



**HAL**  
open science

## Les limites du BDI pour rendre compte du comportement humain en situation de crise

Maël Arnaud, Carole Adam, Julie Dugdale

### ► To cite this version:

Maël Arnaud, Carole Adam, Julie Dugdale. Les limites du BDI pour rendre compte du comportement humain en situation de crise. Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA 2017), Jul 2017, Caen, France. hal-01561962

**HAL Id: hal-01561962**

**<https://hal.science/hal-01561962>**

Submitted on 2 Aug 2017

**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.

# Les limites du BDI pour rendre compte du comportement humain en situation de crise

Maël Arnaud<sup>\*1</sup>, Carole Adam<sup>†1</sup>, and Julie Dugdale<sup>‡1, 2</sup>

<sup>1</sup>Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France

<sup>2</sup>University of Agder, Norway

## Résumé

La modélisation et la simulation à base d'agents peuvent offrir un puissant environnement de test pour des scénarios de gestion de crises, qu'elles soient provoquées par la nature ou par l'homme. L'architecture BDI (Beliefs, Desires, Intentions) a été identifiée comme prometteuse pour la création d'agents au comportement humain. Cependant, cette architecture possède aussi des limites, notamment en ce qui concerne la modélisation du comportement plus ou moins irrationnel d'individus en situation de crise. Dans cet article, nous identifions un certain nombre de critiques théoriques et de limitations pratiques de l'architecture BDI dans le contexte de la gestion de crise.

## 1 Introduction

Nos sociétés expérimentent une recrudescence des catastrophes d'origines naturelles (*e.g.* tsunamis) et humaines (*e.g.* attaques terroristes, accident industriels) [19]. L'incapacité à contrôler l'émergence de ces situations s'est transformée en une volonté de les gérer, avec l'idée d'en minimiser les dégâts. Un objectif important est donc l'amélioration des techniques de préparation de la population et de gestion des crises, dont l'une des conditions essentielles est l'accès à des instruments de test.

La modélisation et simulation à base d'agents (MSBA) est une technique venant de l'intelligence artificielle (IA), qui permet, entre autres, la modélisation du comportement humain et sa simulation dans un environnement dont les paramètres peuvent être modifiés. Un système multi-agents (SMA) est créé en modélisant plusieurs agents dans un même environnement. Cette technique permet la création d'interactions entre agents et donc la modélisation de populations. En créant un environnement capable de générer des catastrophes et en le faisant habiter par des agents imitant un comportement humain, il est possible de tester et de comparer différents scénarios sans avoir besoin d'attendre

l'émergence d'un véritable contexte de crise. Par exemple, l'utilisation des systèmes multi-agents pour l'amélioration de la précision des plans d'évacuation est déjà répandue [16, 27] et est un champ actif de la recherche.

La modélisation des agents dans un SMA fait l'objet de questionnements sur le niveau de complexité des agents. Par exemple, on pourra noter la concurrence entre le principe KISS (Keep It Simple, Stupid), introduit par [5], et le principe KIDS (Keep It Descriptive, Stupid), défendu par [17]. Le problème spécifique du niveau de complexité à adopter, lorsqu'il est question de modéliser, le comportement humain a déjà été discuté par le passé [1]. Les auteurs ont analysé plusieurs architectures disponibles pour la modélisation d'agents au comportement humain. Ils concluent que l'architecture BDI (Beliefs, Desires, Intentions) offre des avantages en termes d'adaptabilité, de robustesse, et de programmation abstraite, ce qui la qualifie comme solution intéressante au problème.

L'architecture BDI possède de sérieux atouts comparée à d'autres architectures (automates à états finis, agents réactifs) [1, 3]. Elle n'est cependant pas exempte de critiques et de détracteurs. Ainsi, Georgeff et al. [21] ont identifié en particulier deux manques de l'architecture, l'incapacité à apprendre et à avoir un comportement social. Ils y apportent ensuite chacun leurs réponses, par exemple Pollack affirme que l'architecture BDI, si elle possède ces manques, n'empêche pas la création d'extensions pour y remédier.

Dans cet article, nous souhaitons présenter certaines des critiques faites à l'architecture BDI et au modèle dont elle découle, et mettre en lumière des manques de l'architecture quant à la simulation de comportements humains en situation de crise. Notre hypothèse est que l'architecture possède des fondements intéressants, mais qu'elle profiterait grandement d'un apport des sciences sociales, au delà de la psychologie naïve sur laquelle elle est bâtie.

La section 2 introduit l'architecture BDI et la philosophie sur laquelle elle est basée, ainsi que quelques extensions existantes. La Section 3 expose ensuite certaines des critiques théoriques qui peuvent être adressées

\*. mael.arnaud@univ-grenoble-alpes.fr

†. carole.adam@univ-grenoble-alpes.fr

‡. julie.dugdale@univ-grenoble-alpes.fr

à cette architecture. La Section 4 analyse des témoignages de victimes d'incendies pour en extraire des limitations pratiques de l'architecture BDI pour la modélisation de comportement en situations de crise. La Section 5 conclut l'article.

## 2 Modèle et Architecture BDI

L'architecture BDI et le modèle BDI sont deux entités bien distinctes, avec parfois peu de rapport entre elles [14, 8]. L'architecture computationnelle BDI, que nous appellerons a-BDI, est une architecture faite pour créer des agents dits "intelligents" (*i.e.* une entité autonome, qui agit en fonction de son environnement et en accord avec ses buts propres). L'a-BDI a été construite à partir du travail du philosophe Michael Bratman, auquel il est important de se référer pour la formalisation des éléments de l'architecture. Ce travail propose un modèle de l'action humaine, que nous nommerons m-BDI et que nous allons commencer par expliquer.

### 2.1 Le modèle BDI (m-BDI)

Le m-BDI s'appuie sur une théorie philosophique de la raison pratique. Cette théorie ne s'intéresse à la raison que dans la mesure où elle est appliquée aux actions.

**La rationalité instrumentale.** Le paradigme de la rationalité instrumentale, auquel souscrit Bratman au travers du m-BDI, postule que la raison n'existe que pour trouver et agencer les actions de manière à ce que l'humain atteigne ses buts. Lorsqu'elle est vue comme un instrument, la raison pratique possède deux fonctions distinctes : 1) la délibération et 2) le raisonnement des moyens et des fins [23]. La délibération est l'activité de décider quel but accomplir, quel désir satisfaire. Le raisonnement des moyens et des fins quant à lui consiste à définir de quelle manière accomplir le but choisi. La rationalité instrumentale a fortement influencé Bratman, et les deux activités qui la composent sont présentes dans le m-BDI.

**L'apport de Bratman.** Avant Bratman, un modèle croyances-désirs existait. Bratman y a apporté la notion d'Intentions. Dans le m-BDI, la délibération est l'activité qui permet de choisir quels désirs un individu va réaliser, les faisant passer au statut d'intention. Lorsqu'il adopte une intention, l'individu commence à planifier son action. C'est là qu'intervient le raisonnement des moyens et des fins. Ce processus aboutit à un plan d'action, dirigé vers le futur, lié à une intention. Ce plan peut être composé d'actions, de plans, d'intentions et d'incertitudes [11, 12].

### 2.2 L'architecture BDI (a-BDI)

L'a-BDI reprend les trois catégories d'états mentaux (B, D et I) identifiées dans le m-BDI comme composantes de l'action humaine. Ces trois états mentaux sont expliqués ci-dessous et illustrés dans la figure 1.

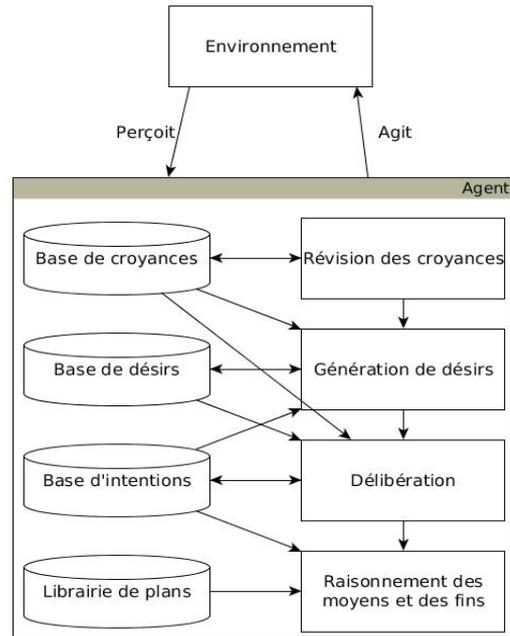


FIG. 1 – Schéma de l'architecture BDI

**Les croyances.** Elles sont stockées dans une "base de connaissances". Elles représentent le monde tel que l'agent se le représente. Les croyances peuvent venir des perceptions de l'agent ou de règles d'inférence propres à l'agent. Les croyances peuvent être hétéroclites, la même base de croyance pouvant contenir par exemple une croyance sur un état interne et externe (*e.g.* la route N45 est bloquée; je me sens nauséeux) ou une certitude et une croyance (*e.g.* je suis sûr que la route N45 est bloquée; je pense que le chien du voisin va m'attaquer).

**Les désirs.** La communauté liée à l'a-BDI s'y réfère souvent comme à des buts ou des choix. Ils correspondent aux états du monde qu'un agent trouve souhaitables. Les désirs d'un agent sont stockés dans une "base de désirs". Ils n'ont pas besoin d'être consistants entre eux. Les désirs peuvent être intrinsèque à l'agent ou générés par des règles internes qui impliquent d'autres états mentaux (*e.g.* croire que j'ai faim implique que je désire manger).

**Les intentions.** Elles sont un sous-ensemble des désirs, que l'agent a sélectionnés pour les réaliser. Contrairement aux désirs, les intentions doivent être consistantes entre elles et avec les croyances de l'agent. Les intentions ne sont pas faciles à adopter et ne peuvent pas être facilement abandonnées. Ainsi un agent peut abandonner une intention si et seulement si : 1) il pense qu'elle est impossible à réaliser, ou 2) il pense que son résultat a déjà été accompli. Une règle supplémentaire vient souvent s'ajouter, permettant à l'agent d'abandonner son intention si l'opportunité de satisfaire un désir avec

une plus grande priorité apparaît.

**Les plans.** Ils sont des séquences de micro et de macro actions. Une micro-action est une action élémentaire (*e.g.* ouvrir une porte), alors qu'une macro-action est une action composée de plusieurs autres (micro ou macro) actions (*e.g.* aller à la cafétéria). Par ailleurs il n'est pas rare dans l'a-BDI (et pour cause, c'est une caractéristique fondamentale du m-BDI) de trouver des macro-actions comportant plusieurs places d'actions "vides". Ces actions seront spécifiées en fonction des états mentaux de l'agent lors de la sélection (ou de la révision) du plan.

### 2.3 Extensions existantes

L'a-BDI est utilisée aujourd'hui depuis plus de quinze ans pour la modélisation de comportements humains. Aussi, de nombreuses limites ont été identifiées [24] et des solutions ont été apportées. Nous présentons ici certaines de ces limites et solutions.

**Les émotions.** L'impact des émotions sur la prise de décision a été démontré [30] mais n'est pourtant pas pris en charge par l'a-BDI standard. L'expérience de la salade de fruit montre que la tendance à considérer les informations cognitives (*e.g.* la salade de fruit contient moins de sucre) et à faire un choix en conséquence est positivement corrélée avec le niveau de disponibilité des ressources cognitives. Inversement, plus la disponibilité des ressources cognitives est faible, plus l'individu aura tendance à faire un choix basé sur des informations affectives (*e.g.* j'aime le gâteau au chocolat) [29]. L'intégration des émotions à l'a-BDI comme "un mécanisme de contrôle des moyens d'un agent" [28] est donc importante pour la modélisation du comportement humain, spécialement en situation de crise où l'individu n'a pas accès à toutes ses ressources cognitives. Des formalisations des émotions pour l'architecture BDI ont été réalisées et devraient donc être intégrées à cet effet [2, 31].

**Les normes et obligations.** Les normes sociales ont été abordées du point de vue des désirs et de la motivation dans l'a-BDI. [14] expliquent que les désirs sont la "Cendrillon de la trinité BDI" et que leur formalisation logique laisse à désirer... En conséquence, les notions de désir et de but sont souvent confondues. Pour pallier à ce problème, les auteurs proposent de voir le désir non comme un but, mais comme une motivation à atteindre un but. Ils proposent ensuite que d'autres types de motivations puissent exister. Parmi elles, les normes et les obligations. La différence fondamentale entre les désirs et les normes, selon les auteurs, est leur provenance. Les premiers viennent de l'agent lui-même, alors que les seconds viennent de l'extérieur, des relations de l'agent avec d'autres agents et de leurs relations. Les auteurs proposent donc une formalisation logique des nouveaux désirs, normes et buts afin

de créer une architecture plus cohérente et permettant de mieux coller, selon eux, au m-BDI.

## 3 Critiques théoriques

Certaines des limites de l'a-BDI pour la modélisation du comportement humain ont été résolues, mais d'autres demeurent. Nous pensons notamment que le m-BDI, s'il possède une base philosophique, ignore largement l'apport des sciences sociales à la compréhension de l'action humaine et de ce qui la dirige. Les critiques adressées au BDI dans ce que nous présentons ensuite visent à surligner ce fait, sans toutefois prétendre à l'exhaustivité.

### 3.1 Des agents rationnels?

Le m-BDI part du principe que l'être humain est rationnellement limité (*i.e.* on considère que l'être humain est un être purement rationnel, mais possède des informations, des ressources cognitives et du temps en quantités limitées). Cependant, ce postulat, en plus d'ignorer toute une partie des découvertes en psychologie [20, 35], n'est habituellement pas correctement intégré à l'a-BDI [8]. Nous nous attacherons ici à la notion de rationalité et de rationalité limitée au sens large (*e.g.* l'individu possède peut-être les informations, mais les possède-t-il dans *toutes* les situations?) et aux critiques qui peuvent être faites au m-BDI et à l'a-BDI sur ces thèmes.

**Les jugements conscients.** L'individu selon le m-BDI, ou l'agent construit à partir l'a-BDI, est dirigé par ses intentions, construites à partir de ses désirs, croyances et d'autres intentions. Cependant, rien ne prend en compte les différents niveaux des attitudes mentales (*e.g.* conscient, implicite, inconscient). Par exemple, quelqu'un peut croire fermement ne pas être raciste et, dans le même temps, acquiescer un avis négatif sur une personne dès qu'elle la voit, basée sur sa couleur de peau [26]. Cette attitude, courante au demeurant, ne peut avoir lieu que si l'individu croit, quelque part, que les individus possédant une couleur de peau X sont "mauvais". Ce qui est parfaitement contradictoire avec la croyance de ne pas être raciste, et donc crée une incohérence logique si ces deux croyances sont placées sur le même plan. Cependant, ce type de croyances généralisatrices et sous-jacentes a sa place dans le processus d'apprentissage humain [6] et est donc à prendre en compte dans la modélisation de comportements humains.

**La dissonance cognitive.** Arrive lorsqu'un individu maintient deux états cognitifs (*i.e.* des états mentaux) incompatibles [18]. La théorie stipule qu'un individu fera un effort proportionnel à la dissonance expérimentée pour retrouver un état de consonance. Prenons l'exemple d'un individu qui a l'intention de se tenir à un régime alimentaire sans sucre et sans gras. Fatigué après sa journée, il passe devant une boutique

avec un dernier beignet disponible à moitié prix, il l’achète et commence à le manger. Cet état de fait implique qu’un individu puisse tenir deux états mentaux incompatibles (*i.e.* je ne mangerai rien de gras ni de sucré; je vais manger un aliment gras et sucré). Ce qui n’est pas possible dans l’a-BDI, qui requiert une congruence dans la base d’intentions. Il faut aussi que l’agent puisse revenir à un état de consonance cognitive, par : 1) la modification de l’une de ses deux intentions (arrêt du régime ou du repas en cours) ou 2) l’ajout d’états mentaux permettant de compenser la dissonance (croyance que l’on peut faire une entorse de temps en temps, intention d’aller courir pour perdre, etc). La modélisation de la dissonance cognitive, au vu de ce que permet actuellement l’a-BDI, paraît forcée et peu réaliste. Il faut que l’agent 1) décide d’abandonner l’une des deux intentions, 2) s’il s’agit de l’intention initiale (*i.e.* le régime), décide s’il abandonne aussi le désir sous-jacent, 3) s’il n’abandonne pas le désir, acquiert un désir ”compensateur” et, 4) une fois l’intention secondaire accomplie, récupère l’intention initiale depuis le désir. L’impossibilité de tenir deux intentions incompatibles oblige à faire un va-et-viens incertain entre les désirs et les intentions, qui ”joue” plus avec les règles du m-BDI qu’il ne les respecte. Ce qui, à notre sens, met en lumière l’une des grandes faiblesses du m-BDI.

### 3.2 Des agents sociaux ?

L’une des critiques couramment faites à l’a-BDI porte sur son incapacité à évoluer dans un contexte social (*e.g.* une simulation multi-agents). Dans cette partie, nous présentons le modèle de Weber comme piste de réflexion pour pallier ce problème, ainsi qu’une critique adressée à l’a-BDI, porteuse d’une solution pour sociabiliser les agents.

**Weber et l’action sociale.** Max Weber [34] a catégorisé l’action sociale et a proposé un modèle influent. La prise en compte de ce modèle et de ses descendants ajouterait une profondeur à l’a-BDI, absente du m-BDI. Premièrement, le modèle de Weber comporte deux types d’actions sociales non rationnelles : 1) l’action affective (dirigée par les émotions) et 2) l’action traditionnelle (dirigée par l’habitude) [34]. Deuxièmement, Weber distingue, en plus de la rationalité instrumentale, la rationalité de valeur. La première est pour lui une sorte de sous-spécification des buts afin d’atteindre un but final plus élevé (*e.g.* finir ses études, pour trouver du travail, pour avoir de l’argent, etc.), qui correspond à peu près à la rationalité instrumentale dont s’inspire Bratman. La seconde tend vers l’accomplissement d’un but qui se suffit à lui-même ou n’est pas atteignable directement (*e.g.* être juste, aller au paradis). Ce second type de rationalité répond à une notion de légitimité chez l’agent, ce qui entraîne des règles différentes de la rationalité instrumentale

(*e.g.* un individu va arrêter de poursuivre une intention s’il vient à croire qu’elle est impossible à réaliser, mais peut-être pas s’il vient à penser qu’elle est contraire à l’une de ses valeurs). L’ajout d’actions sociales basées sur l’affect, l’habitude et sur une rationalité de valeurs seraient donc un ajout conséquent au BDI, dans l’optique de modéliser des agents sociaux.

**Des agents (a)sociaux.** Pour citer une critique récente et plus véhémente de l’a-BDI, [15] expliquent que l’implémentation de l’architecture n’est pas entièrement fondée sur les principes de Bratman et que l’idée d’action humaine derrière l’a-BDI est simpliste. Les auteurs ont par ailleurs un point de vue peu répandu dans la communauté liée à l’a-BDI, quoique parfois soulevé [25] : ils défendent l’idée selon laquelle faire des ajouts à l’a-BDI ne peut être suffisant, dans la mesure où l’architecture diffère trop de la manière dont peut penser un être humain pour être efficace. Ils proposent donc une refonte totale de l’a-BDI, basée sur de nouveaux éléments cognitifs, beaucoup plus ”sociaux”. On retrouve notamment une discussion tournant autour de la notion de ”paysage social”, chère à plusieurs sociologues, dont Pierre Bourdieu [9], ainsi qu’un élément fondamental de l’action sociale de Weber : la notion de valeur.

## 4 Limitations pratiques

En plus de ces critiques théoriques, l’a-BDI a aussi des limites pratiques pour la modélisation de certaines situations. Nous voulons ici en exposer certaines à partir d’une étude de cas.

Le samedi 7 février 2009, au sud-est de l’Australie, une série dévastatrice d’incendies de forêt a fait 173 morts et des centaines de blessés. La politique en termes d’incendies était alors ”prepare, defend or leave early”, impliquant que les individus sont informés (parfois formés) sur la défense de leur propriété contre le feu ; les pompiers ne peuvent pas être présents sur chaque incendie. Nous nous sommes intéressés aux récits de cent personnes [33] rescapées ou proches de personnes décédées à propos de cette journée. Dans ces récits, certaines personnes racontent la manière dont elles se sont mises en danger. Dans cet article nous présentons des exemples concrets à partir de 3 de ces entretiens, qui montrent certains comportements difficiles à capturer avec le m-BDI et remettent en question une partie de l’a-BDI.

### 4.1 Engagement dans les intentions

*Jim Baruta vit avec sa femme et ses deux enfants dans une maison sur une grande propriété que lui et sa femme possèdent à la campagne. Sa femme et ses enfants n’étaient pas à la maison le 7 Février. Jim avait décidé de défendre sa maison contre le feu, s’il devait y en avoir ce jour-là. Cependant, il a dû se rendre en ville pour*

son travail et n'a pas pu en partir avant 15h. Un incendie faisait alors rage près de sa propriété et le long de la route la reliant à son lieu de travail. Jim Baruta savait que "être sur les routes [lors d'un incendie] est la dernière chose à faire". Il avait toutefois l'intention de défendre sa propriété et sa maison. Il a donc outrepassé les avertissements de sa femme, les indices visuels sur la route indiquant qu'il était en danger, et ses connaissances préalables, afin de réaliser cette intention. Il survivra de justesse à cette expérience, sans parvenir à sauver sa maison.

Ce récit pose le problème, bien connu, de l'**engagement** qu'un agent doit avoir dans ses **intentions**. Cohen et Levesque ont étudié ce problème en se posant la question suivante : sous quelles conditions un agent peut-il et doit-il abandonner son intention pour être rationnel ? [13]. Ils présentent des règles d'engagement envers une intention qu'ils jugent trop rigides, nommée "intention fanatique" (*i.e.* une intention que l'agent ne peut abandonner que s'il la pense déjà réalisée ou irréalisable). Par exemple, si je suis en train d'éplucher ma banane et qu'on m'annonce que ma fille a eue un accident, il serait étrange que je termine d'éplucher sereinement ma banane avant d'aller voir comment va ma fille. L'idée des auteurs est donc de rendre une intention dans l'a-BDI plus facile à abandonner, afin d'obtenir un comportement général plus équilibré en terme de rationalité (*i.e.* ce qui est important pour l'agent doit avoir une plus grande priorité dans ses actions). Or, pour modéliser le comportement de Jim Baruta lors de cette journée, il faudrait non pas rendre l'intention plus facile, mais plus difficile à abandonner. Nous avons besoin non seulement des règles d'abandon de l'intention fanatique, mais avons en plus besoin d'un comportement inhibiteur de la perception de la situation. L'agent doit être tellement pris par son intention, qu'il est incapable d'analyser correctement sa situation et de la mettre en résonance avec ses connaissances préalables. Cette incapacité doit l'empêcher, malgré toutes les informations à sa disposition, de former les croyances que son intention est impossible à réaliser et que ses actions vont à l'encontre d'une intention plus importante que celle de défendre sa maison : son intention de rester en vie.

## 4.2 Mise à jour des croyances

*John Benett possède une maison à la campagne. Samedi 7 Février, il s'était préparé et avait préparé sa maison à faire face à un incendie. Il était parfaitement confiant en ses préparations et a passé une grande partie de la journée à essayer d'anticiper la direction du feu. John possédait la croyance qu'un feu ne pouvait, en aucun cas, venir d'une autre direction que celle du vent. Il a donc essayé de récupérer des informations concer-*

*nant un incendie provenant de cette direction. Dans l'après-midi, deux événements auraient pu lui fournir l'information que le feu viendrait de l'ouest : 1) un appel d'un ami lui disant de se préparer car le feu arrivait sur lui depuis le sud-ouest, et 2) des conducteurs exhibant un comportement de fuite paniquée à l'extérieur de sa propriété. John Benett a rejeté ces deux indices sur la base de sa perception du vent et de sa croyance que le feu ne pourrait pas venir de l'ouest ou du sud-ouest en présence d'un vent du nord. Lorsqu'un mur de feu apparut au sud-ouest de sa propriété, il eut du mal à y croire mais fut tout de même dans l'obligation de mettre à jour ses croyances.*

Ce témoignage pose donc le problème **de la source et de la teneur des informations**. Les informations pertinentes peuvent être ignorées sur la base de croyances préalables. Différentes règles, selon lesquelles une croyance peut ou non prendre la place d'une autre, ne sont pas incluses dans l'a-BDI. Ainsi des biais cognitifs peuvent en particulier influencer cette mise à jour des croyances [4]. La mise à jour purement rationnelle des croyances paraît normale pour la création d'agents artificiels, mais est insuffisante pour la modélisation du comportement humain.

## 4.3 Changement de contexte

*Alice Barber vit seule à la campagne. Sa maison se situe au fond d'un cul de sac. Sa protection contre le feu est un mélange entre une bonne préparation, rendue insuffisante par sa passion pour le jardinage, qui l'empêche d'enlever les végétaux entourant sa maison, et un plan d'action insuffisant, augmenté par ses compétences personnelles en matière de prévention d'incendie. Alice Barber possède initialement l'intention de quitter sa maison et de s'enfuir pour se mettre à l'abri dans des lieux qu'elle a préalablement repérés. Cette idée est au coeur de son plan d'évacuation. Cependant, lorsque son frère l'appelle, Alice lui fait part de son intention de rester pour combattre le feu et défendre sa maison. Peu de temps après, après avoir vu sa voisine s'enfuir, elle décide à nouveau de s'en aller, ce qu'elle finira par faire.*

Ce témoignage pose, en apparence, le problème inverse de celui de Jim Baruta. L'engagement envers ses intentions paraît faible et pourrait être modélisé grâce à une a-BDI possédant des règles pour faire varier le niveau d'engagement envers les intentions. Cependant une seconde lecture est possible selon laquelle Alice Barber n'a pas, lorsque son frère l'a appelée, abandonné l'intention de s'enfuir, mais s'est inscrite dans un **contexte différent**. Elle est passée du contexte de la femme seule, préparant sa survie face à un grand incendie, à celui de la soeur dans une situation offrant

une opportunité de faire la preuve de son courage. Prise dans ce second contexte, elle a formé l'intention de défendre sa maison, s'appuyant sur des croyances et des désirs déconnectés du premier contexte. Mais les éléments composant le contexte s'effacent lorsqu'elle raccroche le téléphone et, lorsqu'elle voit partir sa voisine, Alice Barber se réinscrit dans le premier contexte, avec ses croyances, désirs et intentions propres. La situation est alors critique et il faut agir vite, ce qui ne lui permet pas de déceler l'incohérence entre les croyances, désirs et intentions des deux contextes.

## 5 Conclusion

Nous avons dans cet article expliqué pour quelles raisons l'architecture BDI est intéressante afin de répondre au problème de la modélisation d'agents devant posséder un comportement humain. Nous avons expliqué l'architecture et donné quelques clés quant à ses racines philosophiques, qui font une grande partie de sa force. Cependant, nous avons montré que le modèle BDI fait fi de l'apport des sciences sociales à la compréhension de l'action humaine et que, par conséquent, l'architecture BDI qui en découle possède des difficultés à relever le défi qui nous occupe, malgré les efforts déjà déployés pour l'étendre.

D'un point de vue pratique, il existe cependant des outils pour développer des modèles de comportement humain à base d'agents BDI, enrichis de divers aspects psychologiques. Ainsi la plateforme GAMA [22] est une plateforme pour la modélisation et la simulation à base d'agents. Cette plateforme dispose d'une architecture BDI pour les agents [32] mais aussi d'émotions [10] et de liens sociaux [7].

Le travail futur, à notre sens, qu'il s'inscrive dans la création d'un nouveau modèle pour la création d'agents au comportement humain ou dans l'extension de l'architecture BDI, se doit de prêter une plus grande attention aux sciences sociales, notamment à l'étape de formalisation logique.

## Références

- [1] C. Adam and B. Gaudou. BDI agents in social simulations: a survey. *Knowledge Engineering Review*, 31(2002):1–27, 2016.
- [2] C. Adam, A. Herzig, and D. Longin. A logical formalization of the OCC theory of emotions. *Synthese*, 168(2):201–248, 2009.
- [3] C. Adam, P. Taillandier, and J. Dugdale. Comparing agent architectures in social simulation: Bdi agents versus finite-state machines. In *Hawaii International Conference on System Sciences (HICSS-50)*, 2017.
- [4] M. Arnaud, C. Adam, and J. Dugdale. The role of cognitive biases in reactions to bushfires. In *ISCRAM*, Albi, May 2017.
- [5] R. Axelrod. *The complexity of cooperation: agent-based models of conflict and cooperation*. Princeton University Press, 1997.
- [6] M. R. Banaji, R. Bhaskar, D. L. Schacter, and E. Scarry. Implicit stereotypes and memory: The bounded rationality of social beliefs. *Memory, brain, and belief.*, page 139, 2000.
- [7] J. Bangate, J. Dugdale, C. Adam, and E. Beck. A review on the influence of social attachment on human mobility during crises. In *ISCRAM*, Albi, May 2017.
- [8] G. Boella. Decision theoretic planning and the bounded rationality of BDI agents. Technical report, AAAI, 2002.
- [9] P. Bourdieu. *Raisons pratiques: sur la théorie de l'action*. Points (Éditions du Seuil): Essais. Seuil, 1994.
- [10] M. Bourgeois, P. Taillandier, and L. Vercouter. An Agent Architecture Coupling Cognition and Emotions for Simulation of Complex Systems. In *SSC*, 2016.
- [11] M. Bratman. Two Faces of Intention. *The Philosophical Review*, 3, 1984.
- [12] M. Bratman. *Intention, plans, and practical reason*, chapter Introduction. Harvard University Press, Cambridge, MA, 1987.
- [13] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, 1990.
- [14] F. Dignum, D. Kinny, and L. Sonenberg. From desires, obligations and norms to goals. *Cognitive Science Quarterly*, 2(3-4):405–427, 2002.
- [15] F. Dignum, R. Prada, and G. J. Hofstede. From Autistic to Social Agents. *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), Paris, France, May 5-9, 2014*, pages 1161–1164, 2014.
- [16] J. Dugdale, N. Bellamine-Ben Saoud, B. Pavard, and N. Pallamin. Simulation and emergency management. In *Information Systems for Emergency Management*, Advances in Management Information Systems. Sharp, 2010.
- [17] B. Edmonds and S. Moss. From kiss to kids - an 'anti-simplistic' modelling approach. In *Multi-Agent Based Simulations Conference*, pages 130–144, 2004.
- [18] L. Festinger. *A Theory of Cognitive Dissonance*. Stanford University Press, June 1957.
- [19] C. for Research on the Epidemiology of Disasters (CRED). The international disaster database. <http://www.emdat.be/database>. Accessed: 2017-03-28.
- [20] S. Freud. *The Unconscious*. Penguin Modern Classics. Penguin Books Limited, 2005.
- [21] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge. The belief-desire-intention model of agency. *Intelligent Agents V: Agents Theories, Architectures, and Languages. 5th International Workshop, ATAL'98.*, pages 1–10, 1998.
- [22] A. Grignard, P. Taillandier, B. Gaudou, D. A. Vo, N. Q. Huynh, and A. Drogoul. Gama 1.6: Advancing the art of complex agent-based modeling and simulation. In B. T. R. S. F. D. M. K. P. G. Boella, E. Elkind, editor, *Principles and Practice of Multi-Agent Systems (PRIMA)*, volume 8291 of *Lecture Notes in Computer Science (LNCS)*, pages 117–131. Springer, Dunedin, New Zealand, 2013.

- [23] N. Kolodny and J. Brunero. Instrumental rationality. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.
- [24] E. Norling. Folk psychology for human modelling: Extending the bdi paradigm. In *Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2004.
- [25] E. Norling. *Modelling human behaviour with BDI agents*. PhD thesis, RMIT Melbourne, 2009.
- [26] S. A. Ottaway, D. C. Hayden, and M. A. Oakes. Implicit Attitudes and Racism: Effects of Word Familiarity and Frequency on the Implicit Association Test. *Social Cognition*, 19(2):97–144, 2001.
- [27] X. Pan, C. Han, K. Dauber, and K. Law. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI and society*, 22:113–132, 2007.
- [28] D. Pereira, E. Oliveira, N. Moreira, and L. Sarmento. Towards an Architecture for Emotional BDI Agents. *2005 Portuguese Conference on Artificial Intelligence*, pages 40–46, 2005.
- [29] B. Shiv and A. Fedorikhin. Hear and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making. *Journal of Consumer Research*, 26(3):278–292, 1999.
- [30] P. Slovic, M. Finucane, E. Peters, and D. G. Macgregor. The Affect Heuristic. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases, The Psychology of Intuitive Judgment*, pages 397–420. Cambridge University Press, 2002.
- [31] B. Steunebrink, M. Dastani, and J.-J. Meyer. A formal model of emotions: Integrating qualitative and quantitative aspects. In *ECAI'08*, pages 256–260. IOS Press, 2008.
- [32] P. Taillandier, M. Bourgais, P. Caillou, C. Adam, and B. Gaudou. A situated bdi agent architecture for the gama modelling and simulation platform. In *Multi-Agent Based Simulation (MABS) workshop @ AAMAS*. MABS, 2016.
- [33] R. Teague. Final report - vol. iv : the statements of lay witness. Technical report, 2009 Victorian Bushfires Royal Commission, 2009. Accessed 2016-12-20.
- [34] M. Weber. The nature of social action. In W. G. Runciman, editor, *Selections in translation*. Cambridge University Press, 1991.
- [35] T. D. Wilson. *Strangers to ourselves: discovering the adaptive unconscious*. Belknap Press of Harvard University Press, Cambridge, Mass. [u.a.], 2002.