johnson, D., and Weaver, R. (2011). Rich socio-cognitive agents for immersive training environments : case of NonKin Village. *Autonomous Agents and Multi-Agent Systems*, 24(2) :312–343.

[Traum et al., 2008a] Traum, D., Marsella, S. C., Gratch, J., Lee, J., and Hartholt, A. (2008a). Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *Intelligent Virtual Agents*, pages 117–130. Springer.

[Traum et al., 2008b] Traum, D., Swartout, W., Gratch, J., and Marsella, S. (2008b). A virtual human dialogue model for non-team interaction. In *Recent trends in discourse and dialogue*, pages 45–67. Springer.

[Traum et al., 2005] Traum, D., Swartout, W., Marsella, S. C., and Gratch, J. (2005). Fight, flight, or negotiate : Believable strategies for conversing under crisis. pages 52–54. Springer Berlin Heidelberg.