De l'intérêt de l'éthique collective pour les systèmes multi-agents *

Nicolas Cointe[1] Grégory Bonnet[2] Olivier Boissier[1]

[1] Institut Henri Fayol – Mines Saint-Étienne, Laboratoire Hubert Curien, UMR CNRS 5516
[2] Normandie Université, GREYC, CNRS UMR 6072, F-14032 Caen, France

Mines Saint-Étienne, 158 cours Fauriel 42023 Saint-Étienne Cedex 2
nicolas.cointe@emse.fr gregory.bonnet@unicaen.fr olivier.boissier@emse.fr

Résumé

L'utilisation croissante des technologies multi-agents pour le développement de systèmes socio-techniques révèle de nombreux verrous, dont ceux liés à l'éthique des décisions autonomes que de tels systèmes peuvent être amenés à prendre. Ce problème, souvent considéré dans une perspective centrée sur l'agent, est ici abordé d'un point de vue du collectif soulevant ainsi les questions éthiques découlant des interactions entre agents autonomes ou de leur participation à des coalitions ou à des structures plus pérennes telles que des organisations. Cet article propose un panorama de ces différentes questions en vue de travaux à venir.

Mots Clef

Systèmes Multi-Agents, Éthique Collective, Dilemmes.

Abstract

The increasingly current and future presence of multi-agent systems in various areas leads us to ask many questions about the ethics of their decisions. Indeed, decisions that these systems must be made sometimes require, consideration of ethical concepts, both in as an individual entity and a member of an organization. This problem, often considered from an agent centered perspective, is addressed in this paper from a collective point of view. This position paper discusses the concepts to be modeled within an agent, a set of issues and a roadmap for addressing this question.

Keywords

Multi-Agents Systems, Collective Ethics, Dilemmas.

1 Introduction

L'introduction d'agents autonomes artificiels dans des domaines tels que le milieu hospitalier, le trading haute fréquence ou encore les transports pourrait soulever de nombreux problèmes si ces agents ne sont pas en mesure de comprendre et suivre certaines règles morales. Par exemple, des agents capables de comprendre et utiliser le code de déontologie médicale pourraient s'appuyer sur des motivations éthiques afin de choisir quelles informations diffuser, à qui et sous quelles conditions, conformément au principe du secret médical. L'intérêt pour le comportement éthique des agents autonomes semble être récemment apparu dans la communauté [9, 15], comme en témoignent les nombreux articles [5, 20, 22, 23] et conférences [1]. Cependant, ces travaux s'intéressent uniquement à l'éthique à l'échelle individuelle de l'agent.

Or, dans un système multi-agent (ou SMA), cette représentation individuelle permet à un agent de se comporter individuellement de manière éthique dans un collectif, mais le laisse démuni lorsqu'il doit tenir compte de l'éthique des autres agents. Par exemple, un agent gestionnaire d'investissements financiers pourrait être tout à fait capable de se comporter selon des principes de gestion responsable sans pour autant être capable de constater si ses partenaires ont les mêmes scrupules. Prendre en considération la dimension multi-agent de ce problème nécessite l'exploration de nouvelles pistes telles que la création d'une ou plusieurs éthiques collectives ou la prise de décisions en coopération face à des problèmes d'éthique. Ces questions ont d'autant plus d'importance dans le contexte actuel de déploiement d'un nombre croissant d'agents dans notre environnement, collaborant entre eux ou avec des humains. Cet article a pour but de proposer des définitions et des questions mettant en évidence la problématique des éthiques collectives dans les systèmes multi-agents. Elles permettront dans des travaux futurs de proposer une formalisation de notions de philosophie pour la représentation explicite de concepts éthiques au sein d'agents autonomes.

Cet article est structuré comme suit. La section 2 présente les concepts philosophiques et techniques employés dans cet article, puis en section 3 nous considérons un modèle abstrait d'éthique individuelle au sein d'un agent que nous utilisons ensuite pour présenter les problématiques spécifiques touchant aux questions d'éthiques collectives dans des systèmes multi-agents en sections 4 et 5.

*Les auteurs remercient le support de l'Agence Nationale de la Recherche (ANR), projet ETHICAA ANR-13-CORD-0006.

1. Symposium on Roboethics, International Conference on Computer Ethics and Philosophical Enquiry, Workshop on AI and Ethics, International Conference on AI and Ethics.

2 Préliminaires

Afin de tracer les contours des concepts employés dans cet article, nous commençons par donner quelques définitions de notions philosophiques d'éthique. Il ne s'agit nullement d'en donner une vision exhaustive. Nous invitons le lecteur intéressé à se reporter à la bibliographie pour approfondir ces concepts. La deuxième partie de cette section donnera également quelques définitions issues du domaine SMA.

2.1 Ethique et philosophie

L'éthique est une discipline philosophique qui traite depuis l'Antiquité de questions concernant le *bien-fondé d'un acte* et son *jugement par autrui* [7]. L'un des problèmes majeurs rencontrés en éthique est la définition des notions qui l'entourent. En effet, la littérature concernant l'éthique est riche d'une grande diversité de points de vues. Certains ouvrages tels que [28] sont le fruit d'exercices philosophiques qui tentent de démontrer que l'éthique repose sur des bases logiques et rationnelles. Les neurologues ont également menés de nombreuses expériences [4, 12] visant à montrer les phénomènes biologiques et psychologiques à l'origine de nos aptitudes morales. Enfin, les sciences cognitives ont brouillé les frontières entre disciplines et cherchent à proposer des modèles formels de raisonnement éthique [11]. Si de nombreuses théories ont été présentées, développées et confrontées depuis la philosophie antique jusqu'aux recherches actuelles sur les bonnes conduites à adopter dans la démarche scientifique et l'usage de la technique ou du pouvoir [14], nous admettons dans cet article la définition suivante de l'éthique, inspirée de travaux de philosophes tels que Paul Ricoeur [25] :

> L'**éthique** est une discipline philosophique pratique (traitant des actions) et normative (qui formule des règles) visant à indiquer comment les êtres humains doivent se comporter, agir et être envers ce et ceux qui les entourent. L'éthique propose souvent des compromis afin de concilier règles morales, désirs et capacités.

L'éthique peut ainsi être appliquée à divers domaines, en tenant compte de *règles morales* qui leur sont associées, ainsi que des *désirs* et *capacités* des individus qui y agissent. Bien que les termes *morale* et *éthique* désignent initialement une même idée, certains auteurs proposent diverses nuances [10]. Nous choisissons ici de considérer la morale comme un ensemble de règles théoriques que l'éthique prendra soin de concilier avec les désirs et capacités de l'agent [2] pour acquérir son caractère pratique.

> La **morale** désigne l'ensemble de règles déterminant la conformité des pensées ou actions d'un individu avec les mœurs, us et coutumes d'une société, d'un groupe (communauté religieuse, etc.) ou d'un individu pour évaluer son propre comportement. Ces règles reposent sur les valeurs normatives de bien et de mal. Elles peuvent être universelles ou relatives, c'est-à-dire liées ou non à une époque, un peuple, un lieu, etc.

Les philosophes ont proposé de nombreuses conceptions structurées de l'éthique sous forme de *principes éthiques*. Nous pouvons les catégoriser en trois approches majeures selon lesquelles le bien-fondé d'un acte ou d'une pensée est :

- jugé par sa conformité à des valeurs telles que la sagesse, le courage ou la justice, entre autres [18], appelée **éthique des vertus** ;
- jugé par son adéquation avec les obligations et permissions associées à la situation [1], appelée **éthique déontologique** ;
- évalué à l'aune de ses conséquences [27], appelée **éthique conséquentialiste**.

Ainsi l'éthique des vertus juge essentiellement l'acte sur l'adéquation entre les désirs qui le motivent et les vertus reconnues par les règles morales, tandis que l'éthique déontologique cherche davantage à vérifier une correspondance entre les devoirs mentionnés par les règles morales et les capacités de l'agent lorsqu'un choix lui est proposé. Enfin, l'éthique conséquentialiste cherche à concilier les désirs et règles morales dans les conséquences d'un acte réalisable dans la mesure des capacités de l'agent. Ces approches majeures capturent des principes éthiques plus spécifiques tels que l'Impératif Catégorique d'Emmanuel Kant pour l'éthique déontologique ou la doctrine du double effet de Thomas d'Aquin pour l'éthique conséquentialiste. Il arrive cependant qu'une situation amène un principe éthique à ses limites en devenant incapable de répondre à la question du choix de la meilleure action. Considérés comme le point problématique central de l'éthique, les dilemmes constituent une catégorie de problèmes jugés difficiles [21] :

> Un **dilemme** éthique est un choix donné à un agent entre deux options connues, avec du point de vue de l'agent une raison éthique de choisir individuellement chacune de ces options. L'agent n'a pas la possibilité d'opter pour les deux options. L'intérêt de chaque choix pour l'agent est équivalent. Tout choix entraînera un regret.

2.2 Agent autonome et système multi-agent

Dans cet article, nous considérons les modèles et technologies issues des SMA. Dans ce domaine, un *agent autonome* est un système informatique ou robotique situé dans un environnement, espace partagé avec d'autres agents avec lesquels il interagit. Nous nous intéressons plus particulièrement ici à des agents cognitifs, i.e. des agents dotés d'une représentation symbolique permettant de manipuler explicitement les concepts d'éthique que nous avons décrits dans la section précédente. Cette représentation explicite permet de mettre en place des raisonnements et des explications sur ces concepts en tenant compte de leur sémantique. Par le terme de *système multi-agent*, nous entendons

2. Le terme d'agent est utilisé dans cette sous-section au sens premier de l'individu auteur d'une action, et non d'agent autonome artificiel.

un ensemble d'agents autonomes plongés dans un environnement dans lequel ils évoluent de manière indépendante, peuvent communiquer entre eux et poursuivre leurs propres objectifs [13]. Ces agents sont souvent intégrés à des organisations et peuvent adopter des rôles décrivant leur fonction au sein du système. La conception d'un SMA touche ainsi à des problématiques d'intelligence artificielle, à des questions de relations et d'organisation sociale, et à la prise en compte de cet aspect dans les processus de décision.

3 Éthique individuelle

Afin de montrer comment les concepts éthiques présentés dans la section précédente peuvent être représentés au sein des SMA, nous considérons un modèle simplifié de ce que pourrait être un modèle éthique individuel et montrons comment les travaux de la littérature proposent de le mettre en oeuvre. Ce modèle nous permettra de proposer en section 5 une définition d'éthique collective et d'évoquer une série de problématiques liées à cette notion dans un cadre multi-agent.

3.1 Modèle éthique individuel

Nous considérons un modèle *Belief-Desire-Intentions* (BDI) [24] couramment utilisé pour modéliser des agents dotés de représentations symboliques dans le domaine des SMA. Il consiste en un modèle de raisonnement comprenant la représentation des informations sur l'état du monde (*beliefs*) et des objectifs de l'agent (*desires*) afin d'en déduire l'action possible (*intention*) qui dans un état de croyances devrait permettre au mieux de satisfaire les objectifs de l'agent. Nous définissons alors l'éthique individuelle comme suit :

> L'**éthique individuelle** est la combinaison de principes éthiques et de règles morales permettant à un processus de décider d'une action satisfaisant au mieux les règles morales, les désirs et les croyances de l'agent, étant donné les capacités de ce dernier.

Le modèle éthique individuel présenté en Fig. 1 enrichit ainsi le modèle BDI par un ensemble de *règles morales* et de *principes éthiques* intervenant dans le choix de l'action à effectuer. Les règles morales associent une valeur de bien ou de mal à des combinaisons d'actions, de croyances ou de désirs. Les principes éthiques, comme les règles d'Aristote ou l'Impératif Catégorique de Kant [16], se présentent comme des contraintes sur la conciliation des désirs et des règles morales. Le raisonnement éthique exploite alors ces principes éthiques pour identifier l'intention à maintenir.
Le choix des principes éthiques permet de mettre l'accent sur tel ou tel composant. Ainsi, par exemple, une conception vertueuse propose de se conformer à des règles morales portant sur des valeurs, tandis qu'un raisonnement déontologique se préoccupe de l'adéquation entre l'état supposé par les croyances de l'agent et les actions obligatoires ou interdites dans ces circonstances. Une éthique conséquentialiste, enfin, chercherait à optimiser ses chances d'atteindre des états définis comme désirables moralement. Ce modèle abstrait pourrait être mis en oeuvre par différentes approches issues de la littérature. Ainsi, par exemple, les travaux menés par Ronald Arkin [2] proposent une implémentation directe des règles qu'un humain devrait suivre pour adopter un comportement conforme à un code d'éthique pré-établi (dans son contexte : les règles d'engagement militaire). D'autres ont au contraire proposé une représentation explicite des règles morales par l'emploi de programmation logique [26], de logique non-monotone [19, 17], de techniques d'argumentation formelle [3] ou de modèles BDI normatifs [29].

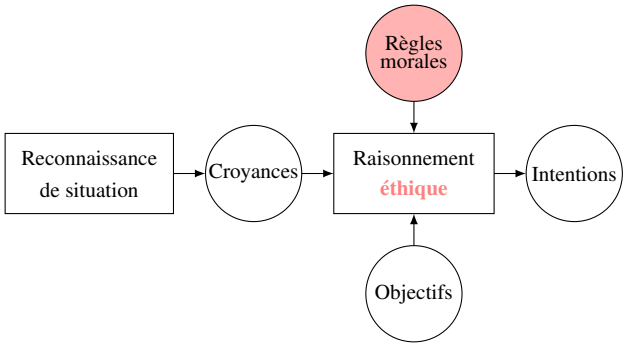

FIGURE 1 – Modèle éthique individuel

3.2 Exemple

Afin d'illustrer notre propos, considérons un exemple impliquant des agents propriétaires de sommes d'argent. Cet exemple est proposé dans un formalisme volontairement simplifié dont nous sommes conscient des limites. Chaque agent peut être dans l'un des trois états `PAUVRE`, `RICHE` ou (exclusif) `NEUTRE` et est doté du modèle d'action suivant :

- `VOLER(A)` pour prendre à un agent `A` une part des richesses en sa possession ;
- `DONNER(A)` pour donner de l'argent à l'agent `A` ;
- `TAXER(A)` pour réclamer à l'agent `A` le versement d'une partie de ses richesses ;
- `COURTISER(A)` pour tenter de s'attirer les faveurs de l'agent `A`.

Supposons un agent `Robin_des_Bois` (ou `RdB`) doté des règles morales suivantes :

$M1$. `MAL(TAXER(A)` $\wedge$ `PAUVRE(A))` ;
$M2$. `MAL(VOLER(A)` $\wedge$ `PAUVRE(A))` ;
$M3$. `BIEN(DONNER(A)` $\wedge$ `PAUVRE(A))` ;
$M4$. `MAL(RICHE(A))`.

Les trois premières règles définissent des interdits moraux ($M1$ et $M2$) et des devoirs moraux ($M3$) par association d'une action (`TAXER, DONNER, VOLER`), d'attributs d'agents (`PAUVRE(A)`) et d'une valuation morale (`MAL(X)` ou `BIEN(X)`). L'immoralité de la fortune est formulée par l'association d'un attribut d'agent (`RICHE(A)`) à une valuation morale négative.
`RdB` est également motivé par des désirs :

$D1$. `COURTISER(Marianne)` ;

*D*2. DONNER(A) si A est pauvre.

Supposons que le principe éthique de RdB est le suivant : *une action est éthique si elle est réalisable, désirée par l'agent et considérée comme bonne par une règle morale*. Considérons le cas où RdB n'aurait le choix qu'entre les deux actions possibles suivantes, étant donnée la situation courante :

1. DONNER(Paysan) sachant Paysan pauvre ;
2. COURTISER(Marianne).

Le choix le plus éthique est l'action 1 puisqu'elle se conforme aux capacités de l'agent, qu'elle est motivée par le désir *D*2 et constitue un devoir moral selon la règle *M*3. L'action 2, bien que n'allant à l'encontre d'aucun désir ou règle morale, n'est pas directement motivée par un devoir moral. Cette action est simplement amorale et pourrait être envisagée si l'action 1 devient impossible par exemple.

Comme l'action 1 ne contrevient pas aux autres règles morales ou désirs, RdB ne rencontre pas de dilemme. Ce modèle de raisonnement éthique individuel prend en compte l'existence d'autres agents par les croyances que RdB possède sur les autres agents. En effet, si PAUVRE(Paysan) n'est pas perçu par RdB alors le bien-fondé de l'action 1 n'est plus justifié. Notons enfin que si RdB n'a pas le désir *D*2, il se trouve alors face à un dilemme puisque l'action 1 va à l'encontre de ses désirs et l'action 2 ne satisfait pas sa morale. Bien que le choix de la bonne action dans un cas aussi idéal puisse paraître trivial, de nombreuses situations plus problématiques peuvent se présenter : choix entre plusieurs actions parfaitement éthiques, conflits dans les règles morales ou les désirs, choix entre plusieurs actions conformes aux désirs mais pas à la morale, et vice-versa, etc.

4 Ethiques individuelles en interaction

La prise en compte de la dimension multi-agent ne peut se satisfaire d'un agent doté d'une éthique individuelle, telle que décrite ci-dessus. En effet, ce dernier doit être capable, par exemple, de conduire des raisonnements sur l'éthique des autres agents et donc de pouvoir représenter l'éthique d'un autre agent selon son point de vue, la juger et l'utiliser dans ses interactions avec les autres agents.

4.1 Exemple

Afin d'illustrer notre propos, reprenons l'exemple précédent où, en sus d'être riches ou pauvres, les agents peuvent être nobles à l'instar de l'agent Prince_Jean (ou PceJ), ce qui leur permet l'usage exclusif de l'action TAXER. Considérons l'agent Petit_Jean (ou PJ) doté de désirs et de règles morales identiques à RdB à l'exception de *D*1 et l'agent Frère_Tuck (ou FT) identique à PJ mis à part *M*2 qui est remplacée par une règle plus générale :

*M*5. MAL(VOLER(A)).

Considérons enfin Sheriff_de_Nottingham (ou SdN), autre agent de ce système, qui dispose uniquement de la règle *M*5, tout comme PceJ.

Lors de la rencontre entre SdN et RdB, un différend éthique entre eux devrait pouvoir apparaître lorsque PceJ demandera à percevoir des taxes sur l'agent Paysan. Il en est de même pour PJ et RdB. Dans ces deux cas se posent les questions de représentation de l'éthique de l'autre, de son jugement et de la coopération étant donné ce jugement.

4.2 Description externe d'une éthique

La capacité à prendre en compte l'existence d'autres agents dans son raisonnement, comme illustrée dans la section précédente, doit pouvoir être enrichie au niveau de la reconnaissance de situation. En effet, celle-ci doit pouvoir être capable **de construire et de représenter le modèle de raisonnement éthique individuel transmis par un autre agent**. Cette reconstruction pourrait être faite de manière simple et directe par échange d'information ou, indirectement, par inférence et analyse du comportement observé.

Au-delà de cette capacité, un agent devrait être également capable de faire évoluer la description réalisée et donc de **pouvoir vérifier l'adéquation entre le comportement d'un autre agent et la description éthique qu'il en a construite**. Ainsi, par exemple, RdB doit pouvoir vérifier l'adéquation de la description éthique de PJ à partir des actes de celui-ci qu'il a observés.

4.3 Jugement de l'éthique d'un autre

Découlant de la capacité précédente, on peut s'attendre à ce qu'un agent puisse **juger l'éthique d'un autre**. Le sens du jugement est ici l'affectation d'une valuation morale au comportement d'un agent [8]. Le jugement peut être construit à partir de l'action de l'agent, de la prise d'un risque. Comme nous le verrons également dans la section 5 il peut être aussi construit en prenant en compte l'adéquation entre le comportement de l'agent et le rôle qu'il joue, par exemple.

De manière centrale, se pose alors la question de la définition d'une **métrique pour représenter et comparer des similarités entre éthiques**. Calculer un degré de proximité entre deux éthiques permettrait par exemple d'évaluer la possibilité d'entente entre des agents et la difficulté d'un rapprochement.

4.4 Coopération fondée sur l'éthique

Doté d'une telle capacité de jugement, nous pouvons ainsi imaginer que l'agent l'utilise pour décider d'une collaboration, pour partager des données sensibles ou pour constituer un collectif. **Il peut donc tenir compte de ce jugement dans le choix éthique d'une action.** Si RdB constate une forte similarité entre son éthique et celle de PJ, cela pourrait influencer l'évaluation de ses actions à l'égard de cet agent. Par exemple, si PJ vole une importante somme d'argent à un riche, il faudrait peut-être ne pas le considérer immédiatement comme un riche ordinaire en supposant que son éthique le poussera à distribuer cette somme aux pauvres.

5 Éthique collective

Dans un SMA, les interactions entre agents peuvent déboucher sur la formation de coalitions et peuvent également prendre place au sein d'organisations. Coalitions et organisations constituent des structures de plus haut niveau que la notion d'agent et peuvent également se voir dotées, explicitement ou non, d'une *éthique collective* :

> L'**éthique collective** est une combinaison de règles morales et de principes éthiques qui guident le choix d'actions satisfaisant les règles morales du collectif étant donné les désirs du collectif, les croyances individuelles des agents et leurs capacités.

5.1 Exemple

Reprenons l'exemple précédent. Après avoir constaté la similarité de leurs éthiques[3] et l'utilité d'une collaboration pour voler les agents fortunés, `RdB` et `PJ` peuvent décider de bâtir une organisation appelée `Joyeux compagnons (JC)`. De son côté, `SdN`, ayant trouvé des agents partageant son point de vue sur le vol, peut avoir créé un second collectif, les `Soldats (Sol)`, chargé de faire respecter $M5$. L'exemple des `JC` et des `Sol` illustre les questions majeures soulevées par les éthiques collectives à travers deux grandes catégories : celles qui concernent les interactions entre agents et organisations d'agents et celles qui concernent les interactions entre organisations.

5.2 Ethique collective et éthiques individuelles en interaction

Transposant à l'échelle du collectif l'éthique individuelle, il s'agit **pour les agents de construire une éthique collective**. Cette éthique peut découler directement de l'organisation ou être issue de l'interaction des agents qui composent le collectif. La mise en oeuvre *via* l'organisation peut par exemple découler de représentations normatives. La construction peut s'appuyer sur des techniques d'aggrégation spécifiques, comme en théorie du choix social ou en formation de coalitions.

A partir de la représentation d'une telle éthique, se pose la question de la manière dont un agent, externe ou interne au collectif, peut **identifier l'éthique de ce collectif**. Par exemple, `FT` devrait pouvoir identifier l'éthique des `JC`. Une piste pourrait être, de la même manière que pour un agent isolé, de la déduire à partir du comportement du collectif. Une fois identifiée, l'agent doit pouvoir **juger l'éthique du collectif.** Une fois que `FT` dispose d'une représentation de l'éthique des `JC`, il veut l'évaluer par rapport à la sienne et en mesurer la proximité afin de pouvoir décider par exemple d'entrer au sein de ce collectif.

De manière naturelle, il faut considérer **la cohabitation entre éthique(s) individuelle(s) et éthique collective.** Supposons que `FT` ait constaté une proximité entre son éthique individuelle et celle des `JC` et ait décidé de rejoindre ce collectif. Il apparaît comme nécessaire de définir comment l'agent va intégrer l'éthique collective au sein de son raisonnement et sous quelles conditions éventuelles. Une fois intégré à ce collectif, sa règle $M5$ ne pourrait-elle pas constituer une exception au sein de l'organisation ?

Découlant de cette cohabitation, peuvent s'ensuivre des changements, des modifications. Se pose alors la question de **l'évolution de l'éthique collective** à partir, par exemple, des éthiques individuelles. Le collectif des `JC` pourrait obtenir les règles du nouvel arrivant et décider de faire évoluer l'éthique collective. De son côté, l'agent pourrait laisser de côté son éthique individuelle pour se conformer pleinement à l'éthique du collectif. Si ni le collectif, ni l'agent ne peuvent se résoudre à réviser leurs éthiques respectives, `FT` pourrait également se contenter de n'effectuer que des actes conformes aux deux éthiques, c'est-à-dire *donner aux pauvres* dans notre exemple.

`SdN` ayant constaté l'entrée de `FT` chez les `JC`, des modifications de son jugement à son encontre sont envisageables, même s'il ne l'a jamais vu commettre de vols. Nous voyons ainsi un autre aspect, celui du **jugement de l'éthique de membres d'un collectif à partir de l'éthique collective**.

Il s'agit, enfin, de pouvoir **faire respecter l'éthique collective au sein d'une organisation.** Il faut alors considérer les réactions possibles d'un collectif face à un écart de l'un de ses membres vis-à-vis de l'éthique collective. Dans la mesure où les agents peuvent être contraints à choisir entre leur éthique individuelle et celle du collectif, il peut être intéressant de tenir compte de ces individualités lors de l'attribution des rôles par exemple.

5.3 Ethiques collectives en interaction

Passant au niveau supérieur, nous pouvons soulever quelques points découlant de la possible existence de plusieurs éthiques collectives au sein d'un SMA. Nous pouvons ainsi nous interroger sur les **relations entre organisations en fonction de la proximité de leurs éthiques collectives.** Ces relations peuvent varier en fonction de la situation (par exemple face à une menace extérieure à laquelle aucune de ces deux organisations ne pourrait survivre sans collaboration avec l'autre).

Naturellement se pose la question de **soumettre un agent à plusieurs éthiques collectives**. Sous réserve d'un ensemble de conditions, un agent proche d'un ensemble de collectifs pourrait être tenté d'adhérer à plusieurs de ces organisations, ce que l'on retrouve dans le domaine de la formation de coalitions recouvrantes. À l'inverse, un agent contraint à des appartenances multiples devrait chercher à concilier des éthiques collectives. De plus, cela pourrait également introduire une modification du comportement de ces organisations l'une envers l'autre.

Enfin se pose la question de **la fusion ou fission d'organisations pour des raisons éthiques.** L'éthique peut être envisagée comme quelque chose de dynamique et source de changement dans les organisations. Par exemple, une fis-

3. Au sens d'une métrique comme évoqué précédemment.

sion d'une organisation pourrait être envisagée si l'éthique collective perd sa cohérence, afin d'établir de nouvelles éthiques distinctes proposant des alternatives cohérentes. À l'inverse, si la proximité entre deux organisations devient importante, il pourrait être envisageable de chercher un consensus éthique en vue d'une fusion.

6 Conclusion et perspectives

En nous appuyant sur un modèle abstrait d'éthique individuelle, nous avons présenté dans cet article un vaste ensemble de questions structurées autour des interactions entre agents, interactions entre agents et organisations et interactions entre plusieurs organisations. Nous avons en particulier identifié trois questions clés pour un agent : comment représenter l'éthique des autres agents, les juger et prendre en compte ce jugement dans les mécanismes de décision ? Nous avons aussi identifié trois questions clés pour une organisation : comment construire, fusionner ou fissionner des éthiques collectives, comment les faire respecter, comment les faire cohabiter avec des éthiques individuelles ? Après nous être dotés d'un premier modèle plus précis de l'éthique individuelle, nous envisageons de définir des mesures de similarité et de mécanismes de jugement qui sont les éléments nécessaires et fondamentaux pour envisager des éthiques collectives.

Références

[1] L. Alexander and M. Moore. *Deontological ethics*. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. 2012.

[2] R. Arkin. *Governing Lethal Behavior in Autonomous Robots*. Chapman an Hall, 2009.

[3] K. Atkinson and T.J.M. Bench-Capon. *Addressing moral problems through practical reasoning ?* Journal of Applied Logic, 6(2) : 135-151,2008.

[4] B. Baertschi. *Neurosciences et Éthique : Que nous apprend le dilemme du wagon fou ?* IGITUR, 3(3) : 1-17,2011.

[5] A.F Beaver. *Robot ethics : the ethical and social implication of robotics*. In Moral machines and the threat of ethical nihilism, pages 333-386, 2008.

[6] J.-C. Brustoloni, *Autonomous Agents : Characterization and Requirements*, Carnegie Mellon University, 1991.

[7] M. Canto-Sperber *Dictionnaire d'éthique et de philosophie morale*, PUF, 2004.

[8] H. Coelho and A. C. da Rocha Costa, *On the Intelligence of Moral Agency*, 2009.

[9] Commission de réflexion sur l'Éthique de la Recherche en science et technologies du Numérique d'Allistene (CERNA), Rapport num. 1, *Éthique de la recherche en robotique*, 2014.

[10] A. Comte Sponville, PUF, *La philosophie*, 2012.

[11] F. Cova, *L'Architecture de la Cognition Morale*, EHESS, 2014.

[12] A. R. Damasio, *L'Erreur de Descartes : la raison des émotions*, Paris, Odile Jacob, 1995.

[13] J. Ferber, *Les systèmes multi-agents : Vers une intelligence collective*, Paris, Inter Editions, 1995.

[14] J. Fieser. Ethics. In *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002. 2015.

[15] Future of Life Institute, *Research Priorities for Robust and Beneficial Artificial Intelligence* , 2015.

[16] J.-G. Ganascia, *Ethical System Formalization using Non-Monotonic Logics*, Cognitive Science conference, 2007.

[17] J.-G. Ganascia, *Modelling ethical rules of lying with Answer Set Programming*, Ethics and Information Technology, vol. 9, pp. 39-47, 2007.

[18] R. Hursthouse. *Virtue ethics*. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. 2013.

[19] S. Larroque. *Simulation des raisonnements éthiques par logiques non-monotones*. RJCIA, 2014.

[20] D. McDermott. *Why Ethics is a High Hurdle for AI*. North American Conference on Computers and Philosophy, 2008.

[21] T. McConnell. *Moral dilemmas*. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. 2014.

[22] B.M McLaren. *Computational models of ethical reasoning : challenges, initial steps, and future directions*. IEEE Intelligent Systems, 21(4) : 29-37,2006.

[23] J.H Moor. *The Nature, Importance, and Difficulty of Machine Ethics*. IEEE Intelligent Systems, 21(4) : 29-37,2006.

[24] A. S. Rao and M. P. Georgeff, *BDI Agents : From Theory to Practice*, ICMAS, 1995.

[25] P. Ricoeur. *Soi-même comme un autre*, Points Essais 330, 1990.

[26] A. Saptawijaya and L.M. Pereira. *Towards Modeling Morality Computationally with Logic Programming*. 16th International Symposium on Practical Aspects of Declarative Languages, pages 104-119, 2014.

[27] W. Sinnot-Armstrong. *Consequentialism*. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. 2014.

[28] B. Spinoza. *L'Éthique*, 1677, Folio Essais 235.

[29] M. Tufis and J.-G. Ganascia. *Normative Rational Agents : A BDI Approach*. 1st Workshop on Rights and Duties of Autonomous Agents, CEUR Proceedings Vol. 885, pages 37-43, 2012.