Raisonner sur l'autonomie d'un agent au sein de systèmes multi-agents ouverts : une approche basée sur les relations de pouvoir

Cosmin Carabelea

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

THESE

présentée par

Cosmin Carabelea

Pour obtenir le grade de Docteur
de l'Ecole Nationale Supérieure des Mines de Saint-Etienne

Spécialité : Informatique

Raisonner sur l'autonomie d'un agent au sein de systèmes multi-agents ouverts – Une approche basée sur les relations de pouvoir

Soutenue à Saint-Etienne, le 3 décembre 2007

Membres du jury:

Rapporteurs :	
Amal EL FALLAH-SEGHROUCHNI	Professeur, Université Paris VI
Catherine TEISSIER	Ingénieur de recherches, CERT Toulouse
Rapporteur externe :	
Cristiano CASTELFRANCHI	Professeur, ISTC-CNR, Rome
Directeurs de thèse :	
Olivier BOISSIER	Professeur, ENS Mines de Saint-Etienne
Adina Magda FLOREA	Professeur, Univ. "Politehnica" de Bucarest
Examinateurs :	
Andreas Herzig	Professeur, CNRS IRIT, Toulouse
Vincent Louis	Docteur, France Telecom R&D

Remerciements

Je tiens à remercier tout un ensemble de personnes grâce à qui j'ai pu mener ce travail à bien. Je commence par remercier mon jury de thèse et mes rapporteurs pour avoir pris le temps de critiquer mon travail. Je remercie aussi toutes les personnes qui, d'une manière directe ou indirecte, ont encadré mon travail et m'ont positionné dans les bonnes directions. Je pense ici tout particulièrement à Cristiano, mais aussi à Adina, Michael et bien d'autres.

Merci aussi aux membres de l'equipe SMA qui m'ont aidé et avec qui ça a été un plaisir de travailler, ainsi que toutes les personnes avec qui j'ai interagit pendant ma thèse et avec qui j'ai pu collaborer, trop nombreuses pour les énumerer . Sur un plan personnel, je tiens à remercier ma femme qui a toujour été là pour me soutenir et me remonter le morale.

Plus formellement, je tiens à remercier l'Ambassade de France en Roumanie pour avoir partiellement financé ce travail, ainsi que la Région Rhône-Alpes pour avoir financer mes séjours à Rome.

Finalement, mes plus chauds remerciements vont à Olivier, qui a toujours été le moteur qui m'a poussé vers l'avant. Il a cru sans arrêt en mes idées et il a fait preuve d'une confiance et d'une patience enormes. Il est pour beaucoup dans ce travail et résultat et pour ceci je lui suis reconnaissant.

Résumé

Ces dernières années, les systèmes multi-agents ont été utilisés dans différents domaines d'application qui ont bénéficié de cette approche basée sur la coordination et l'autonomie. Parmi ces domaines, cette thèse s'est plus particulièrement intéressée à des scénarios d'applications d'*intelligence ambiante*. Dans un scénario type de ce domaine, comme celui que nous avons implémenté et qui est décrit dans cette thèse, des agents se déplacent d'un système ouvert à un autre pour mieux satisfaire les objectifs de leurs utilisateurs. Dans cette thèse nous ne nous intéressons pas à la conception d'un système ouvert, ni d'un agent appartenant à un tel système. Nous prenons le point de vue d'un agent sur le point d'entrer dans une organisation multi-agent, où il aura à choisir parmi plusieurs rôles disponibles. *Notre objectif est de donner à cet agent les moyens d'évaluer avant d'entrer dans une organisation les implications que cette action aura sur ses capacités de satisfaire les désirs de son utilisateur.*

Quand il entre dans une organisation, un agent devient le sujet de plusieurs contraintes imposées par des entités externes (d'autres agents, l'organisation même, la société). Ces contraintes limitent sa prise de décision et son comportement. Elles proviennent du besoin de coordination dans le système. Comme le montre cette thèse, l'*autonomie sociale de décision* est une caractéristique d'un agent, qui est directement liée à ces contraintes sociales. En conséquence, nous basons l'évaluation faite par un agent des contraintes qui lui sont imposées sur le concept d'autonomie,. Nous proposons un *modèle formel* qui utilise des concepts tels que engagement social, dépendance, pouvoir individuel et social,. Ce modèle nous permet de donner une définition formelle de l'autonomie. Equipé de ce modèle, un agent est capable d'évaluer les relations de pouvoir dont il est l'objet et d'en calculer son autonomie. Il peut également évaluer la manière dont elles évoluent en fonction de son comportement. Nous analysons dans cette thèse comment un agent peut utiliser ce modèle à base de pouvoir social et autonomie pour enrichir son raisonnement sur les implications d'entrer dans une organisation multi-agent. Cette approche est illustrée à l'aide d'un scénario d'intelligence ambiante qui a été implémenté et déployé sur des objets portables.

Abstract

In the last years the multi-agent paradigm has been used in a variety of application domains that have benefited from its approach based on coordination and autonomy. Among these domains, this thesis is particularly interested in *ambient intelligence* applications. In typical scenarios of this type, such as the one implemented and described at the end of this work, agents move from an open system to another in order to satisfy their users' objectives. In this thesis we are not interested in designing an open multi-agent system, nor an agent part of such a system. We consider the point of view of an agent about to enter a multi-agent organization or system, or simply about to play a role or to choose between two possible roles. *Our aim is to provide an agent with means to evaluate before entering the organization or playing the role the implications this action will have on its ability to satisfy its user's objectives*.

By entering an organization or playing a role, an agent is subject of constraints imposed from external sources (other agents, the organization itself, the society). These constraints limit its decision-making and behaviour and are issued from the need to have coordination in the system. As we argue in this thesis, the *social autonomy in deliberation* is an agent's characteristic directly linked to these constraints originating from social sources. Therefore, we base an agent's evaluation of the possible roles to play on the concept of autonomy, although this concept does not have a commonly agreed definition in the multi-agent community. This thesis proposes a *formal model* that uses concepts such as social commitments, dependencies, individual and social powers, model that allows us to provide a formal definition of autonomy. Equipped with this model, an agent is able to evaluate its current powers, power relations and autonomy and how they will change depending on its behaviour, e.g., when it plays a role. We discuss in this thesis how an agent can make use of the social power and autonomy model we propose to *enrich its reasoning* on the implications of entering a society or playing a role. This approach, of using our formal model to improve an agent's reasoning, is illustrated with an ambient intelligence scenario implemented and deployed on handheld devices.

Table des matières

Introduction

Introduction générale

Ces dernières années, les *systèmes multi-agents* (Lesser 1995, Ferber 1995) ont proposé un paradigme informatique basé sur la *coordination* et l'*autonomie*. Un agent intelligent (Wooldridge, 1999) est généralement considéré capable d'un comportement autonome, réactif, proactif et social en vue d'atteindre les objectifs pour lesquels il a été conçu. Etant immergé dans un système multi-agent, un agent doit agir sur son environnement qu'il partage avec d'autres agents, interagir avec ceux-ci et obéir aux règles d'une organisation. Une des problématiques centrale des systèmes multi-agents est de coordonner les agents entre eux afin d'établir un comportement global cohérent du système tout en permettant aux agents de contrôler localement leur comportement (autonomie). Le problème provient du fait que le comportement choisi au niveau local par un agent peut résulter en un comportement global différent de celui désiré par le concepteur du système.

Malgré les problèmes qu'il provoque, ce contrôle local des agents représente un des atouts du paradigme multi-agent. Il est essentiel aux applications cibles de ce paradigme. Ainsi par exemple dans le cas de problèmes de logistique (p.ex. Valckenaers, 2004) – chaîne de production, réseaux de transport, etc. – l'état global du système n'étant pas accessible dans des délais raisonnables, chaque élément du système doit prendre des décisions localement. Un autre domaine d'application, dans lequel l'approche multi-agent est de plus en plus utilisée, est celui de l'*intelligence ambiante* (Weiser, 1991). Les scénarios typiques de ce domaine montrent des situations où des services disponibles dans l'environnement sont fournis aux utilisateurs d'une manière parfois réactive et parfois proactive grâce aux nombreux capteurs ou dispositifs situés dans l'ambiant des utilisateurs. Ces dispositifs prennent localement des décisions et interagissent les uns avec les autres pour offrir aux utilisateurs des services adaptés à leurs besoins.

Il existe des correspondances évidentes entre les besoins des applications d'intelligence ambiante et les caractéristiques du paradigme multi-agent. Cependant, l'utilisation de ce dernier dans les scénarios d'intelligence ambiante (et même dans d'autres domaines d'application), pose un défi de conception : l'ouverture. Par leur nature, les applications d'intelligence ambiante représentent des *systèmes ouverts* et *très dynamiques*, dont les éléments peuvent varier souvent. Par exemple, un agent peut se déplacer (souvent virtuellement) entre plusieurs systèmes afin de proposer les meilleurs services à son utilisateur ou bien l'ambiant de l'utilisateur (et donc le système dans lequel son agent personnel agit) peut changer suite à la mobilité de l'utilisateur.

Deux grandes problématiques de conception se posent ainsi dans les systèmes multi-agents. D'un coté, comment concevoir des systèmes qui fonctionnent correctement malgré leur composition dynamique et changeante : des agents différents, inconnus au moment de la conception (potentiellement dangereux) entrent et sortent du système ou changent de comportement. D'un autre coté, comment concevoir des agents qui fonctionnent correctement malgré les changements qui se produisent autour d'eux : appartenance à des systèmes différents avec des règles différentes, interaction avec des types d'agents différents.

La conception des systèmes multi-agents ouverts pose plusieurs problématiques spécifiques dont, par exemple, celui de la gestion de l'ouverture : comment accueillir les agents qui entrent dans le système, comment fournir aux agents des services qui leur permettent de retrouver d'autres agents, comment gérer le départ des agents, etc. D'autres problématiques s'intéressent aux problèmes liés à l'hétérogénéité des agents : dans un système ouvert il est possible que des agents conçus différemment entrent et il est donc impossible pour le concepteur du système de faire des hypothèses sur le mode de raisonnement et/ou de comportement de ces agents. Des règles sont ainsi mises en place, règles qui essaient d'assurer une certaine cohérence dans le système (Vázquez-Salceda, 2004). Cependant, en général ces règles ne contrôlent pas complètement le comportement des agents du système, ce qui entraîne un autre problème : comment s'assurer que des agents malveillants ne rentrent pas dans le système ou s'ils rentrent, comment limiter les dégâts qu'ils peuvent y produire (Muller, 2006). Par exemple, un agent malveillant peut ne pas respecter les règles du système ou les utiliser dans son intérêt – même si une telle désobéissance est détectée, il est parfois difficile de sanctionner un tel agent, comme nous le verrons plus tard.

Dans ce manuscrit, nous ne proposons pas des solutions à ces problèmes liés à la conception des systèmes multi-agents ouverts (comment concevoir un tel système, comment assurer une cohérence de comportement, etc.). En revanche, nous nous intéressons aux conséquences de ces problèmes et à leurs solutions possibles sur la conception des agents. Autrement dit, notre travail se situe dans *la conception d'agents intelligents capables d'agir à l'intérieur de ou par rapport à un système multi-agent ouvert ou dynamique*.

Objectifs

Nous essayons de répondre dans ce manuscrit à des questions liées aux implications de la dynamique d'un système sur le raisonnement ou sur le comportement d'un agent qui en fait partie. Etant donnés les changements du cadre d'action d'un agent (changement lié à l'environnement, aux autres agents ou au système même), nous voulons être capables de *concevoir un agent qui représente et analyse ces changements et leurs implications* sur son comportement et son raisonnement. Pour donner un exemple, nous voulons avoir un agent capable de comprendre comment un changement de rôle ou de groupe dans une organisation lui facilitera ou le gênera dans la satisfaction de ses objectifs . Il s'agit donc de prendre une décision plus informée.

Pour répondre à ce genre de questions, nous considérons nécessaire de pouvoir représenter les changements qui ont lieu dans la prise de décision et le comportement d'un agent quand il change de rôle ou de groupe (par exemple) dans un système. Ces changements sont principalement liés aux *interactions* diverses qu'il a avec d'autres agents et à la modification des règles du système qu'il doit respecter (comme nous le verrons plus tard, ces règles sont modélisées généralement sous la forme d'*organisations*). Comme nous allons l'argumenter dans la suite de ce manuscrit, les interactions et les organisations créent ou imposent des *contraintes* qui limitent le processus de décision et le comportement des agents. Notre approche s'intéresse à la représentation de ces contraintes qui limitent le comportement d'un agent. L'agent pourra ainsi raisonner dessus et prendre des décisions telles que celles d'entrer ou non dans un système ou organisation, de changer ou non de rôle, etc.

Les deux objectifs principaux de ce travail sont donc :

- *de proposer une modélisation des différentes contraintes externes sur le comportement et la prise de décision d'un agent*
- *démontrer les avantages d'utiliser une telle modélisation dans le raisonnement d'un agent*

Nous verrons par la suite qu'il existe plusieurs types de contraintes qui limitent les agents. Plusieurs modèles différents ont été proposés dans la littérature pour ces contraintes. Cependant, comme nous argumenterons dans la partie suivante, nous considérons que le concept d'autonomie est directement lié à ces contraintes. Cette relation entre l'autonomie et ces contraintes nous sera très utile dans la modélisation proposée dans cette thèse. Notre approche dans ce manuscrit sera ainsi de *représenter les contraintes qui limitent la prise de décision et le comportement des agents à travers la notion d'autonomie*. Pour cela nous nous appuierons notamment sur les relations de pouvoir qui tissent un réseau explicitant ces différentes contraintes au sein duquel un agent doit pouvoir se positionner et calculer son autonomie.

Nous voulons souligner que cette représentation est interne à un agent et peut être différente des modèles utilisés dans la conception du système qui contient l'agent. La modélisation que nous proposons permet à un agent de représenter les mécanismes de coordination dans un système multi-agents, même si ces mécanismes sont représentés différemment au niveau du système lui-même. De ce point de vue, notre travail reste assez original dans la communauté scientifique des systèmes multi-agents, où généralement est prise une perspective externe qui se place du point de vue de la conception du système, c.-à-d., de la modélisation des mécanismes de coordination. Autrement dit, nous ne proposons pas une approche pour concevoir le monde, mais proposons une approche permettant de percevoir certains aspects de ce monde d'un point de vue local.

Dans ce travail, nous proposons donc qu'un agent regarde le monde autour de lui (son environnement, les interactions avec d'autres agents, l'organisation dans laquelle il évolue) à travers des lunettes qui représentent son degré d'autonomie. Par ailleurs, nous ne voulons pas seulement qu'un agent puisse se percevoir en terme d'autonomie qu'il peut avoir dans le système mais nous voulons aussi qu'il raisonne sur l'autonomie. Ce *raisonnement à base d'autonomie* (possible seulement si une définition formelle de l'autonomie est donnée) représente en fait un raisonnement sur les contraintes qui limitent le comportement et la prise de décision de l'agent. Nous décrirons plusieurs situations dans lesquelles ce raisonnement peut s'avérer très utile pour un agent, surtout dans le cas des systèmes ouverts et dynamiques. Par exemple, si un agent a le choix d'appartenir à différents groupes, il pourra représenter les contraintes qui le limiteront dans ces groupes, et donc calculer quelle sera son autonomie. Il pourra ainsi choisir le groupe qui lui laissera l'autonomie nécessaire pour satisfaire ses objectifs.

Structure du manuscrit

La structure de ce manuscrit suit les idées présentées dans la section précédente: nous nous intéressons aux différents types de contraintes qui limitent le comportement et la prise de décision d'un agent. Nous proposons une modélisation de la relation entre ces contraintes et le concept d'autonomie, ce qui résultera dans une définition formelle de ce concept. Finalement, nous décrivons le raisonnement d'un

agent sur ces contraintes, donc à base d'autonomie, dans plusieurs situations que nous considérons intéressantes.

Etat de l'art: autonomie, contraintes interpersonnelles et institutionnelles et raisonnement

Dans la première partie de ce manuscrit nous faisons un état de l'art sur les travaux existants liés au notre. Nous nous intéressons ainsi au concept d'*autonomie* dans les systèmes multi-agents, en faisant ressortir pourquoi ce concept joue un rôle central dans notre approche. Comme nous l'expliquerons, les contraintes qui limitent le comportement et la prise de décision d'un agent surgissent du besoin d'avoir une coordination dans un système multi-agent. Nous divisons les approches existantes sur la coordination dans deux catégories : *coordination institutionnelle* (à base des concepts tels que rôles, normes, organisations, etc.) et *coordination interpersonnelle* (à base des interactions entre agents) et nous décrivons les différentes modélisations proposées dans la littérature. Finalement, comme nous voulons enrichir le raisonnement d'un agent en lui permettant d'utiliser la notion d'autonomie, nous nous intéressons également aux architectures d'agent (et modèles de prise de décision) les plus utilisées dans les systèmes multi-agents.

Modèle d'autonomie à base de pouvoir social

Dans la deuxième partie, nous présentons la première contribution de ce travail, un modèle formel des différents types de contraintes externes qui limitent un agent. En utilisant quelques-unes des modélisations les plus utilisées dans la littérature, telles que les engagements sociaux ou le pouvoir social, nous construisons peu à peu une représentation uniforme de toutes ces contraintes, représentation interne à un agent. Cette représentation à base de relations de pouvoir social débouche sur une définition formelle du concept d'autonomie.

Mise en œuvre et illustrations : Raisonnements à base d'autonomie

La troisième partie consiste en l'illustration de différents raisonnements à base d'autonomie et en la mise en œuvre de ce modèle sur l'application ADOMO. Parmi les raisonnements illustrés, nous nous intéressons plus particulièrement à celui que nous considérons très important dans les systèmes ou organisations dynamiques : le choix d'un rôle au sein d'une organisation. Nous décrivons ainsi ce que signifie pour un agent de jouer un rôle au sein d'une organisation, quelles sont les contraintes qui en dérivent et comment le concept (maintenant formel) d'autonomie peut quantifier ces significations pour faire le choix. Nous présentons ensuite l'utilisation de cette approche dans le cadre d'une application d'intelligence ambiante que nous avons implémentée, où l'agent personnel d'un utilisateur s'exécute sur un téléphone portable et est amené à choisir entre des rôles possibles afin de négocier des contrats de publicité.

I. Etat de l’art

Introduction

L'objectif de cette thèse est de proposer un modèle de raisonnement qui permettra aux agents d'évaluer les implications de leurs actions sur leur prise de décision ou leur comportement. Les agents doivent être ainsi capables d'identifier les contraintes qui limitent leur prise de décision ou leur comportement et de raisonner comment ces contraintes se modifient en fonction de leurs actions et choix. Afin de s'intéresser aux différents modèles de raisonnement existants et à leur applicabilité dans notre travail, nous devons premièrement identifier les types de contraintes possibles et comment ils limitent la prise de décision ou le comportement d'un agent. Un agent est généralement limité dans ses choix par des facteurs de provenance interne (p.ex. : le manque de connaissances, etc.), mais aussi de provenance externe, comme par exemple le manque de ressources dans l'environnement, les interactions avec d'autres agents ou bien l'organisation à laquelle il appartient. Comme nous nous intéressons surtout au raisonnement social, les deux derniers types de contraintes nous intéressent particulièrement : comment les interactions avec d'autres agents et l'organisation à laquelle il appartient limitent la prise de décision ou le comportement d'un agent et comment il peut raisonner dessus.

Malgré le manque d'une définition communément admise dans la communauté multi-agents, la notion d'*autonomie* semble adresser les mêmes problèmes que ceux qui nous intéressent. Dans le Chapitre 1 nous analysons de plus près cette notion pour identifier plusieurs formes d'autonomie, parmi lesquelles une très pertinente pour notre travail : l'autonomie sociale en délibération. Nous passons ainsi en revue les approches existantes dans la littérature concernant cette forme d'autonomie, approches qui s'intéressent à l'autonomie sociale d'un agent par rapport à un autre agent (dans les interactions) ou par rapport à une norme (dans des organisations). Comme nous le verrons dans ce Chapitre, selon la perspective considérée, l'autonomie sociale en délibération d'un agent représente la liberté qu'il a pour prendre une décision (sans être contraint par une entité externe) ou son comportement autonome (qui ne respecte pas les contraintes qui lui sont imposées). L'autonomie d'un agent est ainsi limitée par des contraintes externes, contraintes qui sont généralement issues du contrôle (ou des tentatives de contrôle) que d'autres agents ou l'organisation essaient d'exercer sur l'agent. Comme nous allons l'argumenter, ce contrôle représente au niveau de chaque agent le besoin de coordination existant dans le système – l'autonomie des agents est donc limitée par les contraintes issues de la coordination existante dans le système multi-agents.

Nous sommes donc amenés à analyser les différents types de coordination possibles, afin d'identifier les types de contraintes qui limitent l'autonomie d'un agent et sur lesquels l'agent pourra raisonner. Dans le Chapitre 2 nous nous intéressons à la coordination et aux contraintes institutionnelles, coordination qui utilise des concepts comme par exemple les rôles, les normes ou les organisations multi-agents. Cette approche institutionnelle offre aux agents un cadre formel et explicite pour coordonner leurs activités, en décrivant pour chacun d'entre eux le comportement nécessaire pour atteindre la coordination dans le système et en les stimulant de suivre ce comportement. Comme nous le verrons, le concept de *norme* n'est pas forcément institutionnel, mais il est souvent associé au concept de *rôle* dans une organisation/institution. Généralement les agents doivent jouer des rôles,

rôles qui spécifient le comportement attendu de leur part par l'organisation (par exemple, en imposant des buts à satisfaire et des normes à obéir) ; les institutions multi-agents utilisent aussi des mécanismes pour motiver les agents à ne pas dévier de ce comportement spécifié. Le fait de jouer un rôle et les spécifications de ce rôle constituent ainsi des contraintes institutionnelles qui limitent la prise de décision et le comportement des agents.

D'autres contraintes limitent aussi les agents dans leur prise de décision ou leur comportement, des contraintes issues non des institutions, mais des interactions entre les agents. Dans le Chapitre 3, nous nous intéressons à ces contraintes (que nous appelons interpersonnelles), contraintes que nous pouvons diviser dans deux catégories : celles qui poussent les agents à interagir avec d'autres et celles qui résultent des interactions menées avec d'autres agents. En ce qui concerne la deuxième catégorie, il semble que l'approche la plus utilisée dans la littérature est d'utiliser le modèle des *engagements sociaux* : suite à une interaction un agent peut devenir engagé envers un autre agent pour effectuer une action ou satisfaire un but. Ceci constitue une contrainte sur son comportement qui doit maintenant être orienté envers la satisfaction de cet engagement, dans le cas contraire étant considéré autonome (et comme nous le verrons il peut subir des conséquences). Concernant les contraintes qui poussent les agents à interagir les uns avec les autres, nous décrivons la *théorie de la dépendance* : quand un agent n'est pas capable de satisfaire tout seul son but, il dépend des autres agents pour l'aider à satisfaire ce but et il doit donc interagir avec eux afin d'obtenir l'aide nécessaire. Nous décrivons aussi dans ce Chapitre la *théorie du pouvoir social* qui étend la théorie de la dépendance en s'intéressant aussi au pouvoir qu'un agent a d'influencer un autre pour l'aider et qui, comme nous le verrons, nous offre des pistes pour la représentation des contraintes à la fois interpersonnelles, mais aussi institutionnelles.

Après avoir analysé plusieurs modèles de contraintes institutionnelles et interpersonnelles et avoir identifié comment ces contraintes influencent l'autonomie de décision et le comportement autonome des agents, dans le Chapitre 4 nous nous intéressons au raisonnement que les agents peuvent effectuer sur ces contraintes. Dans ce but, nous analysons quelques modèles et architectures d'agents proposés dans la littérature, ce qui nous permettra d'identifier plusieurs types de raisonnement effectués par les agents. Comme nous le verrons, parmi ces types, deux sont particulièrement intéressants pour notre travail, types que nous appelons respectivement *raisonnement a priori* et *raisonnement a posteriori*. Le premier type est utilisé par les agents par exemple pour choisir un partenaire d'interaction (avec qui il a le plus d'autonomie) ou pour choisir parmi plusieurs rôles à jouer dans une organisation (celui qui lui laisse le plus d'autonomie), tandis que le deuxième est utilisé pour décider d'obéir ou pas à une contrainte établie (d'être ou pas autonome). Nous finirons cette partie en synthétisant et en identifiant les exigences pour obtenir un raisonnement à base d'autonomie dans les systèmes multi-agents.

1. Autonomie dans les systèmes multi-agents

Un des objectifs de ce travail est de proposer une modélisation des contraintes issues des interactions entre agents ou des organisations, contraintes qui limitent le choix de comportement ou la prise de décision des agents. Pour des raisons qui deviendront plus claires dans ce chapitre, notre approche base cette modélisation sur la notion d'autonomie. Malheureusement, cette notion n'a pas de définition communément admise dans la communauté scientifique de systèmes multi-agents. Dans ce chapitre nous identifions des propriétés de la notion d'autonomie, propriétés qui seront ensuite utilisées pour définir plusieurs types d'autonomie et expliquer les différences entre ces types. Parmi les types d'autonomie identifiés, dans ce manuscrit nous nous intéressons surtout à l'autonomie sociale concernant la prise de décision d'un agent, autonomie qui sera analysée par la suite. Notre argumentaire continuera avec une analyse de la relation existante entre la notion d'autonomie et les notions de contrôle et coordination dans les systèmes multi-agents, ce qui nous permettra d'identifier plusieurs types de contraintes sociales qui limitent l'autonomie décisionnelle d'un agent.

1.1. Définitions de l'autonomie dans les systèmes multi-agents

Bien que la notion d'autonomie apparaisse dans la plupart des définitions d'agent, jusqu'à ce jour, il n'existe pas de définition communément admise de ce qui apparaît être la caractéristique fondamentale d'un agent dans les systèmes multi-agents. Une quantité importante de travaux dans la communauté existe afin d'essayer de la clarifier et de la formaliser (voir par exemple (Hexmoor *et al.* 2003)). Au manque de définition unique vient s'ajouter le fait que les définitions proposées semblent faire référence à des concepts totalement différents. Dans la suite, nous présentons quelques propriétés du concept d'autonomie, ce qui nous permettra d'identifier plusieurs types d'autonomie considérés dans la communauté multi-agents et nous analysons les principales définitions des types d'autonomie qui ont de l'intérêt pour ce travail.

1.1.1. Propriétés de l'autonomie

Malgré les différentes définitions de l'autonomie existantes dans la littérature, ce concept semble avoir quelques propriétés communément admises, telle que son caractère relationnel et sa dépendance du contexte. Nous présentons ces propriétés, ce qui nous permettra ensuite d'identifier plusieurs types d'autonomie et leurs définitions associées.

(1) Autonomie – une notion relationnelle

Nous suivons dans ce travail le point de vue de Castelfranchi (Castelfranchi 1995) qui considère que l'autonomie ne peut être considérée qu'en termes relationnels : *l'autonomie d'une entité doit être définie en relation avec ou par rapport à une autre entité*. Il ne suffit donc pas de dire si un agent est ou n'est pas autonome, mais il faut aussi préciser par rapport à qui et pour quoi il l'est ou ne l'est pas. L'autonomie d'un agent doit être donc définie par rapport à une autre entité, telle qu'un autre agent, l'utilisateur, le concepteur de l'agent, le système, etc. Nous verrons par la suite comment différents types d'autonomie peuvent être considérés en fonction de l'entité qui influence l'autonomie d'un agent.

L'autonomie d'un agent a aussi un *objet pour lequel l'agent est ou n'est pas autonome* et qui peut être tout élément cognitif définissant l'agent, par exemple un but à satisfaire, un plan à exécuter, une action à effectuer, etc. Dans (Castelfranchi 2000) les auteurs précisent que l'objet de l'autonomie est strictement lié à l'architecture d'un agent : par exemple, pour un agent cognitif qui choisit ses buts et qui planifie pour les atteindre, nous pouvons parler de l'autonomie pour planifier, l'autonomie pour choisir des buts, pour raisonner, etc. Pour un agent situé dans un environnement, nous pouvons parler de l'autonomie pour acquérir de l'information, l'autonomie pour agir, l'autonomie pour des ressources, etc.

(2) Autonomie – une notion située ou contextualisée

Cette propriété relationnelle explique partiellement la multitude de définitions existantes, qui peuvent faire référence à des objets ou entités différents pour lesquels et par rapport à qui un agent peut avoir de l'autonomie. Cependant, Hexmoor (Hexmoor 2000) considère que cette propriété relationnelle de l'autonomie ne suffit pas quand l'autonomie d'un agent est analysée, mais le *contexte* de l'autonomie doit être aussi pris en compte. Un agent peut être autonome dans une situation par rapport à une autre entité et pour un objet et ne pas l'être dans une autre situation (par rapport à la même entité et le même objet). Nous ne pouvons ainsi jamais parler d'une autonomie absolue d'un agent, mais de son autonomie relative à une autre entité, pour quelque chose et dans une situation. Un exemple de type de situation qui influence l'autonomie d'un agent est le rôle qu'il joue dans un groupe : comme nous le verrons dans ce manuscrit, son autonomie change beaucoup en fonction du rôle joué par l'agent ou s'il ne joue pas de rôle.

(3) Degrés d'autonomie

Si un agent n'a pas d'autonomie absolue, les auteurs de (Brainov et Hexmoor 2001) considèrent qu'il faut parler du *degré d'autonomie d'un agent*. En considérant la même entité qui influence l'autonomie d'un agent (par exemple, un autre agent) et un type d'objet de l'autonomie (par exemple, l'adoption des buts), nous pouvons ainsi caractériser différentes situations en fonction du degré d'autonomie qu'elles laissent à l'agent. Par exemple, si dans une situation un agent est autonome par rapport à un autre pour l'adoption de quelques buts et dans une autre situation l'agent est autonome par rapport au même agent pour l'adoption de plus de buts, il a plus d'autonomie dans la deuxième situation. Comme nous le verrons dans ce manuscrit, le degré d'autonomie d'un agent (et donc la différenciation entre plusieurs situations) est très important pour l'évaluation de situations possibles en fonction de l'autonomie laissée à l'agent, ce qui constitue l'objectif de ce travail.

1.1.2. Différents types d'autonomie dans les systèmes multi-agents

Comme l'autonomie d'un agent est relative à l'entité qui l'influence, à son objet et au contexte, différents types d'autonomie peuvent être identifiés en fonction de ces paramètres. Par exemple, Castelfranchi (Castelfranchi 1995) fait la différence entre l'*autonomie en exécution* et *l'autonomie en motivation* (pour des buts). La première forme est le cas classique d'un agent/robot qui obéit à des ordres : l'utilisateur délègue un but au robot et le robot se charge de trouver et d'exécuter un plan pour satisfaire le but, c.-à-d., d'exécuter l'ordre de l'utilisateur. Dans le deuxième cas, l'agent a des

motivations individuelles et il est capable de générer des buts tout seul, sans la demande explicite de l'utilisateur.

Une forme d'autonomie qui se trouve à un niveau intermédiaire entre l'autonomie en exécution et celle en motivation est celle que Castelfranchi (Castelfranchi 1998) appèle *autonomie en délégation*. De manière générale, quand un agent a un but et pour une raison quelconque il ne veut ou ne peut pas le satisfaire, il demande à un autre agent de le satisfaire à sa place, une *délégation* de but du premier agent vers le deuxième est mise en place. De l'autre coté, si un agent décide et essaye de satisfaire un but en considérant que c'est le but d'un autre agent, une *adoption* de but a été effectuée. Comme nous le verrons par la suite, une délégation d'un but n'est pas toujours suivie par une adoption et la raison pour cela réside par exemple dans l'autonomie en motivation des agents. Cependant, pour l'instant considérons le cas d'un agent qui veut déléguer un but à un autre agent. Il peut lui déléguer le but même et laisser l'autre agent trouver et exécuter un plan pour le satisfaire, il peut lui déléguer le but et un plan précis à exécuter ou il peut lui déléguer les actions du plan une après une sans lui communiquer le plan général ou le but final à atteindre. La différence entre ces possibilités est faite en général par la confiance que l'agent qui délègue a dans les capacités de planification ou d'exécution de l'agent qui adopte ou dans sa possibilité de le contrôler pendant la satisfaction du but/exécution du plan. Castelfranchi appèle ce phénomène *l'autonomie en délégation* : c'est l'autonomie d'exécution/planification qu'un agent laisse à un autre quand il lui délègue un but.

Autonomie de décision

L'autonomie en exécution est une propriété désirée pour tous les agents artificiels qui doivent être capables d'exécuter les demandes de leurs utilisateurs. Par contre, l'autonomie en motivation est parfois nécessaire, mais comme nous le verrons par la suite, son existence peut créer des situations non-désirées dans les systèmes multi-agents. Dans ce travail nous nous intéressons à une forme de cette autonomie, *l'autonomie en délibération* (ou *de décision*) d'un agent – l'autonomie d'un agent par rapport à une autre entité dans un contexte de prise de décisions (délibération) concernant quelque chose. En fonction de la nature de l'entité qui influence cette autonomie, nous pouvons parler d'autonomie par rapport au concepteur, par rapport aux stimuli (à l'environnement), par rapport à l'utilisateur ou bien d'autonomie sociale (par rapport à un autre agent ou à une norme). Dans la suite nous présentons brièvement ces types d'autonomie pour analyser ensuite le dernier type d'autonomie qui nous intéresse particulièrement dans ce travail *l'autonomie sociale de décision (en déliberation).*

- *L'autonomie de décision par rapport au concepteur* se réfère à la capacité d'un agent de prendre des décisions différentes de celles désirées par son concepteur. C'est un type d'autonomie rarement pris en considération à cause des inconvénients qu'elle peut provoquer : il n'y a aucune modalité de prédire ou contrôler un agent qui ne respecte même pas les spécifications de conception.

- Comme argumenté dans (Castelfranchi 1995), *l'autonomie par rapport aux stimuli ou à l'environnement* est une caractéristique de tous les êtres humains et une caractéristique désirée des agents artificiels. Cette forme d'autonomie décrit « le problème de Descartes » : les réponses des humains à l'environnement ne sont pas provoquées, ni indépendantes des stimuli externes. C'est-à-dire, le comportement d'un agent ne doit pas être contrôlé par son environnement, mais il doit être dirigé vers la satisfaction de ses buts (ou d'autres motivations internes) en prenant en compte les

stimuli externes. Nous pouvons ainsi préciser que tout agent artificiel capable de satisfaire ses buts (ou ceux de son utilisateur/concepteur) est doté de cette forme d'autonomie.

- *L'autonomie par rapport à l'utilisateur* représente peut-être le sens initial du terme dans les premières définitions d'agents autonomes : c'est l'autonomie d'un agent qui interagit avec l'utilisateur et qui doit choisir son comportement. Par exemple, dans (Bonnet and Tessier, 2007), les auteurs s'intéressent à la conception des constellations de satellites qui collaborent pour mieux observer la Terre. Ces satellites doivent avoir une autonomie par rapport à l'utilisateur pour prendre des décisions sans le contrôle des êtres humains, mais aussi une autonomie en exécution pour pouvoir créer et exécuter des plans qui satisfont leurs objectifs (ou ceux de leurs utilisateurs). En général, cette autonomie par rapport à l'utilisateur d'un agent peut varier en fonction de la situation : parfois l'agent prend des décisions tout seul, parfois c'est l'utilisateur qui prend ces décisions. Le passage du contrôle sur son comportement entre un agent et son utilisateur (ou vice-versa) est appelé dans la littérature *autonomie adaptable*. Par exemple, dans le projet *e-Elves* (Tambe *et al.* 2002), les auteurs. s'intéressent à comment un agent peut apprendre à adapter son autonomie en fonction du problème à résoudre : certaines décisions sont trop importantes pour être prises par l'agent et l'utilisateur doit les prendre, mais parfois l'utilisateur n'est pas joignable ou il prend trop de temps à décider, ce qui menace la coordination entre les agents dans le système. Un agent doit donc décider quand il faut adapter son autonomie de décision en laissant l'utilisateur prendre les décisions ou en les prenant tout seul. *L'autonomie adaptable* pose des problèmes de recherche intéressants et pas encore résolus ; cependant dans ce travail nous ne considérons pas les interactions des agents avec leur(s) utilisateur(s) et nous ne nous intéressons donc pas à cette forme d'autonomie.

- Le type d'autonomie auquel ce travail s'intéresse est *l'autonomie sociale* – l'autonomie de décision d'un agent dans un contexte social, c.-à-d., par rapport à d'autres agents ou par rapport à des entités sociales telles que des organisations, systèmes normatifs, etc. Comme nous le verrons par la suite, l'autonomie sociale est une caractéristique importante d'un agent au sein d'un système multi-agents dans lequel il interagit avec d'autres agents et dans lequel son comportement est limité par des structures organisationnelles ou normatives.

1.1.3. Autonomie sociale dans les interactions

L'autonomie d'un agent par rapport à un autre agent est une des plus fréquentes formes d'autonomie considérées dans les systèmes multi-agents et a une grande importance pour les interactions entre agents. Cette forme d'autonomie est souvent considérée liée à la notion d'indépendance de décision de l'agent. Par exemple, Castelfranchi (Castelfranchi 2000) considère qu'un agent est autonome par rapport à un autre agent pour un de ses buts s'il ne dépend pas de l'autre agent pour la satisfaction de ce but. Sans vouloir donner ici une définition formelle de la notion de dépendance (la théorie de la dépendance sera décrite dans le Chapitre 3), d'une manière informelle, nous pouvons considérer qu'un agent dépend d'un autre pour la satisfaction d'un but s'il ne peut pas satisfaire le but tout seul et il a besoin de l'aide de l'autre agent. Les décisions prises par le premier agent concernant la satisfaction du but doivent prendre en compte cette dépendance et varient en fonction du comportement du deuxième agent. Nous pouvons ainsi dire que le premier agent n'est plus libre de décider comme il veut en ce qui concerne son but parce qu'il dépend d'un autre agent pour satisfaire ce but.

Même si nous avons utilisé l'exemple de la satisfaction d'un but, l'autonomie sociale vue comme une indépendance peut avoir comme objet tout élément cognitif d'un agent. En fonction du modèle d'agent considéré, nous pouvons parler de dépendance pour planifier (un agent a besoin de l'aide d'un autre pour trouver un plan), dépendance pour agir (un agent a besoin de l'aide d'un autre pour exécuter une action), dépendance pour la perception (un agent ne peut pas percevoir l'environnement et a besoin de l'aide d'un autre), etc., et toutes ces dépendances possibles sont liées à des autonomies possibles d'un agent. Cependant, en général un agent est considéré capable de percevoir et d'agir sur son environnement, capable de former des buts et des plans associés à partir de ses motivations, mais pas toujours capable de satisfaire des buts tout seul. C'est la raison pour laquelle l'autonomie sociale dans les interactions est souvent liée aux buts des agents.

Considérer l'autonomie d'un agent par rapport à un autre comme une indépendance facilite beaucoup le travail des concepteurs de systèmes multi-agents : une théorie formelle de la dépendance existe (et sera présentée dans le Chapitre 3), donc nous avons accès à une définition formelle de l'autonomie sociale. Malheureusement, plusieurs auteurs argumentent que la notion d'autonomie sociale inclut la notion d'indépendance sans être limitée à celle-ci. Par exemple, Castelfranchi (Castelfranchi 1995) définit un agent avec une « vraie » autonomie (pas seulement indépendance) comme un agent qui a ses propres buts, qui peut choisir entre plusieurs buts et qui est capable d'adopter des buts des autres agents, mais c'est l'agent même qui décide d'adopter ou non un but en fonction de ses propres buts. Pour être considéré comme autonome, un agent social doit donc prendre les décisions liées aux interactions seulement en fonction de ses propres buts : il adopte seulement les buts qui l'aident à satisfaire les siens.

Cette vision sur l'autonomie est partagée par Luck et d'Inverno (Luck and d'Inverno 2001) qui considèrent qu'un agent est autonome s'il a des motivations. Ils proposent un modèle formel d'un système multi-agents, dans lequel l'autonomie est définie formellement comme l'existence des motivations, qui sont des éléments cognitifs utilisés pour générer des buts. Un agent avec des motivations génère les buts qu'il va poursuivre à partir de ses motivations. Dans l'éventualité où un autre agent lui délègue un but, l'agent autonome décide d'adopter ou pas ce but en fonction de sa relation avec ses motivations. Si le but à adopter satisfait ses motivations, alors il l'adopte, s'il est contraire à ses motivations, alors il le refuse.

Synthèse

Ces deux dernières définitions de l'autonomie sociale la définissent comme une caractéristique de la prise de décision d'un agent : un agent est autonome s'il adopte seulement les buts en concordance avec ses propres buts ou motivations. Comme en général il est impossible pour un observateur externe à un agent de savoir quels sont les buts/motivations de l'agent ou comment l'agent prend ses décisions, il est difficile d'établir le caractère autonome d'un agent. Même si Luck et d'Inverno (Luck and d'Inverno 2000) admettent qu'une différence existe entre l'autonomie sociale et l'indépendance, ils prennent un point de vue pragmatique et argumentent que la distinction entre les deux notions n'est pas intéressante dans la pratique et que la notion d'indépendance peut être considérée comme une approximation suffisante de la notion d'autonomie. Il est intéressant de noter que l'indépendance est un concept externe à un agent, il décrit une relation entre deux agents, tandis que la prise de décision, l'existence des motivations, etc., sont des concepts internes à un agent. Comme un observateur externe

n'a pas généralement accès à la perspective interne d'un agent, mais seulement à celle externe, l'approximation de l'autonomie comme une indépendance (donc point de vue externe) peut représenter une solution pratique. Nous analyserons plus tard dans ce chapitre la distinction entre l'autonomie comme une caractéristique d'un processus interne à l'agent ou comme une relation entre agents, mais avant, nous analysons un autre type d'autonomie sociale : celle d'un agent par rapport à des normes.

1.1.4. Autonomie sociale dans les systèmes à base de normes

Il est communément admis (Dignum 1999) que l'utilisation d'autonomie introduit un degré de non-déterminisme dans le comportement du système, puisque la réponse d'un agent autonome dans une interaction n'est pas connue a priori. Les concepteurs des systèmes multi-agents ouverts et hétérogènes (avec des agents potentiellement différents) ne peuvent donc pas prédire le comportement des agents dans un système. Souvent, pour assurer un comportement global cohérent malgré l'autonomie locale, des normes , des conventions (Tuomela, 1995), des structures organisationnelles (Hubner 2003) sont utilisées. Nous présenterons ces concepts dans le Chapitre 2, mais pour l'instant nous pouvons considérer, dans son sens le plus large, une norme comme une obligation pour un ensemble d'agents d'effectuer ou de ne pas effectuer une action.

Sans normes, un agent autonome décide d'adopter ou non un but d'un autre agent en fonction seulement de ses propres buts. L'existence des normes modifie la prise de décision de l'agent : il peut y avoir des normes qui l'obligent à adopter un but d'un autre agent même si ce but ne correspond pas à ses propres buts ou des normes qui lui interdisent de poursuivre des buts. L'autonomie sociale d'un agent par rapport à un autre agent est donc fortement influencée par les normes existantes dans le système. Des conflits entre les normes sont néanmoins possibles, comme par exemple une norme qui oblige un agent à adopter un but délégué par un autre agent et une norme qui lui interdit de poursuivre ce but. Les agents doivent être donc capables de raisonner sur les normes qui limitent leur autonomie de décision et peut-être de décider de désobéir à de telles normes.

Les agents capables de décider d'obéir ou non à une norme sont appelés des *agents autonomes par rapport aux normes*. Conte *et al.* (Conte *et al.* 1999) identifient plusieurs cas d'autonomie d'un agent par rapport aux normes. Une norme étant un concept du niveau système, donc initialement externe aux agents, les agents doivent adopter les normes, c.-à-d., les reconnaître comme étant des normes et accepter le fait qu'ils sont concernés par elles. *Un agent est autonome par rapport à l'adoption d'une norme s'il n'adopte pas cette norme soit parce qu'il ne la reconnaît pas comme norme ou parce qu'il considère que la norme ne le concerne pas.* Même si l'agent adopte une norme, un agent doit décider de l'obéir (d'avoir le comportement spécifié par la norme) ou de la violer (d'avoir un comportement différent). *Un agent qui décide de violer une norme est considéré autonome par rapport à l'obéissance à une norme*. Finalement, *un agent autonome par rapport aux normes peut prendre l'initiative de proposer ou de modifier des normes* et peut veiller sur la satisfaction des normes par d'autres agents en évaluant leur comportement et en imposant des sanctions si nécessaire.

Synthèse

Dans cette section nous avons identifié quelques propriétés de l'autonomie d'un agent, ce qui nous a permis d'identifier plusieurs types d'autonomie. Parmi eux, dans ce manuscrit nous nous intéressons à

l'autonomie sociale de décision (en délibération) : l'autonomie d'un agent de prendre des décisions dans un contexte social. Ce choix est motivé par un des objectifs de ce travail, celui de permettre à un agent de prendre des décisions pertinentes quand son contexte social change, comme par exemple quand il change de rôle. L'autonomie sociale d'un agent peut être prise en considération dans les interactions avec d'autres agents ou dans un contexte normatif, donc par rapport aux normes. Dans la suite nous synthétisons les définitions présentées auparavant pour distinguer deux perspectives possibles sur l'autonomie sociale et montrer l'importance de cette distinction pour notre travail.

1.2. Deux perspectives sur l'autonomie

Comme nous l'avons vu, les définitions de l'autonomie utilisées dans les systèmes multi-agents tournent autour de quelques éléments communs. Les différences existantes entre ces définitions peuvent être partiellement expliquées par la propriété relationnelle de l'autonomie : des définitions différentes peuvent être obtenues en fonction du référent (par rapport à qui) et de l'objet (pour quoi) un agent est autonome. Par exemple, l'autonomie sociale par rapport à un autre agent (interactions) ou par rapport aux normes (organisations) n'est pas la même que l'autonomie par rapport à l'utilisateur. Cependant, nous pensons que, même dans le cas seulement de l'autonomie sociale, des différences fondamentales existent entre les définitions existantes. Nous partageons le point de vue de (Beavers and Hexmoor 2004) qui considèrent qu'il existe deux perspectives différentes sur l'autonomie et qui utilisent deux expressions pour souligner cette différence : « *avoir de l'autonomie* » et « *être autonome* »[1]. Ces deux perspectives caractérisent des aspects différents d'un agent : son processus de décision (vision interne) ou son comportement (vision externe).

1.2.1. *Etre autonome*

Etre autonome est une caractéristique du comportement de l'agent : un agent est considéré comme autonome quand il a un comportement autonome. Cette caractéristique du comportement d'un agent garde le caractère relationnel de l'autonomie : un agent a un comportement autonome par rapport à une autre entité (autre agent, système normatif, etc.) et pour quelque chose (l'adoption d'un but, l'obéissance à une norme, etc.). Les différentes définitions présentées dans la section précédente utilisent souvent cette perspective.

Pour Luck et d'Inverno (Luck and d'Inverno 2001), un agent *est autonome* par rapport à un autre agent pour l'adoption des buts s'il n'adopte que les buts satisfaisant ses motivations. Pour un observateur externe à l'agent, qui n'a pas accès aux motivations de l'agent ou à sa prise de décision, la seule modalité d'évaluer l'autonomie d'un agent est de se baser sur la réponse de l'agent à une demande d'adoption. Si un agent refuse une demande d'adoption d'un but, son comportement est considéré comme autonome (il a refusé parce que le but ne satisfaisait pas les motivations). Par contre, s'il adopte un but d'un autre agent, un observateur externe ne peut pas caractériser l'autonomie du comportement de l'agent : soit il l'a adopté parce qu'il satisfaisait ses motivations (et donc il est autonome selon Luck et d'Inverno), soit il n'a pas de motivations et adopte tous les buts (et donc il n'est pas autonome selon Luck et d'Inverno). La même perspective est prise par Conte *et al.* (Conte *et*

[1] Termes utilisés dans une conversation entre l'auteur et Dr. Hexmoor au Workshop « Computational Autonomy », workshop associé à la Conférence AAMAS'03, Melbourne, Australia.

al. 1999) qui considèrent qu'un agent a un comportement autonome s'il n'adopte pas une norme ou s'il viole une norme qu'il a adopté. Pour résumer, généralement:

> *Un agent est considéré autonome par rapport à une entité (autre agent, norme) quand son comportement diffère du comportement spécifié (attendu) par cette entité.*

Cette caractéristique du comportement d'un agent est très utile pour un observateur externe à l'agent. Un tel observateur n'a pas accès au processus de décision de l'agent ou a ses éléments cognitifs : il peut caractériser l'agent seulement en se basant sur son comportement (actions effectuées, messages envoyés, etc.). Parce qu'elle dénote un point de vue externe à l'agent, nous avons appelé cette perspective dans un article précédent, une *perspective externe* sur l'autonomie d'un agent (Carabelea *et al.* 2004). Comme nous le verrons dans le Chapitre 2, cette perspective sur l'autonomie a une grande importance dans les institutions multi-agents où les violations des normes sont punies : l'institution doit identifier le comportement autonome d'un agent et agir en conséquence (par exemple, en imposant des sanctions).

1.2.2. *Avoir de l'autonomie*

Avoir de l'autonomie (en délibération) est une caractéristique du processus de décision de l'agent : un agent est considéré avoir de l'autonomie quand son processus de décision est autonome. Certains auteurs considèrent que la prise de décision d'un agent est autonome si elle est réalisée seulement à partir de *ses* propres buts (Castelfranchi 1995) ou motivations (Luck and d'Inverno 2001), ou si elle ne dépend pas du comportement d'un autre agent (Castelfranchi 2000). En ce qui concerne l'autonomie par rapport aux normes, elle existe simplement si l'agent peut prendre des décisions par rapport aux normes (Conte *et al.* 1999). D'autres auteurs, comme Hexmoor (Hexmoor 2000), considèrent qu'avoir de l'autonomie est équivalent au fait que l'agent est libre de décider (il a du libre-arbitre).

Un observateur externe à un agent ne peut pas savoir si l'agent prend des décisions ou si ses réponses sont codées en dur, ni quels sont les facteurs pris en compte dans le processus de décisions (seulement les propres buts, des relations de dépendance, etc.). Pour analyser avec cette perspective l'autonomie d'un agent, nous devons prendre un point de vue interne à l'agent. Nous avons appelé cette perspective, une *perspective interne* sur l'autonomie d'un agent (Carabelea *et al.* 2004).

La première impression est que ce type d'autonomie n'est pas désirable dans les systèmes multi-agents. Si un agent peut décider d'adopter ou pas un but ou d'obéir ou pas à une norme, il existe le risque qu'il prenne des décisions contraires à l'intérêt du système. Le système fonctionnerait donc mieux avec des agents sans autonomie. Ceci n'est pas vrai, comme plusieurs auteurs l'ont mis en évidence. Luck et d'Inverno (Luck and d'Inverno 2001) considèrent qu'un agent sans autonomie sera utilisé comme esclave par les autres agents parce qu'il ne refusera pas leurs demandes d'adoptions de buts. Conte *et al.* (Conte *et al.* 1999) incluent dans leur définition de l'autonomie par rapport aux normes la possibilité d'un agent de proposer ou de modifier des normes. Verhagen (Verhagen 2000) montre l'utilité des agents avec ce type d'autonomie dans l'obtention d'une coordination globale parce qu'ils sont capables de modifier les normes existantes et de les imposer aux autres agents. Pour modifier les normes existantes, un agent doit raisonner sur elles, sur leurs conséquences et sur

comment leur changement améliora les performances du système – autrement dit, il doit avoir une autonomie de décision par rapport à ces normes.

Cette perspective sur l'autonomie s'intéresse à la prise de décision d'un agent pour voir si l'agent peut ou ne peut pas prendre une décision. Si cette prise de décision est impossible (l'agent ne sait pas décider, son architecture interne ne le lui permet pas), l'agent n'a pas d'autonomie pour cette prise de décision. Si elle est interdite, pas libre, imposée de l'extérieur, alors l'agent n'a pas d'autonomie par rapport à l'entité qui lui impose une décision. Pour résumer:

> *Un agent est considéré avoir de l'autonomie par rapport à une entité (autre agent, norme) pour prendre une décision quand sa prise de décision n'est pas influencée par cette entité.*

1.3. Autonomie – contrôle – coordination

Les différentes définitions et perspectives sur l'autonomie sociale présentées ci-dessus mettent en évidence l'importance pour les agents d'avoir de l'autonomie, surtout si, comme la plupart des auteurs, nous considérons l'autonomie en délibération comme la liberté d'un agent de prendre des décisions. Il est ainsi dans l'intérêt de l'agent d'avoir le plus d'autonomie, donc de prendre les décisions qui l'avantagent le plus, mais du point de vue d'un système l'autonomie des agents ne présente pas que des avantages, mais aussi des inconvénients. D'un coté, le comportement autonome des agents peut présenter des inconvénients majeurs pour le concepteur d'un système multi-agents : les agents autonomes peuvent ne pas respecter les normes et peuvent décider d'agir d'une manière pas connue à priori, ne permettant donc pas au système d'atteindre ses objectifs. D'un autre coté, l'autonomie des agents peut améliorer le comportement du système parce que les agents peuvent utiliser leurs capacités délibératives et prendre les décisions les plus adaptées à une situation, comme dans le cas des agents autonomes qui proposent des normes.

1.3.1. Relations entre coordination, contrôle et autonomie

Coordination

Dans notre travail nous suivons le point de vue de Boissier (Boissier 2003) qui considère que le problème principal du concepteur d'un système multi-agents est de s'assurer que le système atteint ses objectifs, ce qui implique le besoin d'une *coordination* entre les comportements des agents qui en font partie. En règle générale, la coordination entre les parties d'un système est nécessaire pour éviter des conflits (utilisation des mêmes ressources) ou pour synchroniser les comportements de ces parties. Le terme coordination est utilisé dans ce manuscrit avec un sens similaire à celui *d'ordre social* de Castelfranchi (Castelfranchi 2005). D'un point de vue externe au système, le système est une entité qui manifeste un comportement orienté vers la satisfaction des buts, par exemple, une équipe d'agents footballeurs est un système multi-agents qui a le but de gagner un match. Pour atteindre ces buts, une coordination est nécessaire au sein du système, coordination pas forcément visible de l'extérieur. Cette coordination est *interne au système, mais externe aux agents* et son obtention représente un des

problèmes fondamentaux pour la communauté multi-agents : comment s'assurer qu'au niveau *global* une coordination se manifeste quand les prises de décision sont *locales*, surtout dans les systèmes ouverts où peuvent agir des agents conçus différemment.

De coordination à contrôle

Par définition, dans un système multi-agents la prise de décision est distribuée entre les agents qui décident chacun sur son comportement local. La propriété globale désirée, la coordination, doit être ainsi traduite au niveau local des agents, agents qui sont concernés par cette propriété. Ceci est réalisé grâce à la notion de *contrôle* que nous utilisons dans ce manuscrit dans le sens d'une influence externe à un agent et exercée sur le comportement de cet agent en réponse à son comportement. Le besoin de coordination globale dans un système nécessite donc un contrôle exercé sur le comportement local de chaque agent. Dans l'exemple d'une équipe d'agent footballeurs, une coordination doit exister entre les agents, par exemple pour passer le ballon, pour se positionner sur le terrain, etc. Cette coordination implique un contrôle sur le comportement de chaque agent, contrôle qui doit assurer qu'un agent passe le ballon quand il faut ou que chaque agent se positionne à sa place. Ce contrôle sur le comportement d'un agent peut prendre plusieurs formes, comme par exemple influencer sa prise de décision ou évaluer son comportement et lui fournir du feedback – renforcement d'un comportement désiré.

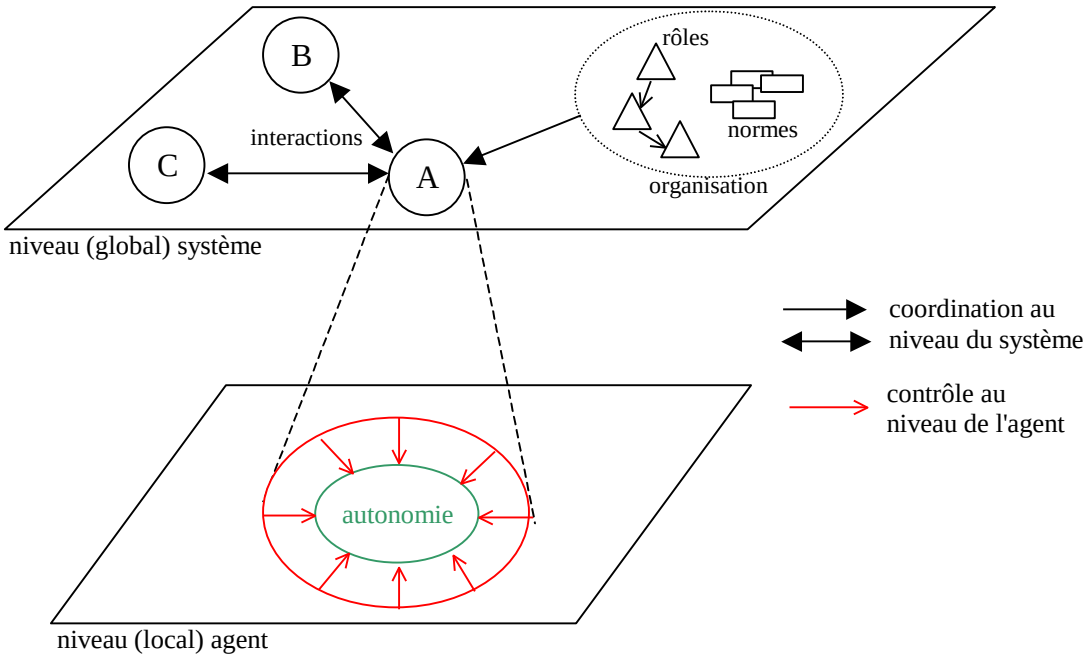


Figure 1.1. Relation coordination – contrôle – autonomie

De contrôle à autonomie

Cette notion de contrôle que nous utilisons ci-dessus est le complément de la notion d'*autonomie* présentée auparavant : plus un agent est contrôlé par une entité externe, moins il a d'autonomie par rapport à cette entité. Si le contrôle prend par exemple la forme d'une influence de la prise de décision d'un agent, il est lié à l'autonomie en décision de cet agent. Le problème revient donc à trouver l'équilibre entre le contrôle exercé sur les agents et leur autonomie. Pour utiliser le même exemple

qu'auparavant, les agents footballeur doivent être contrôlés pour s'assurer qu'ils occupent les bonnes positions sur le terrain, mais ils doivent aussi avoir un degré d'autonomie pour pouvoir décider de changer de position si une opportunité apparaît. Dans un cas idéal il est possible de spécifier un contrôle qui dit aux agents de changer de position – cependant il est souvent impossible d'envisager toutes les possibilités lors de la conception ou de communiquer une telle décision en cours d'action. Les systèmes multi-agents sont généralement utilisés pour résoudre des problèmes où la prise de décision ne doit/peut pas être complètement centralisée afin d'obtenir des solutions optimales – c.-à-d. une autonomie de décision doit être laissée aux agents.

Points de vue similaires dans la littérature

La relation entre la coordination dans un système multi-agents, le contrôle / influence exercé sur les agents et leur autonomie est décrite dans la Figure 1.1. Nous ne sommes pas les seuls à nous intéresser à cette relation et à ses implications. Nous pouvons citer ainsi les travaux de (van der Vecht *et al.*, 2007) ou de (Chopinaud 2007). Dans le premier cas, les auteurs s'intéressent à la conception d'agents capables de changer leur degré d'autonomie, ce qui modifie en fait le type de coordination existante dans le système, en passant d'une coordination explicite à une émergente. Nous verrons ci-dessous à quoi correspondent ces deux types de coordination, pour l'instant nous voulons simplement souligner que si van der Vecht s'intéresse à comment l'autonomie des agents change le type de coordination dans un système, nous nous intéressons plutôt à comment le type de coordination influence l'autonomie des agents.

Quant aux travaux de Chopinaud (Chopinaud 2007), nous partageons avec l'auteur plusieurs points de vue et définitions. Par exemple, la définition de l'autonomie utilisée par l'auteur est similaire avec celle que nous utilisons pour la perspective interne : « un agent *est autonome* s'il est en mesure de prendre seul ses décisions ». Ce travail s'intéresse spécialement à comment détecter un comportement autonome (perspective externe) et à comment vérifier que la coordination dans un système multi-agents est atteinte, c.-à-d., comment s'assurer que le système fonctionne comme prévu. Avec ces outils en main, autonomie, détection des comportements autonomes et vérification du système, l'auteur s'intéresse à la génération des agents autonomes, tout en garantissant que leur autonomie ne nuit pas au système. Par rapport à cette approche, nous ne nous intéressons pas à l'état du système, ni à comment l'autonomie des agents bénéficie ou nuit à cet état ou comment concevoir des agents autonomes – notre objectif est de rendre les agents capables d'évaluer l'autonomie qu'ils ont par rapport à un système existant et comment cette évaluation améliore leur raisonnement.

Synthèse

Les deux travaux présents dans la littérature et mentionnés ci-dessus partagent notre point de vue qu'il faut s'intéresser aux notions de coordination et/ou contrôle pour travailler sur l'autonomie (ou l'inverse). De plus, comme nous, ils soulignent le compromis qu'il faut trouver entre le contrôle exercé de l'extérieur sur un agent et son autonomie. Un système dans lequel la prise de décision des agents est trop influencée devient trop rigide – les agents n'ont pas l'autonomie de prendre les meilleures décisions – tandis qu'un système dans lequel les agents ont trop d'autonomie risque de ne pas être coordonné et donc de ne pas atteindre ses objectifs. En plus, il est souvent difficile d'imposer un contrôle sur le comportement d'un agent sans connaître son fonctionnement (motivations qui orientent

son comportement), comme c'est le cas dans les systèmes ouverts. Par exemple, une sanction financière peut ne pas être vue comme une sanction pour un agent qui n'a pas le but de gagner de l'argent ou qui en a assez. Pour résumer, afin de pouvoir raisonner sur l'autonomie qu'un agent a ou non dans un système, il faut s'intéresser à la coordination présente dans ce système, comme nous le verrons ci-dessous.

1.3.2. Types de coordination

Le raisonnement présenté dans la section précédente nous mène à considérer l'autonomie comme une réponse locale d'un agent au besoin de coordination globale dans le système. Ces deux notions sont strictement liées (par l'intermède de la notion de contrôle), ce qui signifie à notre avis que les différents facteurs qui influencent l'autonomie sont issus des différentes formes de coordination possibles.

La coordination est un domaine de recherche actif et il existe beaucoup de modèles de coordination proposés dans la littérature. Schumacher (Schumacher 1999) classifie les modèles de coordination existants en coordination subjective et coordination objective, même si des travaux comme ceux d'Omicini (Omicini 2000) proposent l'utilisation de modèles hybrides situés entre les deux. La *coordination subjective* est issue des travaux dans le domaine de l'Intelligence Artificielle et représente un point de vue interne aux agents dans le contexte d'un système multi-agents. Dans cette forme, la coordination représente les efforts de chaque agent pour coordonner ses activités avec les autres agents, souvent en terme d'algorithmes ou de protocoles qui peuvent être utilisés pour assurer que les collectivités d'agents atteignent leurs objectifs. La *coordination objective*, met l'accent sur les modèles et mécanismes qui gèrent les interactions entre les agents. Cette forme de coordination sépare les problèmes liés au raisonnement des agents des problèmes de coordination entre agents, ce qui permet l'utilisation de modèles abstraits tels que les organisations d'agents.

Une classification similaire est proposée par Castelfranchi (Castelfranchi 2005) pour décrire les techniques possibles pour atteindre *un ordre social* (la coordination). L'ordre social est défini comme un schéma non-aléatoire d'interactions dans un système, schéma qui permet la satisfaction des objectifs d'une entité. Par exemple, un marché électronique dans lequel tous les agents négocient selon des protocoles bien définis est un système avec un ordre social. Ce que nous avons appelé contrôle (la traduction au niveau local du besoin d'atteindre une coordination au niveau global) est appelé *social control*, notion qui est définie comme toute action d'un agent destinée à renforcer la correspondance du comportement d'un autre agent à une norme sociale (telle qu'un rôle, convention, engagement social, etc.). Castelfranchi fait ensuite la différence entre le *contrôle sociale formel* (imposé aux agents, issu d'une coordination objective) et le *contrôle social informel* (émergent, créant une coordination subjective). Il mentionne aussi d'autres possibilités d'obtenir un ordre social, telle qu'une « main invisible » qui guide les comportements d'agents.

Synthèse

Cette brève revue des approches possibles pour coordonner les agents dans un système met en évidence l'existence de deux catégories de facteurs qui influencent l'autonomie des agents : formels (objectifs) ou informels (subjectifs). La nécessité d'un système multi-agents d'atteindre un ordre social

implique une coordination des agents qui en font partie, ce qui à son tour impose un contrôle sur leurs comportements, ce qui limite leur autonomie. L'autonomie des agents est donc limitée par des contraintes issues du besoin de coordination du système, contraintes subjectives ou objectives que nous analysons par la suite.

1.4. Contraintes sur la décision d'un agent

L'autonomie qu'un agent a dans son processus de décision dépend des facteurs qui influencent ce processus. Dans ce manuscrit nous utilisons le terme « *contrainte* » pour ces facteurs en provenance d'autres entités externes (agents, organisations, etc.) qui réduisent l'autonomie d'un agent. Nous pouvons classer ces contraintes en fonction du type de coordination qui les génère, mais aussi en fonction de ce qu'elles limitent, la prise de décision ou le comportement d'un agent. Pour mieux décrire ces types de contraintes nous allons nous appuyer sur la description d'un système multi-agents que nous utilisons comme exemple ci-dessous.

Ce système est basé sur celui proposé par (Bourne *et al.*, 2000). Il contient des agents qui se déplacent sur une grille pour trouver et exécuter des tâches. Ces tâches peuvent être simples (ST – Simple Task) et effectuées par un seul agent ou coopératives (CT – Cooperative Task) et effectuées par l'effort conjoint de plusieurs agents. L'exécution d'une tâche donne un profit aux agents, les profits des CT étant partagés par les agents coopérant à leur exécution. Les agents ont besoin d'interagir pour former des accords pour s'aider les uns les autres à effectuer des CT. Les agents peuvent faire partie ou non d'une société – les agents qui en font partie sont sujets d'une norme qui leur interdit de violer les contrats avec d'autres membres de la société. Un agent qui ne fait pas partie de la société peut profiter et donc rompre son contrat avec un autre agent, mais d'autres agents peuvent faire la même chose et rompre leurs contrats avec lui.

Considérons maintenant les contraintes qui limitent les agents dans ce système. Ces contraintes sont issues du besoin de coordination. Comme les mécanismes de coordination possibles, nous pouvons les diviser en contraintes subjectives et en contraintes objectives :

- Les *contraintes subjectives* concernent les *contraintes interpersonnelles*. Elles sont issues des interactions ou de besoin d'interagir avec d'autres agents. Par exemple, dans le système décrit ci-dessus, un agent qui veut effectuer une CT ne peut pas le faire tout seul. Il a besoin de l'aide des autres agents pour satisfaire son but, il dépend d'eux. Cette dépendance est issue des relations existantes entre les agents et représente leurs essais de coordonner leurs activités – elle constitue ainsi une contrainte interpersonnelle qui limite l'agent. Etant donné les définitions de l'autonomie sociale présentées auparavant, nous pouvons conclure que cet agent n'a pas d'autonomie par rapport à d'autres agents pour l'exécution d'une CT.

- Les *contraintes objectives* concernent les *contraintes institutionnelles*. Elles sont issues des mécanismes de coordination formels existants dans le système, tels que les structures organisationnelles ou les normes institutionnelles. Par exemple, dans le système décrit ci-dessus, un agent décide de faire partie d'une société dans laquelle il sera sujet d'une norme. Cette norme contraint son comportement et sa prise de décision : chaque fois qu'il forme un accord avec un autre agent pour co-exécuter une CT, la norme l'oblige à respecter cet accord, c.-à-d., d'avoir un comportement orienté vers la co-exécution de la CT. Bien évidemment, un agent peut avoir un comportement autonome par

rapport à la norme et lui désobéir, mais il risque de subir des conséquences (comme par exemple des sanctions).

Nous pouvons aussi faire une différence entre les contraintes qui limitent l'autonomie d'un agent en fonction de la perspective sur l'autonomie qui est limitée par elles. Nous pouvons ainsi parler de contraintes internes (décisionnelles) et contraintes externes (comportementales) :

- Les *contraintes décisionnelles* sont les contraintes qui limitent l'autonomie en délibération d'un agent. Un exemple de ce type de contrainte dans le système que nous utilisons comme exemple est l'impossibilité d'un agent de satisfaire tout seul une CT. Cette dépendance contraint sa prise de décision – il ne peut pas décider librement sans être influencé par d'autres parce qu'il dépend d'eux. Comme nous l'avons discuté ci-dessus, cette contrainte limite l'autonomie de décision d'un agent.

- Les *contraintes comportementales* sont les contraintes qu'un agent doit désobéir pour être considéré être dans un comportement autonome. Dans notre exemple, les agents forment des contrats pour l'exécution en coopération des CT. Ces accords, une fois créés, représentent des contraintes qui limitent les comportements possibles des agents qui sont maintenant obligés de participer à la co-exécution des CT pour lesquelles ils se sont engagés. Un agent qui désobéit à son contrat (ou qui désobéit à la norme le renforçant) pourra être qualifié d'être autonome (c.-à-d. avoir un comportement autonome par rapport à cette contrainte).

Synthèse

Pour résumer, les contraintes comportementales sont des contraintes qui apparaissent suite à des actions d'un agent, telles que l'adoption d'une norme ou l'entrée dans une société. Parce que ces contraintes sont établies par le comportement de l'agent, nous les appelons aussi *des contraintes établies*. Quant aux contraintes décisionnelles, elles représentent le potentiel de prise de décision d'un agent, raison pour laquelle nous les appelons aussi des *contraintes potentielles* ou *inhérentes*. Nous voulons souligner le fait que la différence entre les contraintes décisionnelles et contraintes comportementales n'est pas toujours évidente. Dans l'exemple ci-dessus, avant d'être adoptée, une norme représente une contrainte décisionnelle – elle limitera la liberté de décision d'un agent. Si l'agent adopte la norme, elle représente une contrainte établie – si l'agent refuse d'essayer de remplir son accord, il viole la norme et il est autonome par rapport à elle.

1.5. Synthèse

Malgré la multitude des définitions de l'autonomie existantes dans la littérature, nous avons réussi à expliquer certaines différences entre elles par la propriété relationnelle de l'autonomie : un agent a de l'autonomie par rapport à une autre entité pour quelque chose. Ceci nous a permis d'identifier plusieurs formes d'autonomie, parmi lesquelles une d'une grande importance pour notre travail : *l'autonomie sociale de décision* d'un agent. Nous avons passé en revue des définitions différentes proposées dans la littérature pour cette forme d'autonomie, définitions s'intéressant à l'autonomie sociale d'un agent par rapport à un autre agent (dans des interactions) ou par rapport à une norme (dans des organisations). Nous avons identifié ensuite deux perspectives possibles sur l'autonomie : la perspective externe analysant le comportement autonome d'un agent (p. ex. le fait de n'avoir pas adopté un but ou d'avoir violé une norme) et la perspective interne analysant la prise de décision

autonome d'un agent (p.ex. sa liberté de décider sans influence d'une entité externe). Les deux perspectives sont toutes deux importantes pour le bon fonctionnement d'un système multi-agents et pour le raisonnement d'un agent, surtout dans la problématique d'obtenir une coordination au sein du système.

Nous avons analysé ensuite la relation existant entre la notion d'autonomie en délibération et le contrôle exercé sur le comportement d'un agent par le besoin de coordination existant dans un système multi-agent. Cette coordination crée des contraintes sociales qui limitent l'autonomie d'un agent. Nous avons classé ces contraintes en contraintes subjectives/interpersonnelles (issues de l'interaction) ou objectives/institutionnelles (issues des normes). Nous les avons également distinguées en contraintes comportementales/établies (pouvant être désobéies par un agent autonome) ou décisionnelles/inhérentes (qui limitent l'autonomie d'un agent).

Tableau 1.1. Types de contraintes et d'autonomie

Types de contraintes	**interpersonnelle**	**institutionnelle**
décisionnelle (potentielle, inhérente)	**perspective interne : avoir de l'autonomie (dans sa prise de décision)**	
	ex: la décision d'un agent d'adopter ou pas le but d'un autre agent est influencée par ce deuxième	ex: la décision d'un agent d'adopter une norme est influencée par le rôle qu'il joue ex: la décision d'un agent de poursuivre un but, est influencée par une norme qu'il a
comportementale (établie)	**perspective externe : être autonome (dans son comportement)**	
	ex: un agent s'est engagé envers un autre de satisfaire un de ses buts – s'il viole son engagement il est dans un comportement autonome	ex: un agent a adopté une norme qui l'oblige de satisfaire un but – s'il ne le satisfait pas (viole la norme), il est dans un comportement autonome

L'exemple présenté dans la section précédente illustre comment les décisions prises par un agent (telles qu'adopter ou pas une norme ou un but) sont limitées par différentes contraintes sociales, mais également par d'autres contraintes créées par ces décisions. Par exemple, l'existence de la norme est une contrainte décisionnelle, tandis que le fait de prendre la décision d'adopter cette norme crée une contrainte comportementale. L'objectif de ce travail est de pouvoir évaluer les implications qu'une décision a sur la liberté de décision et de comportement d'un agent (son autonomie). Nous sommes donc menés à étudier ces contraintes sociales issues de la coordination, leur apparition et leur influence sur la prise de décision d'un agent. Les chapitres suivants décrivent les modèles de coordination institutionnelle et interpersonnelle existants.

2. Coordination et contraintes institutionnelles

Une approche possible pour coordonner les agents au sein d'un système est d'utiliser des concepts institutionnels, comme par exemple les organisations multi-agents. Le terme *institutionnel* est utilisé dans ce manuscrit dans son sens large, non réduit au sens de e-institution. Cette approche offre aux agents un cadre formel et explicite pour coordonner leurs activités, en décrivant pour chacun d'entre eux le comportement nécessaire pour atteindre la coordination dans le système et de les inciter à suivre ce comportement. Nous commençons notre analyse de ce type de coordination en passant brièvement en revue quelques modèles organisationnels, ce qui nous permet d'identifier plusieurs concepts qui semblent être centraux dans ce type d'approche, comme les *rôles* ou les *normes*. Nous continuons ensuite à présenter ces concepts, en commençant avec les normes qui, comme nous le verrons, ne représentent pas un concept purement institutionnel, dans la mesure où des normes personnelles et interpersonnelles peuvent également exister. Après avoir étudié quelques approches possibles pour la modélisation des normes et de leurs propriétés, nous analysons des différents modèles de rôles existant dans la littérature en essayant d'identifier les caractéristiques principales de cette notion. Comme nous le verrons par la suite, les rôles ne sont généralement pas considérés indépendamment, mais organisés dans des hiérarchies ou des structures organisationnelles – nous décrivons quelques modèles de ce type également.

Cette analyse des concepts utilisés dans la coordination institutionnelle et ainsi dans le contrôle institutionnel nous permettra d'identifier comment le comportement attendu d'un agent dans une organisation est généralement spécifié. Cependant, les concepteurs des systèmes se méfient du caractère autonome des agents qui les conduit souvent à ne pas avoir le comportement attendu. C'est pourquoi des mécanismes de renforcement sont utilisés. Ces mécanismes sont souvent intégrés dans des institutions multi-agents (nous en présenterons quelques modèles) et des sanctions sont utilisées pour inciter les agents à ne pas dévier du comportement spécifié – une ontologie de sanctions est également présentée. Finalement, nous concluons ce chapitre en analysant les contraintes que les modèles de coordination institutionnelle imposent sur la prise de décision et le comportement des agents.

2.1. Quelques modèles organisationnels

L'utilisation d'une organisation multi-agents a généralement comme objectif de donner aux agents un cadre précis dans lequel ils peuvent coordonner leurs activités. Il existe plusieurs modèles organisationnels proposés dans la littérature. La communauté multi-agents s'intéresse de plus en plus à ces modèles et à leur utilisation. Comme dans ce manuscrit nous prenons le point de vue d'un agent immergé dans une organisation, nous ignorons délibérément les travaux qui s'intéressent au cycle de vie d'une organisation (voir par exemple (Matson and deLoach 2005) sur les états possibles d'une organisation et les transitions possibles entre ces états) et aussi à la re-organisation (Hubner 2003). Nous nous intéressons surtout aux relations qui existent entre une organisation et ses membres : qu'est-ce qui est demandé à un agent qui en est membre, comment l'agent peut comprendre les spécifications organisationnelles et comment l'organisation peut vérifier que l'agent respecte ces

spécifications ? Bien évidemment, les réponses à ces questions ne sont pas simples à donner et la multitude des modèles proposés dans la littérature en est le témoin.

Par la suite, nous présentons brièvement quelques modèles organisationnels parmi les plus connus dans la communauté scientifique multi-agents, ce qui nous permettra ensuite d'identifier les concepts de base utilisés dans la coordination institutionnelle, concepts que nous analyserons ensuite dans le détail.

(1) Le modèle AGR

Un des modèles organisationnels le plus connu et parmi les premiers proposés est le modèle AGR (agent-groupe-rôle) décrit dans (Ferber and Gutknecht 1998). Ce modèle assez simple considère que les *agents* jouent un ou plusieurs *rôles* dans un ou plusieurs *groupes*. Chaque rôle décrit un statut (une étiquette) qui doit être adopté par un agent qui le joue contraignant les communications que l'agent peut avoir avec les autres agents du même groupe. Un agent peut jouer plusieurs rôles ou faire partie de plusieurs groupes, mais généralement ceci est décidé par le concepteur du système, ce qui constitue un des inconvénients de ce modèle – les organisations de type AGR sont statiques, c.-à-d., leur structure ne change pas en cours d'exécution. Ce modèle est d'une grande simplicité (seulement trois notions de base utilisées), ce qui est à la fois un avantage et un inconvénient. Même s'il est facile à utiliser, le modèle organisationnel AGR est de moins en moins utilisé parce qu'il ne permet pas de décrire des organisations complexes et ne s'intéresse pas explicitement à la dynamique d'une organisation.

(2) Le modèle RBAC

Le modèle AGR a été parmi les premiers à utiliser les notions de rôle et groupe pour coordonner les agents d'un système. Ces notions sont utilisées dans la plupart des organisations multi-agents, mais elles ne sont pas propres à la communauté multi-agents. Par exemple, les modèles de *contrôle d'accès à base de rôles* (role-based access control en anglais – RBAC) sont à l'heure actuelle considérés l'approche la plus efficace, pour spécifier le contrôle d'accès dans les systèmes d'informations ou les organisations dynamiques. Ces modèles ont été simplifiés et appliqués aux systèmes multi-agents par les auteurs de (Ricci *et al.* 2004). Sans détailler plus, nous tenons à souligner que la notion de *rôle* est encore une fois centrale : des politiques d'accès sont définies et associées aux rôles – tout utilisateur (le modèle RBAC) ou agent (dans les SMA) qui joue un rôle doit obéir à ces politiques. Un rôle peut hériter d'un autre de telles politiques et des hiérarchies de rôles sont aussi définies, ce qui permet la spécification des organisations plus complexes. Ce modèle permet donc la spécification des organisations plus complexes en manipulant la notion de rôle, mais aussi celle de *norme* – politique d'accès dans la terminologie RBAC.

(3) Le modèle MOISE+

Comme nous le voyons, un rôle ne peut pas être défini seulement par son nom (comme dans le modèle AGR), mais il doit pouvoir être associé à plusieurs concepts, tels que des normes – politiques d'accès (dans le modèle RBAC). Un autre modèle organisationnel, qui utilise aussi ces concepts de rôle, groupe et norme est le modèle MOISE+ (Hübner 2003). Ce modèle divise la spécification d'une organisation en trois catégories distinctes, mais interdépendantes : spécification *structurelle*,

fonctionnelle et *déontique*. La spécification structurelle définit les relations entre agents à l'aide des notions de *rôle*, de *groupe* et de *lien* entre rôles ou groupes. Des notions d'héritage entre rôles, liens d'autorités, etc. sont aussi utilisés. Nous le verrons plus tard lors de l'analyse en détail des hiérarchies de rôles. La spécification fonctionnelle décrit comment le système atteint ses objectifs, en décomposant le but global du système en sous-buts via des plans et en organisant ces sous-buts en *missions*. Une organisation décrite de cette manière fonctionne en supposant qu'un agent qui joue un rôle satisfait les missions associées à son (ses) rôle(s) et qu'il le fera en utilisant les liens existants entre le(s) rôle(s) et groupe(s) de l'organisation. Pour faciliter (mais aussi pour s'assurer de) la satisfaction des missions par les agents jouant des rôles, la spécification déontique décrit les normes (permissions et obligations) associées aux rôles et concernant les missions, en délégant ainsi des missions aux rôles.

La spécification fonctionnelle dans le modèle MOISE+ illustre comment les organisations multi-agents peuvent être utilisées pour atteindre la coordination dans un système : des parties du but global sont déléguées aux agents qui doivent les exécuter d'une manière coordonnée. Pour simplifier le partage de tâche, le concept de *rôle* est généralement utilisé : un sous-but n'est pas délégué à chaque agent, mais aux rôles – les rôles spécifient ainsi les buts que les agents qui les jouent doivent satisfaire. Pour assurer l'exécution coordonnée de ces buts délégués aux rôles, des *normes* sont aussi associées aux rôles, normes qui aident les agents de satisfaire les buts de leurs rôles, mais qui les empêchent dans le même temps de dévier du comportement attendu par l'organisation. Afin de mieux comprendre les contraintes imposées à un agent par l'organisation de laquelle il appartient, nous devons identifier l'influence que ces concepts de norme ou de rôle ont sur l'autonomie ou le comportement d'un agent. Par la suite nous allons analyser ces deux concepts pour ensuite tirer des conclusions.

2.2. Normes

Même si dans certains modèles organisationnels, comme MOISE+, les normes sont associées à des rôles et si en général ces deux notions sont interliées (comme nous le verrons par la suite), ce n'est pas toujours le cas. Il existe plusieurs types de normes et certains de ces types ne sont pas liés aux rôles, ni même aux concepts institutionnels. Dans cette section nous analysons le concept de *norme* en présentant plusieurs de ses caractéristiques et en identifiant ainsi plusieurs types de normes. Nous discutons ensuite plusieurs approches existantes pour modéliser les normes pour les rendre compréhensibles par des agents artificiels.

2.2.1. Différents types de normes

Il existe souvent des confusions concernant le concept de norme, confusions liées surtout aux différents types de normes existants. L'analyse de ce concept est rendue difficile par le fait qu'il est utilisé, parfois avec des sens légèrement différents, dans divers domaines, tels que les sciences sociales ou le droit. La communauté multi-agents ne commençant que relativement récemment à s'y intéresser. Par la suite nous présentons une discussion issue des sciences sociales sur différents types de normes, discussion qui nous permettra ensuite de mieux identifier la notion de norme dans un système multi-agent qui nous intéresse dans ce manuscrit.

(1) Normes sociales et personnelles

Selon Tuomela (Tuomela 1995), il existe plusieurs types de normes dans les sociétés humaines. Ces normes peuvent être divisées en deux grandes catégories : les *normes personnelles (ou potentiellement sociales)* et les *normes sociales*. Les normes de la première catégorie sont des normes qu'un agent utilise et qui ne sont pas partagées par d'autres agents. Cependant, il est possible qu'une telle norme, personnelle à l'agent, soit utilisée par d'autres agents aussi. Ces normes sont appelées *personnelles* parce qu'elles décrivent des concepts propres à chaque agent, mais Tuomela les appelle aussi *potentiellement sociales* parce qu'elles peuvent être transformées en normes sociales suite aux interactions entre les agents. Deux types de normes personnelles existent : les *normes de prudence (ou techniques)* et les *normes morales*. Le premier type (nommé généralement des p-normes) représente les normes « de sens commun » utilisées par un agent, telles que « évite de te faire mal », « maximise ton profit », etc., tandis que le deuxième type (appelé généralement des m-normes) représente des normes similaires mais orientées vers le bien-être des autres, telles que « soit altruiste » ou « ne pas tuer », etc.

Si nous considérons par exemple une norme potentiellement sociale, dans une société d'agents il est possible que chaque agent ait la même norme, mais indépendamment des autres. Cependant, s'il existe une croyance commune dans la société sur l'existence de cette norme et s'il existe une *réponse sociale* (réaction de la société) à la satisfaction ou non-satisfaction de la norme, cette norme devient une norme *sociale*. Il existe deux types de normes sociales, les *règles* (r-normes) et les *vraies normes sociales* (s-normes). La principale différence entre ces deux types est que les r-normes sont émises par une autorité ou un groupe d'agents et qu'elles sont explicitement définies, tandis que les s-normes sont basées sur des croyances communes et partagées entre les agents. Les r-normes sont imposées par une autorité (généralement institutionnelle), elles sont formelles et des sanctions (récompenses ou pénalités) explicites seront imposées par l'autorité quand ces normes sont satisfaites ou violées – nous les appelons ainsi des *normes institutionnelles*. Les s-normes représentent des conventions existant au sein d'une société, conventions parfois implicites et avec des sanctions sociales associées, sanctions telles qu'une baisse de réputation – nous les appelons ainsi des *normes interpersonnelles*.

Si nous considérons l'exemple de la norme morale « ne pas tuer », même si tous les agents ont cette norme, elle n'est pas sociale tant qu'il n'existe pas une croyance commune dans la société sur son existence. S'il existe une telle croyance, c.-à-d., si les membres de la société considèrent qu'il existe la norme de « ne pas tuer », alors cette norme devient une norme sociale. Si une autorité formalise cette norme, l'annonce aux agents concernés, veille à sa satisfaction et impose des sanctions quand elle est violée (par exemple l'exclusion de la société – la prison), alors cette norme représente une loi – une r-norme, sinon elle reste une s-norme.

(2) Obligations, permissions et interdictions

Indifféremment du caractère social d'une norme, Tuomela (Tuomela 1995) propose une autre classification des normes, en s'appuyant sur leur modalité. Il existe ainsi des normes de type « *devrait* » (ought-to en anglais – normes qui sont appelées généralement des *obligations*) et des normes de type « *pourrait* » (may en anglais – généralement appelées des *permissions*). En plus, chacun de ces types peut être encore divisé en deux catégories, en fonction de l'objet de la norme : il

existe par exemple des normes de type « *devrait faire* » et des normes de type *« devrait être* ». Dans le premier cas, un agent est obligé d'effectuer une action, tandis que dans le deuxième un agent est obligé d'atteindre un état du monde (satisfaire un but).

Cette classification de normes est à la base d'une logique modale qui a comme but la représentation et la formalisation de ces concepts : la logique déontique (von Wright 1981). Plusieurs variantes à la logique déontique classique ont été proposées. Ce sont, par exemple, la logique déontique contextuelle (van der Torre 2003),. Des travaux ont été menés pour la conception de systèmes informatiques en utilisant la logique déontique (voir par exemple (Meyer and Wieringa 1991)). Sans détailler plus, nous précisons que la logique déontique utilise deux ou trois opérateurs modaux pour représenter les normes : l'opérateur $O_x p$ représente l'obligation d'un agent de faire en sorte que la proposition p soit vraie, l'opérateur $P_x p$ représente la permission qu'un agent a de faire en sorte que la proposition p soit vraie et l'opérateur $F_x p$ représente l'*interdiction* qu'un agent a de faire en sorte que la proposition p soit vraie. Les permissions et les interdictions d'un agent sont complémentaires et généralement les travaux dans les systèmes multi-agents n'utilisent qu'une des deux modalités – par exemple en considérant que tout ce qui n'est pas explicitement permis est interdit ou l'inverse[1].

(3) Normes interpersonnelles

Cette thèse s'intéresse aux contraintes sociales imposées aux agents et donc aux contraintes issues des autres agents ou des structures institutionnelles. Nous ne nous intéressons donc pas aux normes personnelles – propres aux agents – mais aux normes sociales qui peuvent être des r-normes (institutionnelles) ou des s-normes (interpersonnelles). Même si dans ce chapitre nous nous intéressons principalement à des concepts institutionnels, nous analysons brièvement ici les normes interpersonnelles. Ce type de normes est lié au concept des contraintes interpersonnelles que nous analyserons en détail dans le Chapitre 3.

Dans (Conte and Castelfranchi 1999), Conte et Castelfranchi s'intéressent au cycle de vie d'une norme : comment elle est créée, annoncée, adoptée, renforcée, etc. et comment il est possible de passer d'une norme interpersonnelle à une norme institutionnelle. Les auteurs considèrent qu'une norme sociale n'est pas vraiment sociale sans une reconnaissance générale par les agents. Quand un agent reconnaît une norme comme étant une norme et qu'il décide de lui obéir, le nombre d'agents obéissant à cette norme augmentant son efficacité est accrue.. Plus le nombre d'agents obéissant à une norme est grand, plus les chances que d'autres agents observent ce comportement normatif augmentent. Ces agents forment ainsi des croyances normatives (concernant les normes), c'est-à-dire, en se basant sur le contexte et sur les buts probables d'autres agents, ils arrivent à croire dans l'existence d'une norme. Quand un tel agent reconnaît l'existence d'une nouvelle norme, il est probable qu'il voudra lui obéir, soit parce qu'il considère que la norme lui est utile, soit parce qu'il veut éviter les sanctions sociales (baisse de réputation, par exemple) ou parce qu'il veut s'intégrer, appartenir au groupe et donc être comme les autres agents qui sont sujets de la même norme. Dans ce cas, un agent reconnaît la norme

[1] Dans une discussion privée avec l'auteur, Castelfranchi argumentait que le problème de comment choisir dynamiquement quelle est la modalité explicite dans un système est très intéressant et constitue un aspect négligé à l'heure actuelle par la communauté multi-agents. Par exemple, en fonction de la réputation des sanctions dans un système, un agent qui y entre peut choisir la modalité à utiliser implicitement – un être humain allant dans un pays géré par une dictature peut considérer tout ce qui n'est pas explicitement permis est interdit.

comme étant une norme parce qu'il a observé un comportement normatif mis en œuvre par d'autres agents et décide de l'adopter et de lui obéir, renforçant ainsi ce processus de renforcement collectif.

Conte et Castelfranchi montrent ainsi que le simple fait d'obéir à une norme transforme un agent en un agent qui émet la norme (son comportement sera observé par d'autres) et qui la renforce (les autres voudront lui obéir pour avoir un comportement similaire à cet agent). Ce processus très intéressant ne garantit malheureusement pas un comportement conforme aux normes pour tous les agents d'un système. Les mécanismes qui sont à la base de la diffusion et du renforcement distribué marchent – surtout dans les sociétés humaines, mais il n'y a pas de preuves que ce soit toujours le cas – surtout dans les systèmes multi-agents. Pour cette raison, des normes institutionnelles sont souvent utilisées pour garantir l'obtention de la coordination dans un système, même si des auteurs comme Castelfranchi (Castelfranchi 2005) argumentent qu'aucune des deux approches ne fonctionne isolément à 100% et que des approches hybrides doivent être envisagées.

Synthèse

Ce qui différencie une norme institutionnelle d'une norme interpersonnelle est son caractère clair et formel : une norme institutionnelle est clairement énoncée (les agents ne doivent pas observer le comportement d'autres agents et deviner qu'il est normatif), les confusions possibles sur le but de la norme sont généralement évitées et la décision d'un agent d'obéir ou pas à la norme est motivée par des sanctions précises (et généralement inévitables) qui sont associées à la norme. Nous observons ainsi que pour utiliser des normes institutionnelles dans les systèmes multi-agents, il faut définir formellement les normes (modèle de norme), préciser clairement le contexte dans lequel une telle norme doit être obéie et par qui (attribution de norme) et mettre en place des mécanismes de renforcement des normes (institutions). Par la suite nous analysons le premier aspect (les modèles de normes existants) et dans les sections suivantes nous nous intéressons aux deux autres.

2.2.2. Modèles de normes

Comme nous l'avons précisé, la logique déontique est une logique modale qui permet de spécifier les normes (obligations, permissions ou prohibitions) imposées à un agent. Cette logique est très utile pour décrire formellement les normes et surtout pour vérifier des propriétés d'un système normatif (comme par exemple la cohérence des normes (Ferber 2003), etc.), mais elle a quelques inconvénients. Par exemple, comme montré dans (van der Torre 2003), le contexte d'une norme est important, une norme peut être active seulement dans un contexte spécifique – une logique déontique contextuelle est proposée afin de le prendre en compte. Cependant, le grand inconvénient de cette représentation est le fait qu'étant une logique modale, il est difficile de créer des agents capables de raisonner et de prendre des décisions rapides vis-vis des normes qui leur sont imposées.

Pour permettre aux agents de raisonner sur les normes, des modèles plus simples ont été proposés dans la littérature. Une des approches les plus fréquentes est d'utiliser la logique des prédicats pour la représentation des normes, comme dans les travaux de Stratulat (Stratulat 2002). Il s'intéresse notamment à la représentation des actions et du temps à l'aide des prédicats. Il utilise ainsi trois prédicats spéciaux pour dénoter les obligations, permissions et interdictions d'un agent d'effectuer une action dans un intervalle de temps. Il est intéressant de souligner que, dans cette approche, une norme

concerne une action (p.ex. : « payer ») et non une instance d'une action (p.ex. : « signer le cheque no. 4526 »), ce qui permet de représenter des normes plus générales, mais les agents doivent être capables d'identifier l'action à laquelle appartient une instance. La représentation simple, à base de prédicats, est facilement utilisable en pratique et Stratulat propose aussi des règles d'inférence qui permettent d'identifier les violations de normes.

Un autre modèle de normes utilisé dans les systèmes multi-agents est celui proposé par Lopez y Lopez (López y López 2003). Une norme est considérée comme une structure constituée de plusieurs champs. Une norme est ainsi un ensemble d'agents *cible* – les agents qui devront adopter et obéir à la norme, ensemble qui peut être différent de l'ensemble des agents *bénéficiaires* de la norme – agents bénéficiant de l'obéissance à la norme et qui généralement représentent une institution ou un groupe. Une norme inclut des *buts normatifs* – les buts que les agents cibles doivent satisfaire dans le cas d'une obligation ou ne doivent pas atteindre dans le cas d'une interdiction. De plus, une norme n'étant pas toujours applicable et son activation dépendant du contexte de l'agent, il existe également un *contexte* d'activation d'une norme. Des *exceptions* peuvent également être définies pour spécifier quand les agents ne doivent pas obéir à la norme. Finalement, une norme est associée à un ensemble de *punitions* -sanctions négatives- (resp. des *récompenses* - sanctions positives-) à appliquer aux agents qui la violent, (resp. qui lui obéissent). Nous nous intéressons aux bénéficiaires d'une norme et à ses sanctions dans une section ultérieure. Nous voulons ici souligner deux caractéristiques intéressantes de ce modèle. Premièrement, ce modèle permet de représenter des normes qui ciblent plus d'un agent, ce qui est nécessaire dans les systèmes avec de nombreux agents : il n'est pas pratique de définir des normes pour chaque agent, mais il est nécessaire de grouper ces agents (dans des ensembles dans cette approche ou avec des rôles – comme nous le verrons par la suite). Deuxièmement, ce modèle permet de rendre explicite le contexte d'activation et des exceptions associées à une norme – ce qui permet la conception de systèmes normatifs plus dynamiques et la définition des normes plus générales.

Ces caractéristiques se retrouvent aussi dans le modèle normatif *HarmonIA* proposé par Vázquez-Salceda dans (Vázquez-Salceda 2004). Ce modèle considère l'existence de quatre niveaux de représentation de normes : le niveau abstrait, le niveau concret, le niveau des règles et le niveau procédural. Les normes au niveau abstrait représentent le statut, les valeurs, les objectifs d'une organisation. Ces normes ne sont pas directement utilisables par les agents et doivent donc être transformées en des *normes concrètes* (r-normes) qui sont représentées dans la logique des prédicats. Ceci permet de vérifier diverses propriétés du système normatif, telle que la cohérence des normes. En utilisant des données spécifiques au domaine d'application ainsi que le contexte, les normes concrètes sont ensuite transformées en des *règles* de comportement (normes personnelles) facilement utilisables par le processus de raisonnement d'un agent. Pour permettre aux agents qui n'ont pas un raisonnement à base de règles d'avoir un comportement normatif, au niveau procédural les règles sont interprétées et transformées en des actions à effectuer par les agents. Le grand apport de cette approche ne réside pas dans une nouvelle représentation des normes, mais dans la distinction proposée entre plusieurs niveaux de normes. Le concepteur d'un système peut ainsi définir des normes et vérifier leurs propriétés. Ces normes sont ensuite transformées en des règles de comportement incorporées aux agents. Cependant, un inconvénient à cette approche est que les agents ne peuvent plus reconnaître et raisonner sur les

normes : un agent ne connaît que des règles de comportement, qu'il sait issues des normes, sans avoir accès aux normes proprement dites.

Synthèse

D'autres modèles de normes existent dans la littérature, modèles que nous ne décrivons pas ici mais qui seront mentionnés plus tard quand nous analyserons les structures organisationnelles normatives ou les institutions d'agents. Notre objectif n'est pas de proposer un nouveau modèle de norme ou de système normatif, mais d'identifier les contraintes que l'existence des normes impose aux agents. Nous retenons ainsi qu'il existe des normes sociales interpersonnelles et institutionnelles, normes qui peuvent prendre la forme d'obligations, de permissions ou d'interdictions et qui spécifient quel agent et dans quel contexte est obligé/permis/interdit d'effectuer une action, de satisfaire un but ou en général d'exécuter une tâche. S'il est généralement difficile d'utiliser les normes interpersonnelles dans les systèmes multi-agents à cause des ambiguïtés possibles dans leur cycle de vie, les normes institutionnelles sont plus fréquentes et elles sont généralement définies en utilisant la notion de rôle, comme nous le verrons par la suite.

2.3. Rôles et relations entre rôles

Les rôles représentent un composant de base des modèles organisationnels et constituent généralement une abstraction d'activités et de politiques. Les rôles peuvent être considérés comme des places vides à être remplies par des agents, places décrivant le comportement attendu par l'organisation de la part des agents qui les jouent. La notion de rôle est centrale pour la coordination institutionnelle et dans cette section nous passons en revue quelques définitions données dans la littérature pour nous intéresser ensuite aux relations existantes entre les rôles d'une organisation : les structures organisationnelles.

2.3.1. Définitions des rôles

Dans une section précédente nous avons brièvement décrit le modèle organisationnel MOISE+ (Hübner 2003), modèle qui consiste en trois spécifications : structurelle, fonctionnelle et déontique. La spécification structurelle définit les rôles existants, ainsi que les relations entre eux. Un rôle est caractérisé dans cette approche par son *nom* et par plusieurs caractéristiques, comme par exemple celle de *cardinalité* : le nombre d'agents qui peuvent jouer le rôle. Un autre concept qui caractérise les rôles dans ce modèle est le *lien de communication* entre deux rôles : deux agents peuvent communiquer un avec l'autre seulement s'ils jouent des rôles liés par un tel lien, lien qui peut être considéré comme une norme (une permission) attachée aux deux rôles. Pour simplifier la conception des organisations, le modèle MOISE+ utilise aussi la notion de groupe et les relations d'héritage entre rôles ou entre groupes. Un groupe est constitué de plusieurs rôle et ses caractéristiques sont héritées par les rôles qui le compose – par exemple, un lien de communication entre deux groupes est équivalent à des liens de communication entre tous les rôles d'un groupe avec tous les rôles de l'autre. De plus, une fois un rôle défini, le concepteur de l'organisation peut définir un autre rôle héritant du premier, c.-à-d., ayant les mêmes caractéristiques auxquelles d'autres peuvent être ajoutées.

Si la spécification structurelle identifie les rôles d'une organisation, la spécification fonctionnelle du modèle MOISE+ décrit le comportement attendu des agents qui les jouent en terme de schémas sociaux ensembles de buts décomposés en sous buts. La spécification déontique attribue à chaque rôle

un ensemble de missions (ensemble cohérent de buts) que l'agent qui joue ce rôle doit satisfaire. L'organisation aide ses membres à satisfaire les missions de leurs rôles, généralement en leur donnant accès aux ressources utiles. La spécification déontique exprime dans l'attribution des missions aux rôles des obligations/permissions utilisées pour éviter le fait que les agents puissent ne pas poursuivre les missions de leurs rôles. Pour résumer, le modèle MOISE+ offre des outils pour la spécification des rôles, rôles qui sont considérés en termes des buts (missions) qu'ils doivent satisfaire et normes (permissions ou obligations) qui limitent le comportement utilisé pour la satisfaction de ses buts.

Cette approche est similaire à celle proposée par les auteurs de (Dastani *et al.* 2003), qui définissent un rôle comme étant composé d'un ensemble de buts à satisfaire, un ensemble de normes à obéir et un ensemble de règles d'interactions à respecter. Cette définition d'un rôle est ensuite formalisée et intégrée dans le langage 3APL[2]. Dans cette formalisation un agent est considéré formé d'un ensemble de buts et un ensemble de plans qui peuvent être utilisés pour satisfaire les buts ; ces deux ensembles sont partiellement ordonnés pour représenter les préférences qu'un agent a vis-à-vis de ses buts ou plans. Un rôle est défini de la même manière qu'un agent (ensembles ordonnés de buts et de plans), mais en rajoutant un ensemble d'interdictions et un ensemble d'obligations. Les interdictions d'un rôle représentent les buts qu'un agent qui joue ce rôle ne doit pas essayer de satisfaire, tandis que les obligations d'un rôle représentent des buts qu'un agent qui le joue doit satisfaire. Nous voyons ainsi que cette approche considère aussi un rôle comme étant défini par des buts à satisfaire et des normes à obéir,. Elle pousse cependant le problème plus loin, en s'intéressant aussi à l'adoption d'un rôle par un agent. En fonction des buts propres à l'agent et des buts de son rôle, mais aussi en considérant les normes et les relations de préférence, des conflits peuvent apparaître au sein d'un agent qui joue un rôle, des conflits entre des buts propres et des buts adoptés. Nous analysons le raisonnement d'un agent qui joue un rôle et qui doit choisir entre la satisfaction d'un but propre (et une éventuelle violation de norme) et la satisfaction d'un but du rôle (et une éventuelle non-satisfaction de son propre but) dans le Chapitre 4.

Une approche qui considère aussi un rôle dans les même termes qu'un agent est celle de Boella et van der Torre (Boella and van der Torre 2004b). Cette approche consiste en la définition et la caractérisation d'un rôle avec les mêmes concepts utilisés pour définir un agent : croyances, désirs, buts, etc., auxquels s'ajoutent les obligations d'un rôle. En ce qui concerne les buts et les obligations d'un rôle, cette approche est similaire à celles présentés ci-dessus, mais en plus elle considère aussi les croyances et les désirs d'un rôle. Les désirs d'un agent représentent la source motivationnelle utilisée pour générer des buts et choisir parmi eux ; les désirs d'un rôle sont similaires, mais ils ne sont pas toujours adoptés par les agents. Par exemple, le désir d'un rôle peut être d'assurer le bien de l'organisation, mais l'agent qui joue le rôle et qui satisfait ses buts le fait peut être pour gagner de l'argent (son désir propre), sans considérer le désir du rôle. Un rôle peut avoir aussi des croyances associées, croyances qui, même si elles ne deviennent pas les croyances de l'agent qui joue le rôle, doivent caractériser son comportement. La société humaine offre un exemple intéressant, dans le cas d'un agent qui joue le rôle de politicien – ce rôle a la croyance que tous les gens sont égaux, indifféremment de leur race, mais l'agent qui joue le rôle peut avoir des croyances très différentes. Cependant, tant qu'il joue ce rôle, son comportement est restreint par les croyances du rôle.

[2] 3APL – langage de programmation d'agents : http://www.cs.uu.nl/3apl/

Les auteurs de (Demolombe and Louis, 2006) s'intéressent aussi à la définition d'un rôle dans une organisation multi-agents. Par rapport à d'autres travaux similaires, ce travail met en évidence les conditions qu'un agent doit remplir pour jouer un rôle. De plus, les auteurs font une différence entre les normes dites statiques – toujours valides – et les normes dynamiques qui ont un contexte d'activation. Ces dernières sont représentées dans leur modèle comme des pouvoirs institutionnels – dans le chapitre suivant nous analysons aussi cette notion de pouvoir.

Synthèse

Cette brève étude de quelques approches existantes sur les définitions de rôles nous permet de considérer qu'un rôle est généralement caractérisé par des buts à satisfaire et des normes à obéir. Les buts globaux d'une organisation sont divisés dans des sous-buts qui sont ensuite associés aux rôles et tout agent qui joue un rôle doit essayer de satisfaire ces buts. Pour assurer le fait que le comportement des agents ne dévie pas du comportement désiré (celui qui satisfait les buts), des obligations et des interdictions sont également associées aux rôles. Pour aider les agents jouant des rôles à satisfaire les buts de leurs rôles, des permissions sont aussi associées aux rôles ; ces normes qui caractérisent des rôles représentent des normes institutionnelles. Cependant, parfois, ces permissions ne suffissent pas et les agents ont besoin de l'aide d'autres agents pour satisfaire les buts de leurs rôles – des structures organisationnelles sont ainsi mises en place, comme nous le verrons ci-dessous.

2.3.2. Structures organisationnelles

Comme la plupart des approches concernant la conception d'organisations multi-agents le montre, pour coordonner les agents il ne suffit généralement pas de définir simplement les rôles que les agents peuvent jouer (en termes de buts à satisfaire, normes à obéir, etc.). Ces rôles sont souvent structurés et des relations sont définies entre eux – ces structures et liens organisationnels aident les agents jouant des rôles à mieux satisfaire les buts de l'organisation.

Comme nous l'avons précisé, le modèle organisationnel MOISE+ (Hübner 2003) consiste en trois spécifications, structurelle, fonctionnelle et déontique. La spécification structurelle définit des rôles, mais aussi des liens entre eux et structure ainsi ces rôles dans une *structure organisationnelle*. Les rôles appartiennent à des groupes et il est possible d'avoir des groupes récursifs, c.-à-d., des groupes appartenant à d'autres groupes. Des liens structurels existent entre des rôles ou des groupes, comme par exemple le lien de compatibilité entre deux rôles : un agent peut jouer les deux rôles seulement s'il existe ce type de lien entre eux. Cependant, le lien structurel qui nous semble le plus intéressant est le *lien d'autorité* entre deux rôles : un agent qui joue un rôle avec de l'autorité sur un autre rôle a de l'autorité sur un agent qui joue le deuxième rôle. Quand un agent essaie de satisfaire un but (mission) associé à son rôle, il peut déléguer une partie de ce but à un autre agent s'il a de l'autorité sur lui. Les rôles d'une organisation et les relations d'autorité qui les lient donnent ainsi naissance à une *hiérarchie de rôles* – concept très utile pour la satisfaction coordonnée de buts globaux.

Un autre modèle organisationnel qui utilise une structure similaire est le modèle OperA proposé par Dignum (Dignum 2004). Ce modèle s'étend sur trois niveaux similaires aux niveaux du modèle normatif HarmonIA présenté ci-dessus : le niveau abstrait, concret et d'implémentation. Le premier niveau décrit d'une manière abstraite les objectifs de l'organisation, objectifs qui sont ensuite

décomposés dans des tâches qui sont associées aux rôles dans les structures organisationnelles décrites au sein du niveau concret. La *structure sociale* définit les *rôles* existants, les *groupes* formés des rôles et deux graphes représentant la *hiérarchie des rôles* et les *relations de dépendance* entre ceux-ci. Les rôles sont définis dans ce modèle d'une manière similaire au modèle de rôles de (Dastani *et al.* 2003), les auteurs des deux modèles étant les mêmes. Un rôle consiste ainsi en un identifiant (son nom), un ensemble d'objectifs à atteindre (ses buts – éventuellement avec des sous-buts explicites) et d'un ensemble de droits et devoirs (normes). Les groupes sont utilisés comme dans le modèle MOISE+ pour simplifier la définition de la structure organisationnelle, par exemple en associant une norme à un groupe au lieu de l'associer à chaque rôle du groupe.

La hiérarchie de rôles est similaire à la décomposition des buts globaux en des sous-buts et elle permet ainsi l'attribution de buts organisationnels aux rôles, tandis que le graphe de dépendances contient des relations de dépendance entre rôles. Les concepts liés à la théorie de la dépendance[3] sont présentés dans le Chapitre 3 ; pour l'instant nous nous contentons de préciser que si un rôle n'a pas tous les éléments nécessaires pour satisfaire un de ses buts (par exemple parce qu'il manque une permission), il dépend d'un autre rôle pour ce but. Le niveau concret dans le modèle OperA contient aussi une structure d'interaction – son rôle est de définir des cadres d'interaction (appelés *scènes*) qui décrivent comment les interactions entre les agents doivent être effectuées. Finalement, le niveau d'implémentation spécifie comment les agents sont attachés aux rôles et comment ils suivent la structure d'interaction. L'approche proposée dans ce modèle est de lier un agent au rôle qu'il joue par un *contrat*. Un contrat spécifie les conditions dans lesquelles un agent jouera un rôle (généralement temporaires), mais aussi les sanctions envisagées si l'agent ne respecte pas les spécifications du rôle. Nous verrons par la suite les raisons pour lesquelles, dans ce modèle comme dans d'autres, il est nécessaire de mettre en place un tel contrat ainsi que des sanctions,.

2.3.3. Discussion

La notion de rôle est centrale dans le cadre de la coordination institutionnelle : les rôles sont utilisés pour décrire le comportement attendu des agents appartenant à une organisation. Des buts sont associés aux rôles et tout agent qui joue un rôle doit satisfaire ces buts – dans le cas idéal cette satisfaction des buts locaux implique la satisfaction des buts globaux de l'organisation. Pour aider les agents à satisfaire leurs buts, des structures organisationnelles telles que les hiérarchies de rôles sont souvent utilisées : des liens d'autorité entre les rôles donnent aux agents la possibilité de déléguer une partie de leurs (sous-)buts à d'autres agents sur qui ils ont de l'autorité. De plus, des normes sont aussi associées aux rôles, normes qui ont comme objectif de permettre aux agents de satisfaire les buts de leurs rôles (p.ex. : permissions d'utiliser des ressources) ou de limiter le comportement de ces agents afin d'assurer une coordination des activités locale (p.ex. : obligations de satisfaire un but, interdiction d'exécuter une action). Ces normes représentent des normes institutionnelles (les r-normes de (Tuomela 1995)) – elles sont clairement définies (les agents ciblés par une norme sont ceux qui jouent le rôle auquel la norme est associée) et elles sont issues par une autorité institutionnelle et ont des sanctions explicites associées, comme nous le verrons par la suite.

[3] Même si les dépendances entre rôles sont définies par le concepteur de l'organisation et ne sont pas calculées par les agents comme dans la théorie de la dépendance, les deux types de dépendances reste similaires.

Il est intéressant de souligner que les normes, rôles ou structures organisationnelles représentent des contraintes sur le comportement, mais aussi sur la prise de décision d'un agent. Par exemple, une relation d'autorité entre deux agents jouant des rôles est une contrainte décisionnelle, potentielle : quand l'agent "supérieur" lui délègue la satisfaction d'un but, la prise de décision de l'agent "inférieur" sera contrainte par cette relation d'autorité. Cependant, dans le même temps, la relation d'autorité est une contrainte comportementale, *établie*, aussi : le comportement d'un agent jouant un rôle doit être orienté vers la satisfaction de cette relation qui a été établie par son action de jouer le rôle.

2.4. Renforcement des normes

Si les rôles décrivent le comportement attendu des agents qui les jouent, les normes sont généralement utilisées pour contraindre ces agents à effectuer ce comportement. Cependant, comme nous l'avons précisé dans le Chapitre 1, il existe des agents autonomes par rapport aux normes, c.-à-d. des agents qui n'adoptent pas une norme ou qui ne lui obéissent pas. Il est donc nécessaire pour le concepteur d'une organisation multi-agents non seulement de définir le comportement que les agents doivent effectuer et de les contraindre par des normes de l'effectuer, mais aussi de s'assurer que les agents obéissent à ces normes et si besoin de les renforcer. Le problème devient ainsi de comment vérifier si un agent respecte ou non une norme et de quoi faire dans le cas d'une violation.

Pour qu'une coordination soit effectivement atteinte dans une organisation multi-agents, plusieurs mécanismes doivent être mis en place, en plus de la définition des rôles et des structures organisationnelles. Comme précisé par Dignum (Dignum 1999), plusieurs types d'agents spéciaux sont nécessaires afin d'assurer un comportement conforme aux normes des agents potentiellement autonomes s'exécutant au sein de l'organisation. Un premier type d'agent est celui d'agent *émetteur* de normes. De tels agents se chargent de créer de nouvelles normes ou de modifier des normes existantes. La fonction de ces agents doit être reconnue par les autres agents qui eux n'ont pas le droit de refuser l'adoption d'une norme émise par un agent émetteur. Comme nous l'avons vu, même si un agent adopte une norme institutionnelle, du fait de son autonomie, il est possible qu'il ne lui obéisse pas – ceci est généralement appelé une violation de norme et des sanctions y sont généralement associées. Avant de sanctionner les agents qui ont violé des normes, il faut détecter ces violations. C'est là ou un deuxième type d'agent est introduit : des agents *policiers*. Ils peuvent être utilisés afin de surveiller le comportement des agents et de détecter si ce comportement est conforme aux normes ou non. D'autres agents, *juges*, décident ensuite si une sanction doit être appliquée pour cette violation et l'appliquent aux agents qui ont violé la norme – nous voulons souligner que des sanctions positives (récompenses) peuvent être également envisagées pour les agents obéissant aux normes. Une fois de plus, les agents policiers et juges doivent être reconnus en tant que tels par les autres agents : les sanctions qu'ils imposent doivent être acceptées comme venant d'une autorité et non d'un agent quelconque. Avant de s'intéresser aux sanctions envisageables dans les systèmes multi-agents, nous voulons présenter brièvement quelques modèles d'institutions multi-agents qui s'intéressent à la détection des violations et au renforcement des normes.

2.4.1. Institutions électroniques

Comme nous l'avons précisé, les structures organisationnelles offrent un cadre de coordination en utilisant des rôles, des normes et des relations entre rôles, mais des mécanismes de renforcement des normes sont aussi nécessaires. Une fois qu'une structure organisationnelle est instanciée (des agents sont associés à des rôles), ces mécanismes seront utilisés pour assurer le fait que les agents jouent leurs rôles. Un modèle organisationnel doit être ainsi enrichi avec ces mécanismes qui décrivent son fonctionnement à l'instanciation pour obtenir un modèle institutionnel. Cette approche est utilisée par Vázquez-Salceda (Vázquez-Salceda 2004) qui propose le modèle institutionnel *Omni*. Ce modèle est basé sur le modèle organisationnel OperA et le modèle normatif HarmonIA (tous les deux décrits ci-dessus), auxquels une dimension ontologique est aussi rajoutée – en fonction du domaine d'application, une ontologie différente est utilisée, ce qui augmente la généralité du modèle. Le modèle Omni garde la structure sur plusieurs niveaux des modèles qui forment sa base et contient ainsi le niveau abstrait, le niveau concret et le niveau d'implémentation.

Les rôles dans ce modèle décrivent les buts que les agents qui les jouent doivent satisfaire et les normes qu'ils doivent obéir. Pour que l'institution fonctionne bien il faut que les agents ne dévient pas des spécifications des rôles. Il faut donc spécifier d'une manière formelle qu'un agent joue un rôle (qui a des caractéristiques telles que buts ou normes) et de pouvoir détecter quand le comportement d'un agent ne suit pas les caractéristiques de son rôle. La solution utilisée dans le modèle Omni est l'utilisation de la notion de *contrat* : un contrat est formé entre un agent et l'institution (généralement représentée par un autre agent), contrat concernant le rôle que cet agent jouera et dans quelles conditions (intervalle de temps, rémunération, etc.). Tant que ces conditions sont vraies, l'agent est lié par contrat pour jouer le rôle – si son comportement ne suit pas les spécifications de son rôle, l'institution considère qu'il ne remplit pas son contrat et elle peut le sanctionner. Cette approche transforme ainsi la violation d'une norme par un agent dans une violation des spécifications de son rôle, ce qui implique une violation de son contrat avec l'institution. Ceci permet de traiter toutes les violations possibles au même niveau et donc avec les mêmes sanctions.

La notion de *contrat* semble la plus utilisée dans les institutions multi-agent : à part le modèle Omni, le modèle $Moise^{Inst}$ proposé par Gâteau *et al.* (Gâteau *et al.* 2004) l'utilisant aussi. Ce modèle étend le modèle MOISE+ décrit ci-dessus, en rajoutant à ses spécifications structurelle, fonctionnelle et déontique une spécification contextuelle. Cette dernière est utilisée pour décrire le contexte dans lequel peut se retrouver un agent jouant un rôle : ce contexte influence les normes applicables et donc le comportement des agents. Les contrats sont utilisés ici pour décrire le résultat des interactions entre agents : les agents négocient et forment des contrats qu'ils doivent ensuite respecter. Des agents spéciaux sont utilisés pour *arbitrer* les contrats, c.-à-d., surveiller s'ils sont remplis ou non et prendre les mesures nécessaires selon le cas. Ainsi, même si un contrat est issu d'une interaction entre deux agents, le fait que ces agents jouent des rôles dans une institution le transforme dans une notion *institutionnelle* – il est considéré un contrat seulement s'il existe une institution qui renforce sa satisfaction.

Un point de vue similaire est pris par les auteurs de (Boella and van der Torre 2004a), qui analysent en détail le caractère institutionnel d'un contrat. Pour eux, une institution est composée des contrats créés : le fait de « signer » un contrat est un fait institutionnel, ainsi que les activités qui mènent à la

création d'un contrat (p.ex. : l'accord entre deux agents). De plus, ils analysent une organisation en utilisant des caractéristiques d'agents, telles que croyances, désirs ou buts : la création d'un contrat modifie ainsi l'état mental d'une institution.

Finalement, un autre modèle d'institution électronique est celui proposé par Esteva (Esteva 2003), qui propose aussi ISLANDER – un outil supportant la spécification de règles et de protocoles dans une institution. Les institutions créées avec ce modèle ne s'intéressent pas aux hiérarchies de rôles, leur seul but étant de réguler les interactions entre agents. Les agents négocient dans des marchés électroniques pour former des contrats et les normes institutionnelles imposent des règles à respecter pendant ces négociations (telles que des protocoles à utiliser, etc.). En ce qui concerne la violation potentielle des normes, elle n'existe pas : tout agent entrant dans une institution doit incorporer un module par l'intermèdiaire duquel il communiquera avec d'autres agents – comme ce module est fourni par l'institution, il obéit aux normes et donc les agents ne pourront jamais être autonomes par rapport à celles-ci.

Synthèse

Pour résumer, les modèles organisationnels sont souvent enrichis afin de gérer l'instanciation des structures organisationnelles, c.-à-d., l'association des agents aux rôles. Les modèles institutionnels ainsi obtenus doivent vérifier que le comportement des agents est conforme aux spécifications des rôles qu'ils jouent et de prendre les mesures nécessaires selon le cas. Comme l'objectif de ce manuscrit n'est pas de proposer un nouveau modèle organisationnel ou institutionnel, nous n'allons plus nous intéresser à la conception des institutions, mais à comment les agents qui en font partie sont affectés par les contraintes qui en sont issues.

2.4.2. Sanctions

Comme nous l'avons précisé, les institutions multi-agents essaient d'inciter les agents à obéir aux normes (ou les spécifications de leurs rôles) en définissant des sanctions et en utilisant des mécanismes d'application de ces sanctions. Dans cette section nous décrivons une ontologie des sanctions et des types de punitions, ontologie proposée par les auteurs de (Pasquier *et al.* 2005).

Les sanctions qu'un agent peut subir peuvent être caractérisées par trois dimensions : leur direction, leur type et leur style. Selon leur direction, les sanctions peuvent être divisées en *sanctions négatives* (ou *punitions*) et en *sanctions positives* (ou *récompenses*). La première catégorie est très fréquente dans les institutions multi-agents : des punitions sont associées à la violation des normes pour décourager les agents d'avoir un comportement différent de celui spécifié. Cependant, la deuxième catégorie devrait aussi être envisagée : les récompenses peuvent encourager les agents à obéir aux normes. Il existe trois types de sanctions : matérielles, sociales et psychologiques. Les *sanctions matérielles* consistent en des sanctions physiques (p.ex. : violence, actions de réparation) ou financières (p.ex. : payer une amende). Les *sanctions sociales* affectent des valeurs sociales telles que la confiance, la réputation ou la crédibilité, valeurs qui sont généralement représentées au sein d'autres agents que celui à qui s'applique la sanction et qui n'a donc aucune possibilité de refuser l'application de la sanction. Les *sanctions psychologiques* sont fréquentes dans la société humaine mais moins souvent utilisées dans les systèmes multi-agents et incluent des sanctions comme la culpabilité ou la

honte. En ce qui concerne leur style, les sanctions se divisent en des *sanctions implicites* et en des *sanctions explicites* – ces dernières sont les plus fréquentes dans les institutions multi-agents parce qu'elles ne laissent aucun doute aux agents autonomes sur les conséquences de leurs éventuelles déviations des spécifications organisationnelles.

Comme nous l'avons précisé, il ne suffit pas de définir des normes et même des sanctions associées, mais il est souvent nécessaire de mettre en place de mécanismes d'application des sanctions quand une norme est violée (sanctions négatives) ou quand elle est satisfaite (sanctions positives). Pasquier *et al.* identifient plusieurs types de *politiques de punition* – paradigmes utilisés pour décider si une sanction doit être appliquée ou pas, quand et comment. Parmi ces types, les plus importantes et utiles dans les systèmes multi-agents sont la dissuasion et la rétribution. La *dissuasion* est un paradigme de punition qui spécifie que le but d'une sanction est de prévenir une violation potentielle en imposant des sanctions lourdes et explicites pour décourager les agents. La *rétribution* considère qu'une violation doit être réparée par une pénalité du même montant que le préjudice provoqué par la violation. L'utilisation des punitions du premier type décourage les agents d'avoir des comportements autonomes, ce qui parfois est envisageable (voir ci-dessous), tandis que le deuxième type de punition, où les sanctions sont égales aux préjudices, donne aux agents des stimuli socialement corrects pour prendre des précautions. Mais pour que les sanctions de type rétribution fonctionnent bien dans un système multi-agents, une *hypothèse de responsabilité stricte* (strict liability en anglais) est nécessaire : les agents fautifs sont identifiés et aucun d'entre eux ne peut échapper aux sanctions correspondantes (la même approche fonctionne dans le cas des sanctions positives aussi).

Le concepteur d'une institution multi-agents doit choisir les caractéristiques des sanctions utilisées et les politiques de punitions les plus pertinentes pour son système – même si nous ne nous intéressons pas directement à cette problématique, elle influence partiellement le raisonnement des agents. Quand un agent doit décider d'avoir ou non un comportement autonome par rapport à une norme, il doit évaluer les conséquences que ce comportement aura, incluant ici le type de sanctions probables, leur montant, s'il peut y échapper ou pas, etc. Ce problème devient plus compliqué par le fait que, selon Castelfranchi (Castelfranchi 2005), il existe des violations de normes qui ne doivent pas être punies. Une *violation fonctionnelle* apparaît quand un agent viole intentionnellement une norme parce qu'il considère que son comportement satisfait mieux les objectifs de l'organisation que le comportement spécifié par la norme. Même si pour un agent institutionnel il n'est pas facile de détecter si une violation est motivée ou non par l'intention de satisfaire un but du système, nous voyons ici un exemple de l'apport de l'autonomie des agents à l'amélioration du fonctionnement d'un système : un agent doit être incité à obéir aux normes, mais il ne doit pas être complètement contraint par les sanctions associées afin de lui laisser un choix qui peut s'avérer utile au système[4].

[4] Une direction de recherche intéressante est de mettre en place un mécanisme qui permet aux agents de *justifier* pourquoi ils ont violé une norme et envisager ainsi éventuellement de ne pas les sanctionner si la raison de la violation était motivée par l'amélioration du système.

2.5. Synthèse

Pour coordonner les agents au sein d'un système, l'approche institutionnelle fait appel à plusieurs concepts, tels que les rôles ou les normes. Généralement les agents sont supposés jouer des rôles, rôles qui sont structurés en hiérarchies et/ou qui appartiennent à des groupes. Un rôle contient une liste de buts à satisfaire par les agents qui le joue – si ces buts sont satisfaits (d'une manière coordonnée avec d'autres agents), alors les buts globaux de l'organisation seront également atteints. Pour aider les agents à mieux satisfaire les buts de leurs rôles, différents liens entre rôles existent, comme par exemple le lien d'autorité entre deux rôles qui permet à un agent de déléguer une partie de ses (sous-)buts à un autre. Des permissions sont également associées aux rôles pour donner accès à des ressources aux agents qui les jouent et les aider ainsi à mieux satisfaire leurs buts. A part des permissions, d'autres types de normes institutionnelles peuvent être associés aux rôles : des obligations de satisfaire des buts ou des interdictions d'effectuer des actions (utiliser des ressources, poursuivre des buts, etc.). Tous ces éléments spécifient le comportement attendu d'un agent qui joue un rôle. Les institutions multi-agents utilisent également des mécanismes pour motiver les agents à ne pas dévier de ce comportement (p.ex. : des récompenses pour les agents qui obéissent aux normes ou des punitions pour ceux qui les violent).

Le fait d'appartenir à une organisation dans laquelle il joue un ou plusieurs rôles, impose à un agent plusieurs contraintes sur sa prise de décision et son comportement, limitant ainsi son autonomie. Considérons par exemple un rôle défini par des buts à satisfaire, des normes à obéir et des relations d'autorité vers et envers d'autres rôles. Ces buts représentent des contraintes institutionnelles et comportementales : elles limitent le comportement de l'agent en l'obligeant de les poursuivre. Les normes déjà adoptées par un agent représentent des contraintes comportementales : le comportement d'un agent doit être conforme à ses normes, si non il est considéré autonome par rapport à elles. Mais de nouvelles normes peuvent être à tout moment associées à un rôle par un agent institutionnel ou des normes peuvent devenir actives si elles dépendent du contexte : le fait de jouer un rôle impose aussi des contraintes décisionnelles à un agent qui doit adopter des normes quand un autre agent le décide. D'autres contraintes décisionnelles sont issues des relations d'autorité : si un agent a de l'autorité sur un autre pour un but, la décision de ce dernier concernant le but est limitée par cette relation – chaque fois que l'autre le voudra, il devra adopter ce but et essayer de le satisfaire. Nous voulons également souligner ici la nature duale de ces contraintes : même si une d'entre elles limite la prise de décision d'un agent (et sont donc potentielles), parce qu'elles sont attachées à des rôles et que les agents s'engagent à jouer les rôles (p.ex. : avec des contrats), ces contraintes deviennent aussi des contraintes établies – si le comportement de l'agent ne les suit pas, il est considéré autonome. Dans le chapitre suivant nous verrons d'autres types de contraintes, issues des interactions entre agents, qui sont à la fois potentielles et établies.

3. Coordination et contraintes interpersonnelles

L'objectif de cette thèse n'est pas de proposer un nouveau modèle de coordination (interpersonnel ou institutionnel), mais d'analyser les implications, sous la forme des contraintes imposées, que l'utilisation d'un tel modèle a sur l'autonomie d'un agent. Nous n'essayons donc pas de proposer un nouveau modèle de coordination, mais nous nous intéressons à différents modèles existants, afin d'analyser des possibilités de représentation de ces contraintes. Dans la première partie de ce chapitre nous allons analyser à l'aide d'un cadre de coordination très connu, le modèle GPGP/TAEMS, la problématique de la coordination non-institutionnelle dans les systèmes multi-agents et le concept d'interaction entre agents. Cette analyse nous permettra d'identifier l'existence de deux types de contraintes qu'une telle coordination impose aux agents : les contraintes qui mènent les agents à interagir (pourquoi ont-ils besoin de le faire ?) et les contraintes issues de ces interactions (qu'est-ce qu'ils doivent faire après ?). D'un coté, pour le deuxième type de contrainte, le modèle d'engagements sociaux semble être la représentation la plus utilisée – nous détaillons ce modèle, ses utilisations et nous décrivons ainsi les implications de l'utilisation des méta-engagements (ou politiques sociales). D'un autre coté, la théorie de la dépendance explique et donne un cadre formel de représentation du premier type de contrainte – nous présentons et discutons cette théorie. Nous présentons ensuite une théorie des sciences sociales qui l'étend et qui s'intéresse aux pouvoirs individuels et aux relations de pouvoirs entre agents : la théorie du pouvoir social. Nous finirons par une discussion sur la manière dont ces théories peuvent être utilisées pour décrire les contraintes interpersonnelles qui limitent la prise de décision et le comportement des agents.

3.1. Différentes approches pour coordonner des agents

Le terme *coordination* est utilisé avec plusieurs sens dans la communauté multi-agents. Comme nous l'avons vu au chapitre 1, les approches existantes pour la coordination dans les systèmes multi-agents peuvent être divisées en deux catégories, appelées coordination *objective* et coordination *subjective* dans (Schumacher 1999) et (Omicini 2000). La coordination objective utilise des structures externes aux agents, structures qui peuvent être les structures normatives, organisationnelles ou institutionnelles présentées dans le chapitre précédent, mais aussi des structures interpersonnelles, telles que des langages de coordination (p.ex. : CooL (Barbuceanu and Fox 1995)) ou des cadres (frameworks) de coordination (p.ex. : Tucson (Omicini and Zambonelli 1998)). La coordination subjective représente les efforts de chaque agent pour coordonner ses activités avec les autres et prend souvent la forme d'interactions entre agents. Dans la suite de cette section nous présentons brièvement les deux types de coordination, en décrivant un exemple d'un cadre de coordination et en discutant les interactions au sein des systèmes multi-agents. Ceci nous permettra ensuite d'analyser les implications de ces types de coordination sur les contraintes imposées aux agents et donc sur leur autonomie.

3.1.1. Cadres (framework) de coordination – le modèle TAEMS/GPGP

Parmi les plusieurs cadres de coordination existants et utilisés par la communauté multi-agent, nous avons choisi de présenter ici le modèle GPGP et le modèle TAEMS associé (Lesser *et al.* 2004). Ce choix est justifié par la généralité et la richesse de ce modèle, qui nous permettent ainsi d'illustrer la

problématique des cadres de coordination et d'utiliser ce modèle comme exemple pour illustrer différents aspects par la suite.

La planification globale partiale généralisée (GPGP) et sa représentation associée de structures hiérarchiques de tâches (TAEMS) ont été proposées comme un cadre indépendant du domaine pour la coordination des activités de groupes d'agents travaillant ensemble pour la satisfaction des buts de haut-niveau. L'objectif de GPGP est de maximiser l'utilité globale obtenue par un groupe d'agents qui effectuent des tâches, indépendantes ou non, sans avoir nécessairement une image globale de la coordination. La clé de cette approche est la décomposition des buts de haut-niveau en sous-buts en utilisant les structures de tâches TAEMS. Chaque agent est responsable de quelque sous-buts et essaye de les satisfaire – il doit cependant coordonner ses activités avec d'autres agents parce que des relations peuvent exister entre ces buts. Par exemple, la satisfaction d'un but peut faciliter la satisfaction d'un autre et donc leur bon ordonnancement pourra être bénéfique pour le système. Dans un autre cas, la satisfaction d'un but par un agent peut empêcher la satisfaction d'un autre but par un autre agent et donc les agents doivent interagir pour privilégier la satisfaction d'un seul des buts.

TAEMS est un formalisme qui permet d'exprimer cette décomposition d'un but dans des sous-buts qui peuvent être décomposés à leur tour jusqu'à un niveau contenant des actions élémentaires peuvent être exécutées. Un agent qui planifie de cette manière obtient ainsi une arborescence similaire à des arbres et/ou. Le terme *tâche* est utilisé pour représenter une telle arborescence (un plan partiel ou complet pour satisfaire un but), un but ou sous-but et même une action. Les agents manipulent donc des tâches qui ont des caractéristiques telles qu'une durée d'exécution, une échéance, une utilité probable, etc. De plus, le formalisme TAEMS permet l'utilisation de plusieurs fonctions qui calculent l'utilité d'une tâche (but) à partir de l'utilité de ses sous-tâches. Des fonctions de type somme (l'utilité finale est la somme des utilités de sous-buts), séquence (l'utilité finale est égale à l'utilité du dernier sous-but), max (l'utilité finale est égale à la plus grande utilité de sous-buts), etc., peuvent être utilisées.

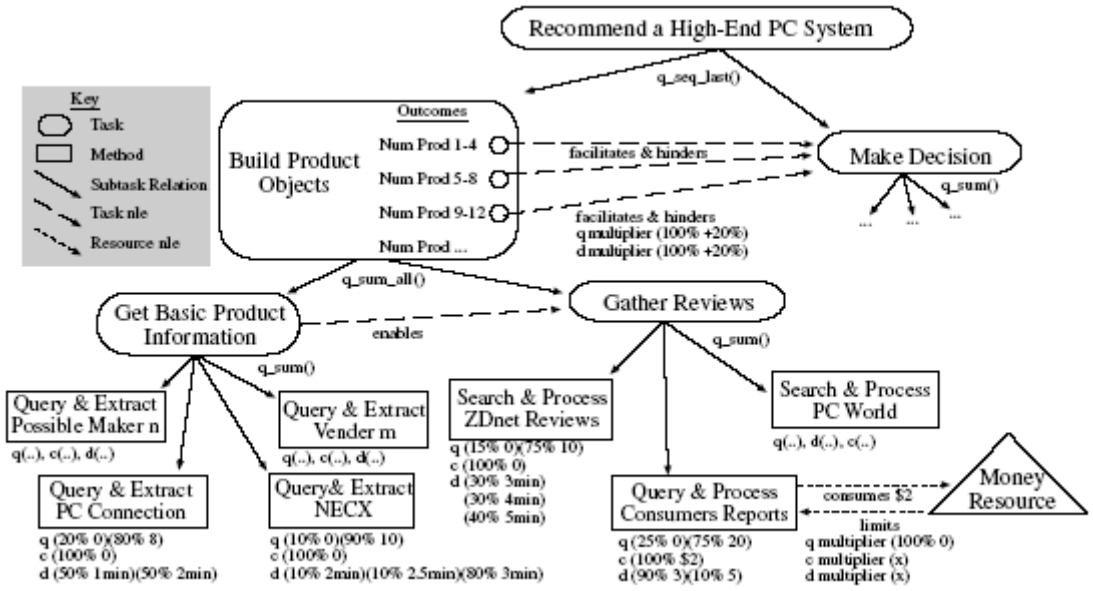


Figure 3.1. Un exemple d'une structure de tâches TAEMS

Le grand apport du modèle TAEMS est l'identification et l'utilisation de relations existantes entre les tâches. Du fait de durées trop grandes par rapport aux échéances, des utilisations de mêmes ressources ou de contraintes dépendantes du domaine, les tâches existantes sont interdépendantes. Plusieurs relations entre tâches peuvent exister, telles que *enables/disables* (l'exécution d'une tâche (ne) permet (plus) l'exécution d'une autre) ou *hinders*/*facilitates* (l'exécution d'une tâche rend difficile/aide à l'exécution d'une autre). De plus, des relations entre une tâche et une ressource peuvent aussi exister, telles que *produces/consumes* (une tâche produit/consume une ressource) ou *limits* (une ressource est critique pour l'exécution d'une tâche). Un exemple, issu du domaine de recherche d'information, d'une structure de tâche TAEMS avec les fonctions de calcul d'utilité (*quality* – notée *q* dans la figure) et des relations entre tâches est présenté dans la Figure 3.1.

Le modèle GPGP considère que la coordination des activités des agents est nécessaire pour assurer une satisfaction optimale des buts de haut-niveau en évitant que les agents effectuent des tâches non-nécessaires ou non-importantes, qu'ils attendent des ressources utilisées ou produites par d'autres agents, qu'ils soient surchargés pendant que d'autres sont en sous-charge ou qu'ils ratent les échéances des tâches importantes. Les agents doivent donc coopérer pour identifier les relations existantes entre leurs tâches et éviter que ces situations arrivent. Des structures de tâches sont ainsi échangées entre les agents afin de pouvoir identifier ces relations, ce qui représente une caractéristique du modèle GPGP. Les structures échangées entre agents représentent des plans partiels (connus au niveau local) et cet échange permet aux agents de construire peu à peu un plan global.

L'utilisation de ce modèle de coordination implique donc une interaction entre les agents, interaction qui utilise le modèle TAEMS pour échanger des structures de tâche. Ce modèle peut être aussi utilisé dans le raisonnement interne des agents, qui peuvent planifier en terme de structures de tâche, mais ceci n'est pas obligatoire. Les interactions entre agents ont comme but d'identifier des relations entre les tâches, mais aussi former des accords entre les agents concernant ces relations. Le modèle GPGP considère que, à la fin de leurs interactions, les agents forment des *engagements* les uns envers les autres : ces engagements représentent les tâches avec leurs caractéristiques que les agents s'engagent d'effectuer afin de coordonner leurs activités avec les autres. Par exemple, si un agent ne peut effectuer sa tâche qu'après qu'un autre ait effectué la sienne, ils interagissent et identifient une relation de type *enables* entre les tâches. Ensuite, le deuxième agent s'engage à effectuer sa tâche avec une échéance qui permettra au premier d'effectuer la sienne avant son échéance. Bien évidemment, des algorithmes de planification sont utilisés pour planifier en fonction des durées, échéances, utilités des tâches, etc.

3.1.2. Interactions entre agents

Le cadre de coordination GPGP présenté ci-dessus s'intéresse surtout au processus de planification et à la représentation de plans. Une condition nécessaire pour son utilisation réside dans la capacité des agents à interagir les uns avec les autres. *L'interaction est donc nécessaire pour arriver à une coordination entre agents, tandis que l'utilisation d'un cadre de coordination est nécessaire pour une coordination efficace*. Plusieurs systèmes multi-agents existants montrent qu'une coordination des activités des agents membres peut être atteinte sans cadres de coordination ou même des raisonnements très complexes. C'est notamment le cas du système MANTA (Drogoul and Ferber 1992), dans lequel des agents réactifs avec le comportement inspiré de celui des fourmis arrivent à

coordonner leurs activités en laissant des traces dans l'environnement. Ceci est un exemple de système dans lequel les agents interagissent à travers leur environnement.

Pour résoudre des problèmes plus complexes que la coordination des fourmis, des agents plus intelligents sont souvent nécessaires et avec eux un modèle d'interaction plus riche. Le modèle le plus utilisé par la communauté multi-agents est de loin la communication par l'envoie de messages, messages qui utilisent souvent le paradigme des *actes de langage*, comme par exemple le langage de communication entre agents appelé FIPA-ACL (FIPA 2005) proposé par l'organisation de standardisation FIPA[5]. Ce langage spécifie pour chaque message envoyé l'agent qui l'envoie, le(s) destinataire(s), l'acte de langage utilisé (tels que *request, inform, propose, accept*, etc.), le contenu du message et d'autres champs qui essayent d'assurer le fait que le destinataire est en mesure de comprendre le message. Une sémantique est également associée à chaque acte de langage pour permettre aux agents de comprendre le cadre cognitif d'un agent qui l'a émis. Par exemple, la sémantique d'un message *inform* est que l'agent qui l'envoie *a la croyance* que le contenu du message est vrai.

Cette sémantique est à la base de la plupart des critiques du langage FIPA-ACL, parce qu'elle se réfère à l'état cognitif *interne* d'un agent qui envoie un message, état qui n'est pas accessible par un observateur externe. Dans le cadre du même exemple, un agent qui reçoit un message *inform* sera obligé de considérer que l'agent qui l'a envoyé a une croyance vis-à-vis du contenu du message, ce qui n'est pas forcément vrai. Des agents menteurs peuvent exister et envoyer des messages sans respecter la sémantique associée, ce qui rend difficile d'arriver à un comportement global cohérent. Même si l'approche FIPA ne s'intéresse pas à ce problème, elle essaie d'améliorer la coordination dans un système en imposant des règles sur l'émission de messages. Des protocoles d'interaction sont ainsi utilisés pour spécifier quel agent peut envoyer quel message à un moment et à qui, ce qui permet la création des conversations cohérentes entre des agents, agents qui peuvent ainsi mieux coordonner leurs activités.

3.1.3. Discussion

Pour arriver à un comportement global cohérent dans un système sans utiliser de concepts institutionnels, les agents doivent interagir pour coordonner leurs activités. Indifféremment de comment cette interaction est effectuée, par l'intermèdiaire de l'environnement ou par envoi des messages, en utilisant des protocoles ou dans un cadre de coordination, cette coordination interpersonnelle impose des contraintes sur la prise de décision et sur le comportement des agents. Si nous considérons par exemple le modèle GPGP, les agents doivent interagir pour identifier des relations entre leurs tâches, relations qui leur permettront de mieux coordonner leurs activités, c.-à-d., de mieux satisfaire les buts de haut-niveau. Dans le cadre du même modèle, suite aux interactions, les agents s'engagent à effectuer des tâches avec des durées et échéances différentes.

Nous voyons ainsi apparaître deux catégories des contraintes : les contraintes qui poussent les agents à interagir (à se coordonner) et les contraintes issues des interactions. Comme nous l'avons discuté dans le Chapitre 1, ces contraintes sont directement liées à la notion d'autonomie qui nous intéresse dans cette thèse – il est donc important de mieux les analyser. La deuxième catégorie représente des

[5] Foundation for Intelligent Physical Agents – FIPA. www.fipa.org

contraintes comportementales qui limitent le comportement d'un agent – par exemple, dans GPGP, il a à effectuer une partie d'un plan commun avec d'autres agents et donc son comportement sera orienté vers l'exécution de ce plan. La plupart des modèles existants utilisent d'une manière plus ou moins implicite le paradigme d'*engagements sociaux* (par exemple, dans GPGP, un agent devient *engagé* à la fin d'une interaction). Nous présentons par la suite ce paradigme en montrant ses différents aspects et ses implications pour la représentation des contraintes qui limitent l'autonomie d'un agent.

Quant à la première catégorie des contraintes, elle contient des contraintes décisionnelles qui limitent la prise de décision d'un agent : par exemple, dans GPGP, il ne peut pas satisfaire tout seul les buts de haut-niveau et il décide ainsi d'interagir avec d'autres. Une théorie qui essaie d'analyser et d'identifier les facteurs qui poussent un agent à interagir avec d'autres agents, à coordonner ses activités avec eux, est la *théorie de la dépendance*, que nous présenterons plus tard dans ce chapitre ainsi que son extension – la *théorie du pouvoir social*.

3.2. Engagements sociaux

Dans cette section nous décrivons le modèle d'engagements sociaux et les différentes approches existantes pour sa représentation ou utilisation. Comme il existe plusieurs types d'engagement utilisés dans les systèmes multi-agents, nous commençons par décrire ses différents types, pour nous concentrer ensuite sur les engagements sociaux. Apres avoir présenté différentes propriétés qui les caractérisent, nous présentons également un type spécial d'engagement social : les politiques sociales. Nous finirons par une discussion de l'importance de ces concepts pour notre travail et pour la représentation des différents types de contraintes interpersonnelles.

3.2.1. Plusieurs types d'engagements dans les systèmes multi-agents

Depuis plusieurs années, la notion d'*engagement* a été reconnue par la communauté multi-agent comme un concept important, avec de nombreuses définitions proposées et diverses approches pour son utilisation au sein des systèmes multi-agents. Dans (Castelfranchi 1995a), Castelfranchi met en évidence les différences existantes entre plusieurs types d'engagements présents dans la littérature, tels que les engagements individuels, sociaux ou collectifs.

Les *engagements individuels* représentent le type d'engagements le plus basique. Ils font référence à la relation entre un agent et une de ses actions : un agent est dit *engagé* pour une action s'il a l'intention d'effectuer cette action (Cohen and Levesque 1990). Ce type d'engagement est très similaire au concept d'intention dans le modèle BDI (Georgeff 1987), modèle très connu dans la communauté multi-agent et que nous décrivons dans le Chapitre 4. Ce type d'engagement est présent dans la plupart des théories d'agents, même si parfois implicitement ou avec un autre nom. Par exemple, Castelfranchi argumente que le terme *interne* soit utilisé et non *individuel*, parce que nous pouvons aussi considérer les engagements internes d'un groupe d'agents, c.-à-d., les intentions du groupe. D'un autre coté, Singh (Singh 1991) propose l'utilisation du terme *engagement psychologique* pour l'opposer à la notion d'engagement social. Les *engagements sociaux* représentent un concept relationnel qui lie deux agents : un agent est dit *engagé envers un autre agent*. Cette relation entre agents est très importante pour l'obtention de la coordination dans un système multi-agents – vue son importance, nous détaillerons par la suite. Pour l'instant nous nous contentons de mentionner que ce

type d'engagement est similaire à la notion de *contrat* que nous avons décrite dans le Chapitre précédent. Castelfranchi souligne aussi le fait qu'un engagement social d'un agent envers un autre représente plus qu'un engagement individuel du premier agent et connu par le deuxième agent et plus qu'un engagement interne d'un groupe d'agents.

Les *engagements collectifs* représentent tout simplement les engagements internes (individuels) d'un collectif d'agents. Même s'il existe plusieurs théories des engagements individuels (ou notions similaires), le passage au collectif n'est pas évident à faire. Certes, vu de l'extérieur, un groupe peut être considéré comme une seule entité avec des caractéristiques diverses. L'une de celles-ci est notamment que le groupe est engagé pour effectuer une action, c.-à-d., le groupe a l'intention de l'effectuer. Mais dans l'intérieur du groupe, il n'est pas forcément clair qui parmi ses membres a l'intention d'effectuer l'action ou quels membres sont conscients de l'existence d'un tel engagement collectif. Plusieurs travaux dans la littérature s'intéressent à ces problèmes, nous pouvons citer ici l'approche de Jennings (Jennings 1993) qui utilise la notion d'*engagement commun* (joint commitments) ou l'approche de Dunin-Keplicz et Verbrugge (Dunin-Keplicz and Verbrugge 2002) qui utilisent la notion d'*intention collective*.

Nous voulons mentionner ici une caractéristique intéressante de la notion d'engagement collectif : un tel engagement n'est pas seulement une somme d'engagements individuels, mais il implique l'existence de plusieurs engagements sociaux. Premièrement, un engagement collectif n'apparaît pas simplement parce que deux (ou plusieurs) agents ont la même intention (engagement individuel) et ils sont conscients de l'intention de l'autre. Castelfranchi utilise l'exemple de deux personnes qui attendent un bus au même arrêt : chaque agent a l'intention de faire un signe au chauffeur d'arrêter le bus et chaque agent sait que l'autre agent a cette intention. Cependant, nous n'appelons pas ces deux agents une *équipe* avec un engagement *collectif*. Quelque chose manque : un accord implicite ou explicite concernant une activité commune, accord basé sur la conscience de l'existence d'une relation entre eux.

Cet accord entre agents qui semble nécessaire pour la notion d'engagement collectif peut être représenté comme un engagement social. Castelfranchi argumente que, au moins dans les systèmes coopératifs, les engagements collectifs peuvent être représentés sous la forme d'ensembles d'engagements sociaux et individuels. Cette représentation se base sur le fait qu'un engagement collectif implique l'existence d'un engagement social de chaque membre du collectif *vers le collectif* : l'existence des nombreux engagements sociaux réciproques entre les membres (qui fait quoi) et l'existence des engagements individuels de chaque membre pour arriver à la satisfaction de l'engagement collectif.

Synthèse

Même si la notion d'engagement individuel n'est pas forcement liée à la coordination dans un système (elle représente un élément *interne* à l'agent), les deux autres types d'engagements sont essentiels pour la coordination au sein d'un système multi-agent. Ces engagements représentent des contraintes imposées au comportement d'un agent, soit par le fait qu'il s'est engagé vers un autre agent, soit par le fait que le groupe auquel il appartient a l'intention d'effectuer une action. Cependant, dans notre travail nous ne nous intéressons pas directement aux engagements collectifs, vu que, comme

Castelfranchi l'argumente, ils ne sont tout simplement qu'une collection d'engagements sociaux. Dans ce manuscrit nous nous intéressons donc aux engagement sociaux qui, comme nous le verrons plus tard, peuvent être mis en relation avec des concepts institutionnels présentés dans le chapitre précédent.

3.2.2. Caractéristiques des engagements sociaux

Il existe de nombreux modèles d'engagements sociaux proposés dans la littérature. Les différences entre ces modèles sont justifiées par les différentes utilisations de ce concept, telles que le domaine des jeux de dialogue à base d'engagements sociaux (Maudet and Chaib-draa 2002), (Morge 2005), pour donner un exemple. Dans ce manuscrit nous nous appuyons sur les travaux de Singh, qui a été parmi les premiers à introduire le concept d'engagement social dans la communauté multi-agent. Nous analysons donc ce concept du point de vue de Singh (Singh 1999), en discutant selon le cas d'autres approches complémentaires.

3.2.2.1. Eléments d'un engagement social

Débiteur et créditeur

La caractéristique principale d'un engagement social est qu'il représente une relation entre deux agents : *un agent est engagé envers un autre*. Ces deux agents sont appelés respectivement le *débiteur* et le *créditeur* : l'agent débiteur a un engagement envers l'agent créditeur. Il est important de souligner ici que l'agent créditeur ne représente pas forcément le *bénéficiaire* de l'engagement, mais seulement l'agent envers lequel l'engagement est orienté. L'agent ou l'entité qui bénéficie de l'engagement peut être représenté explicitement, mais ceci n'est pas obligatoire : généralement la notion de bénéficiaire d'un engagement dépend du problème et n'est pas utilisée dans la définition de l'engagement.

Objet

Au delà du débiteur et du créditeur, un engagement social est aussi défini par son *objet* : *un agent est engagé envers un autre pour quelque chose*. Initialement, l'objet d'un engagement a été considéré comme une action (un agent s'engage à effectuer une action) voir par exemple (Singh 1991). Dans les approches plus récentes (p.ex. : (Singh 1999)), l'objet d'un engagement social est généralement représenté comme *une condition à satisfaire*. Ceci permet une représentation plus générale, parce qu'une condition peut représenter une action à effectuer, un but à satisfaire ou même autre chose. Un agent devient donc engagé envers un autre agent pour la satisfaction d'une condition-objet de l'engagement.

Contexte ou témoin

Il existe un autre élément nécessaire dans la définition d'un engagement social – l'élément qui donne à un tel engagement son caractère social. Le *contexte* d'un engagement représente selon Singh les normes ou conventions existantes dans le groupe dans lequel l'engagement est créé. Cet élément est appelé par Castelfranchi (Castelfranchi 1995a), le *témoin* de l'engagement social : l'agent devant lequel l'engagement a été créé. Sans l'existence de ce témoin, l'engagement n'est pas un engagement vraiment « social » , vu que ce terme dénote généralement un concept reliant au moins trois agents

(Castelfranchi 2005). Cet élément est parfois ignoré dans les différentes approches existantes où il est réduit à d'autres notions, par exemple les sanctions imposées. C'est notamment le cas du modèle de Pasquier (Pasquier *et al.* 2005), qui remplace le composant *témoin* d'un engagement social avec deux sanctions associées à l'engagement. Une sanction est imposée à l'agent débiteur s'il ne remplit pas son engagement, et une est imposée à l'agent créditeur s'il annule l'engagement. Sans entrer dans plus de détails, nous voulons souligner que dans le modèle de Pasquier le témoin ou le contexte ne sont plus nécessaires : tout engagement créé est public et les sanctions possibles sont explicites et font partie de l'engagement.

Condition de validité

Dans l'approche de Singh, un agent est engagé envers un autre pour satisfaire une condition-objet de l'engagement. Pasquier (Pasquier *et al.* 2005) divise cet objet d'un engagement en deux parties : l'objet-même de l'engagement (p.ex. : une action à effectuer) et la condition de validité de l'engagement, qui prend souvent la forme d'une échéance. Un agent est donc engagé envers un autre pour un objet (une condition à satisfaire), mais avec une condition de validité de l'engagement. Quand cette condition n'est plus valable (p.ex. : l'échéance est dépassée), le débiteur n'est plus engagé.

Les éléments décrits ci-dessus représentent les composants de base d'un engagement social : l'agent débiteur est engagé envers l'agent créditeur pour la réalisation d'un objet dans un contexte (devant un témoin), une condition de validité de l'engagement pouvant aussi être utilisée. D'autres éléments peuvent être rajoutés à ceux-ci, en fonction du domaine d'application. Par exemple, dans le cadre des jeux de dialogue, où un engagement est créé quand un agent envoie un message, le temps de la création de l'engagement peut être important et donc utilisé explicitement dans le modèle d'engagement (Muller 2006). En ce qui concerne la représentation formelle d'un tel engagement, il peut être considéré simplement un tuple composé de toutes les caractéristiques mentionnées ci-dessus. Cependant, l'approche qui se retrouve dans la plupart des travaux est de représenter les engagements sociaux dans la logique des prédicats, ce qui permet aux agents de raisonner plus facilement sur les engagements qu'ils ont envers d'autres.

3.2.2.2. Etat d'un engagements social

Singh décrit plusieurs opérations possibles sur un engagement social : la création, la satisfaction (ce qu'il appelle la « libération d'un agent de son engagement »), l'annulation, la disparition (ce qu'il appelle « délier un agent de son engagement »), la délégation et l'attribution d'un engagement. L'opération de création crée un engagement et peut être effectuée explicitement par l'agent débiteur (il s'engage envers un autre) ou implicitement quand une condition est remplie (par exemple, un engagement est crée automatiquement suite à une action effectuée). Un agent peut satisfaire un engagement en satisfaisant la condition-objet de l'engagement ou un engagement peut être annulé par le créditeur. Un engagement peut aussi disparaître, sans être ni satisfait ni annulé. Finalement, le débiteur d'un engagement peut être changé avec l'opération de délégation ou le créditeur peut être changé avec l'opération d'attribution.

Même si Singh ne le décrit pas explicitement, l'existence de ces opérations fait référence au fait qu'il existe plusieurs états possibles d'un engagement. Pasquier (Pasquier *et al.* 2005) identifie ces états et

propose des opérations avec les engagements en fonction de ces états (Figure 3.2). Dans son approche, un engagement peut se retrouver dans un des états suivants : *inactif, actif, violé, satisfait* ou *annulé*. Un engagement passe de l'état *inactif* (état implicite) dans l'état *actif* à l'aide de l'opération de *création.* Un engagement actif peut devenir *violé* (opération de *violation* – sa condition-objet ne peut pas être satisfaite) ou *satisfait* (opération de *satisfaction* – sa condition-objet a été satisfaite). Un engagement actif peut aussi être transformé dans un autre engagement (opération de *actualisation/délégation*) ou il peut être *annulé* (opération d'*annulation*). Ce modèle utilise cet état *annulé* pour dénoter tout état final d'un engagement : une fois un engagement violé ou satisfait, il passe dans cet état ce qui signifie qu'il n'est plus considéré dans le raisonnement de l'agent.

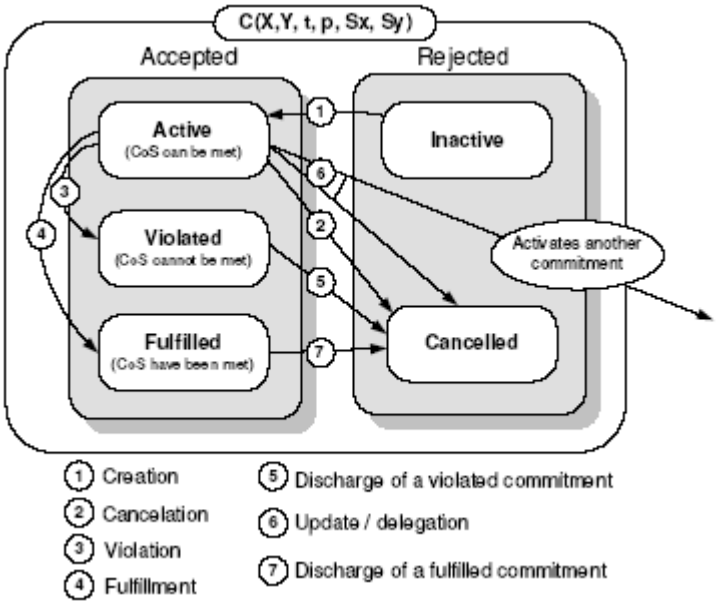


Figure 3.2. Etats possibles d'un engagement social

La définition des états possibles des engagements et des opérations associées permet une meilleure manipulation de ces engagements. Généralement il est considéré que chaque agent garde un *agenda* de ses engagements pour qu'il puisse raisonner dessus. Mais une liste de tous les engagements existants et de leurs états doit être aussi gardée par des agents autres que les débiteurs ou créditeurs de l'engagement pour vérifier leur état. Singh propose l'utilisation de *tableaux d'engagements* (commitment stores) qui contiennent les engagements actifs et qui sont constamment vérifiés pour détecter des changements d'états de ses engagements. C'est ici que l'élément de témoin ou de contexte de l'engagement devient important : quand un engagement est créé devant un témoin, le rôle de celui-ci est d'assurer le fait que l'engagement devient public, c'est-à-dire, d'autres agents peuvent savoir que cet engagement est actif et peuvent également effectuer des actions quand son état change. Notamment, des sanctions sont généralement associées au fait de violer ou de satisfaire un engagement, sanctions qui peuvent être explicitement définies dans l'engagement ou des sanctions implicites ou pas connues *a priori*. Nous voulons souligner que le problème de comment imposer des sanctions aux agents autonomes qui violent leurs engagements est similaire au renforcement des normes dans des institutions (voir Chapitre 2),. Le modèle d'engagements sociaux et des tableaux d'engagements ont l'avantage de rendre visible et explicite la violation d'un engagement (le changement de son état).

3.2.2.3. Politiques sociales

Généralement, l'état d'un engagement social change suite à une action effectuée par un agent ou à l'absence de cette exécution. Par exemple, un engagement devient *satisfait* quand son objet est satisfait (s'il dénote une action à effectuer, quand l'action est effectuée) ou il devient *violé* quand l'échéance associée est dépassée sans que l'objet soit satisfait. Cependant, la création d'engagements sociaux est une opération plus délicate. Un agent peut s'engager d'une manière proactive, c.-à-d., créer un engagement envers un autre agent de sa propre initiative (l'accord du deuxième agent est souvent nécessaire) ou créer un engagement social quand un autre agent le lui demande. Etant donné l'autonomie des agents, un problème apparaît : comment garantir le fait qu'un agent s'engage quand un autre le lui demande ?

Singh (Singh 1999) utilise le concept de *politique* (policy) qui représente une expression conditionnelle sur des engagements ou sur des opérations avec les engagements. Un exemple d'une telle politique est la politique d'un agent de créer un engagement social de payer une amende chaque fois qu'il a violé un autre engagement. Cette politique peut être interne à l'agent ou elle peut être externe – reconnue socialement. Dans ce dernier cas Singh parle de *politiques sociales*, concept qu'il analyse en détail pour tirer la conclusion qu'il ressemble à celui de norme (voir Chapitre 2).

Une telle politique sociale est représentée comme un engagement social qui a comme objet une expression logique contenant des engagements sociaux et des opérations associées. Par exemple, Singh introduit une nouvelle opération sur les engagements sociaux, l'opération de *demande* (request) qui signifie qu'un agent demande à un autre de créer un engagement social envers lui. Un exemple de politique sociale est celle qui a comme objet l'expression que toute opération de demande doit être suivie de l'opération de création. Un agent qui a cette politique sociale est engagé (envers un agent quelconque) de créer un engagement social à chaque fois qu' il lui est demandé de le faire. Les politiques sociales peuvent être donc représentées comme des engagements de niveau supérieur (meta-engagements) qui ont comme objet des expressions avec des engagements sociaux simples (de niveau inférieur). Cette représentation uniformisée des engagements et des politiques sociales est très intéressante et permet par exemple le traitement des politiques sociales (concept similaire aux normes) de la même manière que les engagements sociaux : gestion d'un tableau d'engagements, états de violation ou de satisfactions, sanctions associées, etc.

3.2.3. Discussion

Une des caractéristiques des plus intéressantes des engagements sociaux est qu'ils offrent une description explicite du comportement attendu d'un agent. Quand un agent s'engage envers un autre agent à satisfaire une condition, les conséquences sur le comportement de l'agent deviennent claires : il doit satisfaire son engagement ou il sera sanctionné. De plus, si cet engagement est inséré dans un tableau d'engagements, d'autres agents peuvent être conscients de l'engagement pris et de son état (s'il a été violé ou rempli, par exemple). Cette caractéristique fait que ce paradigme est attirant pour la communauté jeux de dialogue, où un modèle qui explicite les effets des messages échangés entre agents est nécessaire. Ces approches considèrent généralement (voir (Maudet and Chaib-draa 2002) ou (Morge 2005), par exemple) qu'un engagement social est créé chaque fois qu'un message est envoyé par un agent : l'agent devient engagé envers le(s) destinataire du message pour le contenu du message.

Mais le paradigme des engagements sociaux est plus général : un engagement peut représenter non seulement l'effet d'une action de communication, mais aussi le résultat d'un dialogue complet, d'une interaction. Par exemple, à la fin d'une négociation, un agent peut s'engager envers un autre à satisfaire une condition.

Indifféremment s'il est créé suite à un message envoyé, à une action effectuée ou à une interaction menée avec un autre agent, un engagement social représente une contrainte interpersonnelle imposée sur le comportement d'un agent. Cette contrainte est *interpersonnelle* parce qu'elle est générée par des interactions entre agents et l'aspect institutionnel n'est pas nécessaire (même s'il peut exister). Nous considérons que cette contrainte est comportementale (cf. Chapitre 1) parce qu'elle limite le comportement futur de l'agent – de plus, un agent qui viole son engagement (n'obéit pas à la contrainte) est souvent appelé *autonome* dans la littérature.

Un autre avantage du paradigme des engagements sociaux vient de la représentation unifiée avec les politiques sociales. Singh (Singh 1999) considère ces politiques sociales comme étant des normes, ce qui permet ainsi de représenter des concepts institutionnels (les normes) et des concepts interpersonnels (les engagements sociaux) de la même manière. Cette représentation unifiée permet aux agents d'utiliser un seul type de raisonnement pour deux concepts de nature différente et le renforcement des normes est simplifié en utilisant des tableaux d'engagements qui explicitent le changement d'états des engagements/politiques.

Sans vouloir entrer dans des détails ici (nous discutons ces aspects plus tard), nous voulons souligner la nature duale du concept de politique sociale. D'un coté, une politique sociale n'est qu'un engagement social (concept *interpersonnel*) de niveau supérieur, d'un autre coté, une politique sociale peut représenter une norme (concept souvent *institutionnel*). De plus, en tant qu'engagement social, une politique sociale est une contrainte créée par une action d'un agent (contrainte *comportementale* par rapport à qui un agent peut être autonome), mais elle représente aussi une contrainte qui limite la prise de décision d'un agent concernant une opération avec un engagement, donc une contrainte *décisionnelle*. Nous analyserons plus tard l'importance que ces dualités ont sur notre travail ; avant cela nous nous intéressons à d'autres formes de contraintes décisionnelles issues des interactions entre agents : les relations de dépendance et de pouvoir.

3.3. Théorie de la dépendance

Les agents d'un système multi-agents ne sont généralement pas capables de satisfaire leurs buts individuellement, ce qui pousse les agents à interagir les uns avec les autres afin de s'aider réciproquement. Ce besoin d'interaction existe dans tous les agents, qu'ils soient coopératifs –aider les autres est un de leurs objectifs – ou compétitifs –aider les autres n'est qu'un moyen d'obtenir de l'aide pour eux-mêmes. Dans les deux cas, les agents essaient de maximiser les chances d'avoir une interaction qui satisfait le mieux leurs objectifs, c.-à-d., de mieux aider les autres ou respectivement d'obtenir le plus d'aide possible. Il est donc important pour un agent de comprendre pourquoi il doit interagir avec d'autres agents (dans quel but) et de pouvoir choisir le meilleur partenaire d'interaction possible – l'agent qui lui permettra de mieux satisfaire ce but.

Dans (Sichman 1995), Sichman appelle ce raisonnement effectué par un agent avant d'interagir avec d'autres un *raisonnement social* et propose l'utilisation de la *théorie de la dépendance* comme mécanisme qui reste à la base d'un tel raisonnement. Il utilise le concept de description externe d'un agent, qui représente les caractéristiques d'un agent connues par les autres agents et utilise cette description pour permettre aux agents d'identifier les relations de dépendance qui les lient. Un réseau de dépendance est ainsi créé, réseau qui peut être utilisé par les agents pour mieux choisir leurs partenaires d'interaction afin de mieux satisfaire leurs buts. Comme la théorie de la dépendance reste à la base d'autres théories sociales et peut être utilisée dans le cadre d'un raisonnement social tel qu'il nous intéresse dans ce manuscrit, nous détaillons les éléments qui le compose.

(1) Description externe

A la base de la théorie de la dépendance nous trouvons le concept de *description externe* – une structure de donnée représentant les caractéristiques d'un agent que les autres peuvent connaître. Sichman définit cette description comme étant composée des *buts* qu'un agent veut satisfaire (qui peuvent être actifs ou non), les *actions* qu'un agent est capable d'effectuer, les *ressources* qu'il possède et les *plans* qu'il connaît – un plan étant considéré comme formé des actions à exécuter et des ressources à utiliser afin de satisfaire un but. Ces éléments sont les seuls à être visibles de l'extérieur – un agent ne connaît pas les mécanismes de décision d'un autre agent, ni son architecture interne, mais seulement cette description externe. Il faut mentionner que même ce qu'un agent connaît sur un autre peut ne pas être tout à fait correct ou complet. Un agent peut avoir des croyances fausses à propos des buts, plans et même actions ou ressources d'un autre agent ou il peut ignorer certains de ces éléments qui caractérisent l'autre.

Le raisonnement type d'un agent considéré de cette manière est de choisir, pour chaque but actif qu'il a, un plan qui satisfait ce but. Dans le cas idéal, un agent connaît un tel plan et il est capable d'effectuer tous les composants du plan, c.-à-d., d'exécuter les actions et de posséder les ressources qui en font partie. Malheureusement, dans le cas général ces hypothèses ne tiennent pas et les agents se retrouvent souvent dans l'impossibilité de satisfaire tout seuls leurs buts.

(2) Relations et réseaux de dépendance

Sichman utilise la notion d'autonomie pour décrire la capacité d'un agent à satisfaire ses buts, mais avec un sens plutôt indépendance en *exécution* que d'autonomie de décision (voir Chapitre 1). Il considère qu'un agent est *r-autonome* (autonome par rapport aux ressources) pour un but et étant donné un ensemble des plans s'il existe un plan dans cet ensemble, plan qui satisfait le but et qui n'utilise que des ressources en la possession de l'agent. De manière similaire il définit un agent *a-autonome* (autonome par rapport aux actions) pour un but et étant donné un ensemble des plans et il utilise la notion de *s-autonome* (autonomie sociale) pour dénoter un agent qui est à la fois r- et a-autonome.

Comme nous l'avons précisé, un agent s-autonome, donc capable de satisfaire tout seul un de ses buts, est un cas idéal – dans le cas général les agents ne sont pas s-autonomes. Un agent n'est pas a-autonome (r-autonome) pour un but si, en utilisant n'importe quel plan dans un ensemble de plans donné, il n'est pas capable d'exécuter toutes les actions (il ne possède pas toutes les ressources ou

actions) demandées. Sichman utilise ainsi la notion de *dépendance* pour décrire la relation qui lie un agent qui n'est pas a-, r- ou s-autonome à un autre agent. Un agent est *a-dépendant* d'un autre agent pour un but et étant donné un ensemble des plans s'il a le but mais n'est pas a-autonome pour lui et, de plus, l'autre agent est capable d'exécuter au moins une des actions nécessaires que le premier ne peut pas exécuter. Une définition similaire est donnée pour la relation de *r-dépendance* et un agent est considéré *s-dépendant* d'un autre s'il est a- ou r-dépendant de lui.

Nous voyons ainsi pourquoi le concept de description externe est nécessaire : pour calculer des relations de dépendance entre agent, il faut connaître au moins l'ensemble d'actions possibles et de ressources possédées par les autres agents. Si de plus un agent connaît toute la description externe d'un autre agent, il peut identifier les relations de dépendance de l'autre agent vers d'autres, en identifiant les buts pour lesquels il n'est pas s-autonome. Des *réseaux de dépendance* peuvent être ainsi construits, réseaux qui représentent des graphes ayant comme nœuds des agents et comme arcs des relations de dépendances. Nous voulons souligner que ces réseaux peuvent être objectifs s'ils sont construits par quelqu'un qui connaît les vraies descriptions externes des agents (tel que le concepteur du système, par exemple) ou ils peuvent être subjectifs et souvent incomplets et même incorrects s'ils sont construits par les agents-mêmes.

(3) Situations de dépendance

Un réseau de dépendance peut s'avérer utile pour l'agent qui l'a construit : il peut servir à identifier des situations de dépendance dans lesquelles l'agent se retrouve par rapport à d'autres agents, situations qui sont ensuite utilisées dans un raisonnement social. Nous décrivons ici les divers types de situation de dépendance possible, tandis que le raisonnement social qui les utilise sera décrit dans le Chapitre 4.

Considérons le cas d'un agent qui a un but et qui n'est pas s-autonome pour ce but, c.-à-d., qu'il lui manque au moins une action ou une ressource pour pouvoir le satisfaire tout seul. Il existe plusieurs situations dans lesquelles il peut se retrouver par rapport à un autre agent. L'agent est *indépendant* par rapport à un autre agent si celui-ci ne peut pas lui fournir le(s) élément(s) qu'il lui manque (action, ressource) et il est *dépendant* par rapport à lui dans le cas contraire. Suivons maintenant le cas de la dépendance : en utilisant la description externe qu'il connaît sur l'autre agent, un agent peut aussi calculer la (les) dépendances de l'autre vers lui. Si un agent dépend d'un autre mais l'autre ne dépend pas de lui, cette relation est appelée une *dépendance unilatérale*. Si un agent dépend d'un autre pour un but et l'autre dépend aussi à son tour de lui pour un autre but, les agents se retrouvent dans une situation de *dépendance réciproque*. Un cas particulier de cette réciprocité est quand les deux buts pour lesquels les agents dépendent l'un de l'autre représentent le même but : les agents se retrouvent ainsi dans une situation de *dépendance mutuelle*.

Parce que les descriptions externes connues par les agents peuvent être incorrectes ou incomplètes, les situations de dépendance calculées par un agent peuvent être différentes de celles calculées par d'autres. Un agent peut ainsi se considérer dans une situation de dépendance réciproque (resp. mutuelle) avec un autre agent en considérant qu'il dépend de l'autre pour un but et que l'autre dépend de lui pour un autre but (resp. pour le même but), sans que cette situation soit reconnue par l'autre agent, qui peut considérer la situation comme une dépendance unilatérale ou même une indépendance.

Sichman utilise les notions de dépendance réciproque/mutuelle *localement connue* et *mutuellement connue* pour faire la différence entre les situations de dépendance dont un seul agent est conscient et celles connues par les deux agents impliqués. Comme nous le verrons dans le Chapitre 4, cette taxonomie de situations de dépendances possibles permet à un agent de raisonner socialement et de prendre des décisions plus informées concernant avec qui et pour quoi interagir.

(4) Relations de dépendance entre les rôles

La théorie de la dépendance telle que proposée et formalisée par Sichman décrit les relations et les situations de dépendances existantes entre des agents, relations et situations identifiées et calculées à partir des descriptions externes des agents. Ce travail a été étendu par les auteurs de (Hannoun *et al.* 1998) aux dépendances entre rôles. La description externe d'un agent le considère en termes des buts qu'il a, plans qu'il connaît, actions qu'il sait exécuter et ressources qu'il possède. Comme nous l'avons vu dans le chapitre précédent, tous ces éléments peuvent faire partie de la description d'un rôle : un rôle peut être obligé de satisfaire des buts, il a des connaissances associées (des plans), des permissions ou des prohibitions d'exécuter des actions ou d'utiliser des ressources. De plus, dans le cas des rôles, leur description externe est généralement connue complètement par les membres de l'organisation, ce qui n'est pas toujours vrai pour la description externe des agents.

Hannoun *et al.* définissent des relations de dépendance entre les rôles d'une organisation, dépendances issues des actions que les agents ont (ou n'ont pas) le droit d'exécuter et des missions qu'ils doivent remplir (concept similaire à une obligation de satisfaire un but – voir le modèle MOISE+ décrit dans le Chapitre 2). Ils arrivent ainsi à définir un *rôle autonome* pour une de ses missions comme étant un rôle qui ne dépend pas d'un autre rôle pour la satisfaction de cette mission. S'il dépend d'autres rôles, les auteurs le considèrent autonome pour une mission avec la coopération d'autres rôles. Cette relation d'autonomie d'un rôle avec la coopération d'autres rôles est très importante pour la conception des structures organisationnelles : une telle structure est consistante si tout rôle qui en fait partie a des missions associées pour lesquelles il est autonome, éventuellement avec la coopération d'autres rôles. En plus, il existe aussi un calcul similaire qui prend en compte les agents (et leurs capacités) qui jouent les rôles, ce qui aide à la bonne répartition agent-rôle dans l'organisation. Ce travail représente une extension importante faite à la théorie de la dépendance (implicitement *interpersonnelle*) en la considérant et utilisant un point de vue *institutionnel.*

(5) Limitations de la théorie de la dépendance

La théorie de la dépendance représente un mécanisme utile pour l'analyse du comportement d'un système multi-agent. L'existence des relations de dépendance explique en grande mesure le comportement interpersonnel des agents, notamment ce qui pousse un agent à interagir avec d'autres agents. Mais cette théorie, telle que présentée et formalisée par Sichman (Sichman 1995), est limitée par quelques hypothèses qui ont été faites. Par exemple, les aspects déontiques sont complètement ignorés dans cet approche : un agent est s-dépendent d'un autre pour un but s'il ne sais pas exécuter une action (ou n'a pas une ressource) nécessaire. Cependant, une dépendance similaire existe dans le cas où l'agent *sait* effectuer l'action (*a* la ressource), mais il n'a pas la permission de l'exécuter (l'utiliser) et l'autre agent a cet permission ou peut lui la donner. Cet aspect déontique qui manque à la théorie de la dépendance originale est partiellement pris en considération par les travaux de Hannoun

et al. (Hannoun *et al.* 1998) qui considèrent des dépendances entre rôles, dépendances générées partiellement par le manque de permissions. Nous considérons néanmoins que les dépendances déontiques doivent être analysées dans plus de détails afin d'obtenir une théorie de la dépendance plus complète.

Une autre hypothèse qui limite la généralité de cette théorie est le fait que toutes les relations de dépendances sont calculées par rapport à un but à satisfaire et *un ensemble de plans*, donc le caractère s-autonome d'un agent peut changer en fonction du plan pris en compte. Le manque d'autonomie d'un agent peut ainsi être provoqué par le fait que l'agent ne connaît pas le bon plan et donc une nouvelle source de dépendance peut être envisagée : un agent dépend d'un autre pour un but s'il ne connaît pas un plan qui satisfait le but et par rapport auquel il est s-autonome, mais le deuxième agent peut lui fournir un tel plan. Cette relation de dépendance provoquée par le manque de connaissance (plans) n'est pas prise en compte dans les travaux de Sichman.

La présence des plans dans la définition de la s-autonomie nous conduit aussi à considérer cette forme d'autonomie comme une autonomie *en exécution*. Un agent est s-autonome pour un but s'il est capable d'*exécuter un plan* connu et qui satisfait le but. Cette approche ne s'intéresse pas à l'autonomie en délibération (est-ce qu'un agent peut décider librement concernant un but), malgré le fait que cette autonomie est influencée par les relations de dépendance. Ces relations représentent des *contraintes interpersonnelles décisionnelles* qui sont générées par le manque de savoir-faire ou de ressources. La décision d'un agent de poursuivre ou pas un but n'est plus libre si l'agent ne peut pas satisfaire le but tout seul, mais il dépend d'un autre agent : l'agent doit premièrement interagir avec l'autre agent (et dans le cas général lui offrir quelque chose en échange) afin de pouvoir satisfaire son but. Cet aspect est pris en considération et analysé plus rigoureusement dans la théorie du pouvoir social qui étend la théorie de la dépendance.

3.4. Théorie du pouvoir social

Le pouvoir est un élément important dans les relations, organisations ou sociétés humaines et beaucoup de théories des sciences sociales s'y intéressent. Il n'est donc pas surprenant que ce concept intéresse la communauté multi-agents, qui essaye de concevoir des systèmes complexes à l'aide des modèles inspirés de la société humaine. Il est loin des objectifs de cette thèse de vouloir proposer un état de l'art exhaustif sur les théories de pouvoir dans les sciences sociales : nous nous contentons ici de décrire les différents concepts qui sont à la base de telles théories et leur utilité pour les systèmes multi-agents. Dans la suite de cette section nous décrivons ces concepts tels qu'ils sont présentés par Castelfranchi dans (Castelfranchi 2002), où ils sont groupés dans trois catégories : pouvoirs individuels, relations de pouvoirs entre agents et pouvoirs institutionnels. Nous présentons ensuite d'autres points de vue existants dans la littérature et des essais de formalisation de ces concepts.

3.4.1. Pouvoirs individuels

Contrairement à la plupart des points de vue et des définitions de sciences sociales, Castelfranchi considère que le pouvoir n'est pas une notion intrinsèquement sociale qui ne limite pas sa référence aux sociétés des agents. Elle exprime évidemment une relation sociale très importante, mais à la base le pouvoir se réfère à la relation entre un agent, ses buts et ses capacités et ressources. La notion de

base d'une théorie du pouvoir est donc le *pouvoir de* (*power-of* en anglais), notion qui est relative à quelque chose, généralement un but. Il est donc possible d'exprimer le fait qu'un *agent a le pouvoir de satisfaire un but*, ce qui signifie que l'agent est en condition (est capable) de faire en sorte que l'état du monde décrit par le but soit atteint. Ce concept de pouvoir s'applique aussi aux actions (un agent a le pouvoir d'exécuter une action) ou même aux état de l'environnement (un agent a le pouvoir d'arriver à un état).

Il existe plusieurs caractéristiques de ce concept, caractéristiques qui permettent d'identifier plusieurs types de pouvoirs existants, tels que pouvoir objectif ou subjectif, pouvoir interne ou externe, pouvoir d'une action ou d'un but, pouvoir direct ou indirect, etc. :

- *pouvoir objectif et pouvoir subjectif*. Même si dans le cas général un agent est conscient de ses capacités et ressources, ceci n'est pas toujours vrai. Il est possible qu'un agent a le pouvoir d'exécuter une action (il sait le faire), sans qu'il sache qu'il a ce pouvoir. Il existe donc une notion *objective* de pouvoir – power-of qu'un agent a – et une notion *subjective* – power-of qu'un agent a et sait qu'il l'a.

- *pouvoir interne et pouvoir externe*. Les pouvoirs d'un agent peuvent aussi être divisés en deux catégories, selon leur relation avec l'agent. Les pouvoirs *internes* représentent les pouvoirs basés sur des concepts internes à l'agent, tels que ses capacités, ses actions, ses connaissances, etc. Quant aux pouvoirs *externes*, ils font référence aux ressources qu'un agent possède (p.ex. : argent) ou à des conditions favorables dans l'environnement.

- *pouvoir d'une action et pouvoir d'un but*. Un agent a le *pouvoir d'une action* s'il sait effectuer l'action (l'action fait partie du répertoire d'actions de l'agent) et s'il est en condition de l'effectuer, c.-à-d., les conditions nécessaires sont remplies. Un agent a *le pouvoir d'un but* s'il connaît un plan qui satisfait le but et s'il sait exécuter un tel plan (sait effectuer toutes les actions et possède toutes les ressources prévues dans le plan). Nous voyons ainsi que pour avoir le pouvoir d'un but, un agent doit généralement avoir le pouvoir d'une ou de plusieurs actions. Comme généralement les agents sont des entités téléonomiques (leurs actions sont orientées vers la satisfaction des buts), Castelfranchi propose ainsi de parler du pouvoir d'une action dans le cadre d'un but – le pouvoir d'un agent de satisfaire un but en exécutant une action.

- *pouvoir direct et indirect*. Quand un agent sait effectuer une action et qu'il est en condition de l'effectuer, il a le pouvoir de cette action et ce pouvoir est un pouvoir *direct* – l'agent peut l'effectuer directement, sans l'intermédiaire d'un autre agent. Un agent peut aussi avoir des pouvoirs *indirects*, en passant par d'autres agents : dans ce cas, l'agent ne peut pas effectuer directement une action, mais il peut convaincre un autre agent (qui peut le faire) de le faire. Nous verrons plus tard la signification du point de vue de la théorie du pouvoir du fait qu'un agent peut convaincre un autre d'utiliser son pouvoir.

Pour résumer, il existe plusieurs caractéristiques de pouvoirs qu'un agent peut avoir, caractéristiques que nous avons brièvement décrits ci-dessus. Nous voulons également souligner que le *manque de pouvoir* est une notion importante aussi, surtout pour les relations de pouvoir entre agents. Quand un agent n'a pas le pouvoir d'une action ou d'un but, il manque le savoir-faire, des ressources, des opportunités, etc. pour effectuer l'action ou satisfaire le but. Un cas intéressant est celui quand un agent a un pouvoir objectif, mais qu'il n'a pas le pouvoir subjectif (il ne sait pas qu'il a le pouvoir). La notion de *pouvoir total* (ou *complet*) est alors utilisée pour dénoter le fait qu'un agent a un *pouvoir de*

et il est également conscient de ce fait et peut prendre des décisions le concernant. Si nous prenons un exemple, un agent a le pouvoir d'un but s'il connaît un plan qu'il sait exécuter et qui satisfait le but. Si l'agent ne sait pas qu'il a ce pouvoir de, il n'a pas le pouvoir total pour ce but – il ne le poursuivra probablement pas parce qu'il ne sait pas qu'il peut le satisfaire. Si l'agent a d'autres buts (ou motivations) plus importants, il n'a également pas le pouvoir total pour ce but – il ne le poursuivra pas parce qu'il décidera probablement de poursuivre d'autres buts.

3.4.2. Relations de pouvoir entre agents

Les notions de pouvoir individuel présentées ci-dessus permettent l'apparition de diverses relations de pouvoir entre agents. Un exemple de relation sociale de pouvoir est le *pouvoir relatif* de deux agents. Il est possible de quantifier le pouvoir des agents et donc de comparer qui a le plus de pouvoir, ce qui représente une relation sociale. Ceci est très important dans les sociétés compétitives, où plus un agent a de pouvoir, plus il a de chances de satisfaire ses buts. Ce pouvoir relatif est souvent appelé *pouvoir de négociation* parce qu'il est utilisé sur des places de marché : un agent avec plus de pouvoir que d'autres sera généralement préféré comme exécutant des contrats parce qu'il a plus d'options à sa disposition pour satisfaire des buts. Cette relation de pouvoir est assez simple, surtout par rapport à d'autres relations telles que les relations de dépendance, le pouvoir sur un autre, le pouvoir d'influencer, etc., relations que nous analysons par la suite.

(1) Relations de dépendance

Comme nous l'avons précisé, la théorie de la dépendance est à la base de la théorie du pouvoir social. Cette dernière définit ainsi pour un agent *sa dépendance d'un autre agent pour une action dans un but* si le premier agent a besoin de l'action du deuxième agent pour satisfaire le but. C'est-à-dire, le premier agent manque de pouvoir d'action dans le but et le deuxième a ce pouvoir. La plupart des caractéristiques des relations de dépendances présentées dans la section précédente sont intégrées dans la théorie du pouvoir, telles que des dépendances réciproques ou mutuelles, etc. Nous voulons souligner aussi que la dépendance est considérée comme une notion objective – elle existe même si les agents ne se rendent pas compte, mais que des situations intéressantes peuvent apparaître si nous considérons les croyances des agents. Des dépendances inconnues apparaissent quand un agent qui dépend d'un autre ignore cette dépendance et des dépendances illusoires apparaissent quand un agent qui n'est pas dépendant d'un autre croit qu'il l'est.

(2) Avoir du pouvoir sur un autre et pouvoir d'influencer

Considérons maintenant une relation de dépendance entre deux agents et le point de vue du deuxième agent : il peut être ou ne pas être conscient de l'existence de cette relation de dépendance. Si un agent croit qu'un autre agent dépend de lui (pour une action dans un but), il a du *pouvoir sur* cet agent. Il est important de souligner que cette relation de pouvoir est *subjective*, contrairement à celle de dépendance sur laquelle elle est basée. Notamment, un agent peut croire qu'un autre dépend de lui sans que ce soit le cas – il a du pouvoir sur l'autre agent, mais il ne sera pas capable d'utiliser ce pouvoir. Parmi les utilisations les plus fréquentes de ce pouvoir sur un autre, on trouve la motivation et la récompense de l'autre. L'agent qui a du pouvoir sur un autre peut utiliser ce pouvoir pour motiver l'autre à faire quelque chose et le récompense ou le pénalise en fonction de ce qu'il fait. Généralement,

la récompense prend la forme de l'exécution de l'action pour laquelle l'autre dépend de lui et la pénalité prend la forme de la non-exécution de cette action ou l'exécution d'une action contraire.

Quand un agent manque du pouvoir d'exécuter une action pour satisfaire un but et un autre agent a ce pouvoir, le premier agent dépend du deuxième et doit le convaincre de vouloir exécuter l'action pour lui. Il peut le convaincre s'il a le *pouvoir de l'influencer* – relation de pouvoir qui représente le fait qu'un agent a du pouvoir sur un autre agent et que l'autre agent en est conscient. C'est-à-dire, un agent a du pouvoir d'influencer un autre si l'autre agent croit qu'il est conscient que l'autre dépend de lui. Ceci peut sembler compliqué, mais ces relations de pouvoir apparaissent souvent dans les interactions humaines et le fait que les croyances d'agents peuvent être fausses génère des situations intéressantes, comme le montre l'exemple ci-dessous.

(3) Utilisations des relations de pouvoir

Prenons par exemple le cas d'un agent *voleur* qui menace avec un pistolet un agent *caissier*. Le voleur dépend du caissier pour lui donner de l'argent (et satisfaire son but d'avoir de l'argent) et le caissier dépend du voleur pour ne le pas tuer (et satisfaire son but de rester en vie) – une situation de dépendance réciproque. Comme le voleur est conscient de la dépendance du caissier vis-vis de lui, il a du pouvoir sur le caissier. Comme le caissier croit que le voleur est conscient de ce pouvoir qu'il a sur lui, le voleur obtient ainsi le pouvoir d'influencer le caissier, pouvoir qu'il utilisera pour satisfaire son but d'avoir de l'argent. Il suffit que le caissier croit qu'il dépend du voleur et que qu'il croit que ce dernier le sait pour que le voleur ait le pouvoir d'influencer le caissier Si le pistolet est en plastique, la première croyance du caissier est fausse (dépendance illusoire), mais les relations de pouvoir sur et pouvoir d'influencer existent toujours. Si le caissier croit que le voleur ne se rend pas compte qu'il le menace avec un pistolet, le voleur n'a pas du pouvoir de l'influencer – cela est la raison pour laquelle les voleurs utilisent aussi des menaces verbales (ils envoient des messages) : pour communiquer aux autres qu'ils sont conscients de l'existence de la relation de pouvoir sur et obtenir ainsi le pouvoir de les influencer.

Cet exemple montre comment les pouvoirs individuels et surtout leur manque donnent naissance à diverses relations de pouvoir entre agents, relations de pouvoir qui sont ensuite utilisées pour acquérir de nouveaux pouvoirs individuels. En effet, un agent qui a le pouvoir d'influencer un agent pour une action (dans un but), acquiert un pouvoir *indirect* d'effectuer l'action (ou de satisfaire le but) : il pourra le faire par l'intermédiaire de l'autre agent. Nous voyons ainsi comment dans une société les pouvoirs des agents se multiplient (l'agent aura plus de pouvoirs grâce aux pouvoirs indirects), les pouvoirs circulent entre les agents (le pouvoir direct d'un agent devient le pouvoir indirect d'un autre qui a le pouvoir de l'influencer) ou les pouvoirs se transforment (si un autre agent dépend d'un autre agent pour un de ses pouvoirs, ce pouvoir se transforme en un pouvoir d'influencer, qui à son tour se transforme en un pouvoir indirect).

3.4.3. Vers des pouvoirs institutionnels

Les relations de pouvoir présentées brièvement ci-dessus donnent une idée de la dynamique des pouvoirs qui existent dans une société. Castelfranchi estime qu'à part la multiplication, circulation ou transformation des pouvoirs, une opération est particulièrement importante pour décrire cette

dynamique : le *don de pouvoir* (empowerment). En effectuant une action, un agent peut donner du pouvoir à un autre agent de satisfaire un but (ou d'exécuter une autre action). Pour parler d'un véritable don de pouvoir, il est nécessaire que cette action soit effectuée avec l'intention de donner du pouvoir, c.-à-d., le don de pouvoir ne doit pas être accidentel. De plus, il est nécessaire que l'agent recevant du pouvoir ait besoin de ce pouvoir, c.-à-d., il ne l'avait pas auparavant. Le don de pouvoir prend souvent la forme d'un don de permission : un agent n'a pas la permission d'effectuer une action, un autre agent lui donne cette permission en lui donnant ainsi le pouvoir d'effectuer l'action. Mais le don de pouvoir n'est pas toujours déontique : un agent peut donner du pouvoir à un autre en lui communiquant des plans ou en lui enseignant des actions, ce qui permettra à l'agent de satisfaire ses buts.

Le don de pouvoir est particulièrement important dans un contexte institutionnel parce qu'il peut expliquer beaucoup de relations entre une institution et ses membres. Notamment, dans le cas le plus évident, une institution donne du pouvoir aux agents qui en font partie, souvent sous la forme du don des permissions ou de mise à disposition des ressources. Un aspect moins évident de ce don de pouvoir de la part d'une institution est l'effet *count-as* : parfois les actions d'un agent ne sont plus interprétées comme ses actions propres, mais comme les actions de l'institution qu'il représente, ce qui lui confère plus de pouvoirs. L'exemple classique est celui d'un policier qui arrête un malfaisant : ce dernier s'arrête non parce qu'il a peur de l'agent policier, mais de l'institution qu'il représente. Une institution donne ainsi des pouvoirs à ses membres simplement parce que le fait qu'ils en fassent partie est visible ou connu par d'autres agents (dans l'exemple précédent l'uniforme de policier donne du pouvoir à l'agent qui le porte). Par exemple, les auteurs de (Demolombe and Louis, 2006), précisent les pouvoirs institutionnels donnés à un agent qui joue un rôle dans une organisation – dans leur approche, ces pouvoirs peuvent être considérés comme des normes contextualisées.

Les dons de pouvoir d'une institution sont généralement faits avec un but précis : acheter du pouvoir de ses membres. L'institution donne du pouvoir aux membres en échange des pouvoirs que les membres donnent à l'institution. L'opération par laquelle ceci est effectué est la *mise à disposition des pouvoirs*. Un agent qui entre dans une institution met à sa disposition quelques-uns de ses pouvoirs (et parfois tous) – ces pouvoirs deviennent ainsi des pouvoirs de l'institution. L'agent garde son pouvoir de satisfaire des buts, mais il perd le pouvoir total pour ces buts : il n'est plus libre de décider comme il veut de les poursuivre ou pas – c'est l'institution qui en décide. L'institution gagne ainsi des pouvoirs indirects de satisfaire des buts par l'intermédiaire de ses membres.

3.4.4. Discussion

Ce bref passage en revue des concepts les plus importants de la théorie du pouvoir social nous montre la complexité de ces concepts. Les diverses relations de pouvoir existantes entre agents, surtout dans un contexte institutionnel et leur dynamique constituent un sujet de recherche encore ouvert dans les sciences sociales – sans vouloir donner une image exhaustive de ce domaine, nous n'avons présenté ici que les éléments de base qui peuvent être utilisés dans le domaine des systèmes multi-agents. Nous pensons qu'il peut être utile pour un agent de pouvoir calculer ses pouvoirs et ses relations de pouvoir avec d'autres agents, afin de pouvoir prendre des décisions plus informées.

Notamment, comme les relations de dépendance sur lesquelles elles se basent, le pouvoir sur un autre et le pouvoir d'influencer représentent des contraintes décisionnelles imposées aux agents. La prise de décision d'un agent est limitée par le fait qu'un autre a du pouvoir sur lui ou a le pouvoir de l'influencer. Ceci devient plus évident dans le cas du pouvoir indirect : quand un agent a un pouvoir indirect par l'intermédiaire d'un autre agent, ce dernier n'a plus de pouvoir total – il ne décide plus concernant son but ou action – une contrainte décisionnelle est représentée par l'existence de ce pouvoir. Malgré le fait que l'origine de ces relations de pouvoir est interpersonnelle (relations de dépendance issues du manque de pouvoir individuel), comme nous avons pu le voir, ces contraintes peuvent aussi être considérées dans un contexte institutionnel. Nous pensons ainsi que la théorie du pouvoir peut fournir des outils pour la modélisation des contraintes décisionnelles interpersonnelles, mais aussi institutionnelles.

Cependant, cette théorie à ses limitations. Du point de vue conceptuel, les relations de pouvoir entre agents ne sont pas très claires quand ceux-ci appartiennent à une institution. Ceci s'explique par la complexité des interactions entre un agent et son institution, ainsi que par la manière parfois très subtile avec laquelle une institution modifie les interactions entre ses membres. De plus, afin de pouvoir utiliser une telle théorie dans les systèmes multi-agents, elle doit être formalisée. A notre connaissance, peu de travaux s'intéressent à cet aspect. Par exemple, Sichman (Sichman 1995) propose une formalisation qui lui permet d'identifier des relations de dépendance, sans aller plus loin, tandis que dans (Boella *et al.* 2004) une formalisation pour les pouvoirs individuels et les relations de dépendance est proposée. Les auteurs de (Lorini *et al.* 2007) proposent eux aussi une modélisation de la théorie du pouvoir social, en s'intéressant surtout aux aspects liés à la théorie de l'action. D'autres auteurs, tels que (Jones & Sergot 1996) s'intéressent à la formalisation seulement des aspects institutionnels (comme l'effet *count-as*), sans considérer les relations interpersonnelles. En conclusion, malgré (et probablement à cause de) son grand potentiel de pouvoir expliquer les interactions entre agents dans un contexte interpersonnel et aussi institutionnel, la théorie du pouvoir social nécessite la clarification et surtout la formalisation de certains concepts afin d'être utilisée par des agents artificiels.

3.5. Synthèse

Pour arriver à un comportement global cohérent dans un système sans utiliser des concepts institutionnels, les agents doivent interagir pour coordonner leurs activités. Indifféremment de comment cette interaction est effectuée, par l'intermédiaire de l'environnement ou par envoi des messages, en utilisant des protocoles ou dans un framework de coordination, cette coordination interpersonnelle résulte en des contraintes sur la prise de décision et sur le comportement des agents. Nous voyons ainsi apparaître deux catégories des contraintes : les contraintes qui poussent les agents à interagir (à se coordonner) et les contraintes issues des interactions.

La deuxième catégorie de contraintes représente les contraintes issues des interactions entre agents. La plupart des modèles existants utilisent d'une manière plus ou moins implicite le paradigme d'*engagements sociaux*, engagements qui représentent le comportement attendu d'un agent à la fin d'une interaction (ou dans certains modèles suite à une action ou à un envoi de message). Parmi les avantages de ce paradigme que nous avons mentionnés ci-dessus, nous voulons rappeler le fait que les mêmes modèles de représentations, de gestion de et de raisonnement sur engagements sociaux peuvent

être utilisés pour des méta-engagements aussi. Ces méta-engagements, appelés généralement des *politiques sociales* représentent des contraintes sur la prise de décision des agents (et pas sur leur comportement comme les engagements de bas-niveau) – même si ces contraintes sont à la base interpersonnelles, souvent les politiques sociales représentent des normes et d'autres concepts interpersonnels.

Quant à la première catégorie des contraintes, nous avons présenté *la théorie de la dépendance* qui est utilisée pour représenter des contraintes interpersonnelles décisionnelles qui poussent les agents à interagir : quand un agent ne peut pas satisfaire tout seul ses buts, il dépend d'autres agents et doit chercher leur aide pour les satisfaire. Cette théorie de la dépendance constitue une partie d'une théorie plus générale la *théorie du pouvoir social* que nous avons aussi brièvement présenté. Celle-ci est une théorie des sciences sociales qui s'intéresse aux pouvoirs individuels d'un agent et aux relations de pouvoir entre agents dans des contextes interpersonnels, mais aussi institutionnels. Grâce aux différents types de pouvoir, il est possible de décrire les actions qu'un agent sait et a le droit d'effectuer et des buts qu'il sait, peut ou a le droit de satisfaire. Différentes relations apparaissent entre les agents en fonction des relations de dépendance qui existent entre eux à cause des pouvoirs qu'ils manquent. Ces relations peuvent générer des situations intéressantes si les agents sont conscients ou non de leur existence. Ces relations de pouvoir entre agents représentent également des contraintes décisionnelles qui limitent la prise de décision des agents, en les poussant par exemple à chercher de l'aide d'un autre agent ou de mettre à sa disposition des pouvoirs individuels. Cependant, la théorie du pouvoir ne s'intéresse pas seulement aux relations de pouvoir dans un contexte non-institutionnel, mais aussi à ces relations dans le cadre d'une institution : il nous semble que cette théorie peut être utilisée pour modéliser de telles contraintes issues d'un coté des interactions entre agents et d'un autre des concepts organisationnels utilisés.

Pour résumer, dans ce chapitre nous avons présenté plusieurs modèles et théories qui permettent la représentation des contraintes interpersonnelles qui limitent la prise de décision ou le comportement des agents. De plus, certains de ces modèles peuvent être plus ou moins naturellement étendus à des contextes institutionnels, ce qui nous offre des pistes vers une représentation plus uniforme des contraintes imposées aux agents. Nous rappelons ainsi que ces contraintes sont directement liées au concept d'autonomie, concept qui reste à la base du raisonnement qui constitue l'objectif de cette thèse et qui sera analysé par la suite.

4. Raisonnement social dans les systèmes multi-agents

Comme cette thèse s'intéresse au raisonnement des agents, nous analysons dans ce chapitre des modèles de raisonnement social proposés dans la littérature. Les agents raisonnent pour résoudre des problèmes et atteindre leurs objectifs. Ils sont généralement conçus en utilisant des architectures ou des modèles. Nous passons en revue quelques-uns de ces derniers, en mettant en évidence leur évolution vis-à-vis de la prise en compte de concepts ou de résolution des problèmes de plus en plus complexes. Ce passage en revue nous permet ainsi d'identifier plusieurs types de raisonnement effectué par les agents, raisonnements que nous synthétisons dans quatre catégories que nous discuterons sur un exemple de système multi-agent. Parmi ces quatre catégories, deux sont particulièrement intéressantes pour ce travail, catégories que nous appelons raisonnement *a priori* et *a posteriori*. Nous détaillons ensuite ces deux types de raisonnement avec des exemples tirés de la littérature en identifiant s'ils s'intéressent au raisonnement concernant des contraintes interpersonnelles, institutionnelles ou les deux. Finalement, nous discutons ces raisonnements en relation avec le concept d'autonomie et nous identifions ce qui manque, selon nous, pour obtenir un raisonnement basé sur l'autonomie dans les systèmes multi-agents.

4.1. Architectures et modèles d'agents

De nombreuses architectures et modèles d'agents ont été proposés et utilisés par la communauté multi-agents. Leur nombre s'explique par la diversité des problèmes résolus par les systèmes multi-agents. Chaque type de problème a en effet ses propriétés qui imposent des choix sur les caractéristiques de l'architecture utilisée pour le résoudre. Les architectures existantes peuvent être classifiées en fonction de différents critères, tels que le type de modèle utilisé (symbolique ou non) ou la prise en compte explicite des relations avec d'autres agents (Wooldridge 1999).

Dans ce manuscrit nous suivons la classification utilisée par Boissier (Boissier 2001) qui identifie trois types d'agents : *agents individuels* (qui raisonnent seulement sur leurs interactions avec l'environnement), *agents sociaux* (qui raisonnent en plus sur les interactions avec d'autres agents) et *agents normatifs* (qui raisonnent en plus sur des concepts organisationnels tels que des normes, rôles, etc.). Cette classification souligne l'évolution des problèmes auxquels la communauté multi-agents s'est intéressée : si initialement les agents devaient résoudre des problèmes tous seuls, sans interagir avec d'autres agents, des agents plus complexes ont été nécessaires ensuite pour résoudre des problèmes plus complexes. Certains domaines d'application ont nécessité une coordination explicite entre les agents, donc un besoin d'interagir avec d'autres agents (sociaux) et puis des règles de comportement ont été introduites pour régulariser les comportements d'agents (normatifs).

Le but de ce travail étant de créer des agents dotés d'un raisonnement sur des contraintes interpersonnelles et institutionnelles, les architectures d'agents normatifs présentent le plus d'intérêt pour nous. Par la suite, nous présentons donc brièvement les plus importantes architectures et modèles d'agents individuels et sociaux et analysons avec plus de détails les architectures d'agents normatifs.

4.1.1. Architectures d'agents individuels et sociaux

Les agents individuels raisonnent et choisissent leur comportement en fonction de leur état interne et de l'état de l'environnement, sans prendre en considération d'autres agents. De ce point de vue, deux types de raisonnement se sont imposés dans la communauté multi-agents : des agents *réactifs* et des agents *délibératifs*. Chacun de ces modèles a ses avantages et inconvénients, ce qui explique l'utilisation fréquente d'un autre type d'agents : les agents *hybrides*, capables de manifester un comportement réactif ou délibératif en fonction du besoin. Pour permettre aux agents de raisonner sur les interactions avec d'autres agents, ces architectures d'agents hybrides ont été enrichies avec des modèles sociaux, pour obtenir ainsi des agents *hybrides et sociaux* :

- *Agents réactifs – l'architecture de subsomption* proposée par (Brooks 1986) est un des exemples les plus connus d'architecture pour les agents réactifs. Un agent est spécifié avec un ensemble des comportements possibles, chacun avec une priorité associée. Un mécanisme de contrôle assure que seul le comportement le plus prioritaire est utilisé – même si plusieurs comportements sont utilisables dans un état de l'environnement, un seul d'entre eux sera utilisé. Un comportement est généralement représenté à l'aide d'un automate à états finis qui prend une entrée de l'environnement et la transforme dans une action à effectuer. Les agents ne maintiennent donc pas une représentation symbolique de l'environnement et n'effectuent pas des raisonnements symboliques. Cette simplicité du raisonnement fait que ces agents sont capables de répondre vite aux changements dans l'environnement, mais qu'ils ne sont pas capables de résoudre individuellement des problèmes complexes.

- *Agents délibératifs – le modèle BDI* (Georgeff and Lansky 1987) est sans doute le modèle d'agents le plus connu et utilisé. Il est à la base d'une des premières architectures d'agents délibératifs, l'architecture PRS (Rao and Georgeff 1991). Ce modèle considère qu'un agent consiste en un ensemble de croyances, un ensemble de désirs et un ensemble d'intentions. Une logique BDI existe et définit des opérateurs modaux pour ces trois concepts, ce qui permet aux concepteurs de prouver des propriétés de leurs agents. Ainsi, les croyances d'un agent peuvent s'avérer fausses, ses désirs contradictoires, mais l'ensemble d'intentions doit être cohérent. L'architecture PRS rajoute à ce modèle un ensemble de plans qui peuvent être activés par des désirs et/ou par des états de l'environnement et un interpréteur (module de raisonnement) symbolique qui choisit le plan le plus adapté et modifie les intentions de l'agent en conséquence. Les agents délibératifs sont capables de résoudre des problèmes complexes grâce à leur raisonnement symbolique, mais leur temps de réponse aux changements dans l'environnement peut s'avérer trop grand et donc inefficace.

- *Agents hybrides et sociaux*. Des agents hybrides ont été proposés et utilisés pour combiner les avantages des agents réactifs et délibératifs et éliminer leurs inconvénients. Les architectures d'agents hybrides sont en général des architectures *en couches* – où chaque couche contient un comportement réactif ou délibératif et s'active en fonction du problème à résoudre et de l'état de l'environnement. Un exemple de ce type est l'*architecture InterRaP* (Muller 1996) qui contient trois couches : une basée comportement (réactive), une pour la planification locale (délibérative) et une pour la planification coopérative. Cette dernière couche contient le raisonnement social de l'agent, raisonnement sur les buts/intentions communs (joint goals en anglais) avec d'autres agents et sur les croyances mutuelles et partagées entre agents. Ces éléments communs à plusieurs agents sont mis à jours à l'aide des interactions entre agents, ce qui fait des agents InterRaP des agents sociaux.

Synthèse

D'autres architectures d'agents ont été proposées dans la littérature – d'une manière générale elles contiennent toutes des composants réactifs et/ou délibératifs et dans le cas d'agents sociaux des composants sociaux qui gèrent les interactions entre agents et qui sont responsables de la modification du comportement de l'agent en fonction du résultat de ces interactions. Peu à peu ces architectures ont évolué pour résoudre des problèmes de plus en plus complexes, en arrivant ainsi à des architectures d'agents qui prennent en compte des concepts normatifs, architectures que nous présentons par la suite.

4.1.2. Architectures d'agents normatifs

Le besoin d'utiliser des mécanismes institutionnels pour la coordination dans les systèmes multi-agents a enrichi les modèles et architectures d'agents avec des modules capables de gérer les contraintes institutionnelles imposées aux agents. Parmi les nombreuses architectures qui existent dans la littérature, nous présentons par la suite l'architecture ADEPT, le modèle B-DOING, une modification apportée à l'architecture DESIRE et l'architecture NoA, en mettant en évidence leur influence sur l'autonomie des agents.

(1) L'architecture ADEPT

L'architecture ADEPT (Jennings *et al.*, 1998) est une architecture d'agents utilisés dans le domaine du commerce électronique. Ces agents font partie d'un système multi-agent constitué de plusieurs *agences*, chaque agence étant composée d'un agent *responsable*, un ensemble d'agences *subsidiaires* et un ensemble de *tâches à accomplir*. Des organisations horizontales, hiérarchiques ou hybrides peuvent être ainsi construites. L'objectif du groupement d'agents dans des agences est de permettre une bonne coordination par la délégation des tâches entre agents. Si un agent délègue une tâche à un agent subsidiaire (appartenant à une agence subsidiaire de l'agence dont le premier agent est responsable), celui-ci ne peut pas refuser l'adoption, il peut seulement proposer une meilleure solution. Par contre, le refus est possible dans les délégations entre agents non-subsidiaires et des mécanismes de négociation doivent être utilisés.

Cette organisation d'un système multi-agent influence l'architecture des agents qui en font partie. L'architecture ADEPT contient plusieurs modules qui permettent à un agent d'être conscient de la structure organisationnelle existante (les agences) et d'agir en conséquence. Un agent doit raisonner sur les contraintes organisationnelles avant d'interagir avec un autre agent parce que l'interaction avec un agent subsidiaire (délégation sans possibilité de refuser l'adoption) diffère de l'interaction avec un agent non-subsidiaire (négociation). Cette différence entre les deux types d'interaction est codée en dur dans l'architecture des agents, où des modules différents sont utilisés pour la prise de décision d'adopter ou pas la tâche en fonction de la relation avec l'agent déléguant la tâche.

L'architecture ADEPT est une des premières architectures d'agents qui prend en compte des concepts institutionnels. Son approche sur l'autonomie d'agents est relativement simpliste. Un agent a de l'autonomie par rapport à un autre agent pour l'adoption d'un but seulement si l'autre agent n'est pas son supérieur dans la hiérarchie existante. L'existence des modules différents de prise de décision en fonction des relations entre agents assure que les agents n'auront pas des comportements autonomes.

Les organisations ainsi construites ne sont pas flexibles et le comportement et la prise de décision des agents sont sévèrement contraints par les spécifications institutionnelles.

(2) Le modèle B-DOING

Le modèle B-DOING (Dignum *et al.*, 2001) a été proposé comme une extension du modèle BDI. Si dans ce dernier les intentions de l'agent étaient créées à partir de ses désirs et croyances, dans le modèle B-DOING une notion intermédiaire est utilisée : les buts. Les intentions sont maintenant générées à partir des croyances et buts et les buts sont générés à partir des croyances et désirs, mais aussi des obligations et normes. Dans ce modèle les obligations sont issues des interactions avec d'autres agents – similaires donc aux engagements sociaux présentés dans le chapitre précédent. Les obligations représentent donc les contraintes interpersonnelles et les normes des contraintes institutionnelles. Comme le montre la Figure 4.1, un agent met à jour les buts qu'il va poursuivre à partir des buts générés de ces trois sources : ses désirs, ses obligations et ses normes, en prenant en compte ses croyances et les buts déjà existants. Ce modèle a à la base une logique B-DOING qui étend la logique BDI en rajoutant trois opérateurs pour les buts, obligations et normes.

Le modèle B-DOING rajoute dans le raisonnement délibératif de type BDI des aspects sociaux : les obligations (contraintes interpersonnelles) et les normes (contraintes institutionnelles) peuvent aussi générer ou modifier les buts d'un agent. Parce que le processus de mise à jour des buts agit comme un filtre pour les buts générés par trois sources, un agent de type B-DOING peut être dans un comportement autonome par rapport à d'autres agents (s'il ne poursuit pas les buts issus de ses obligations) ou par rapport aux normes (s'il ne poursuit pas les buts issus des normes). En général, ce modèle assure une autonomie totale : l'agent a la possibilité de violer toute contrainte qui lui est imposée, mais des agents sans un type d'autonomie peuvent être conçus si le module de maintenance des buts est conçu pour ne jamais refuser un type des buts.

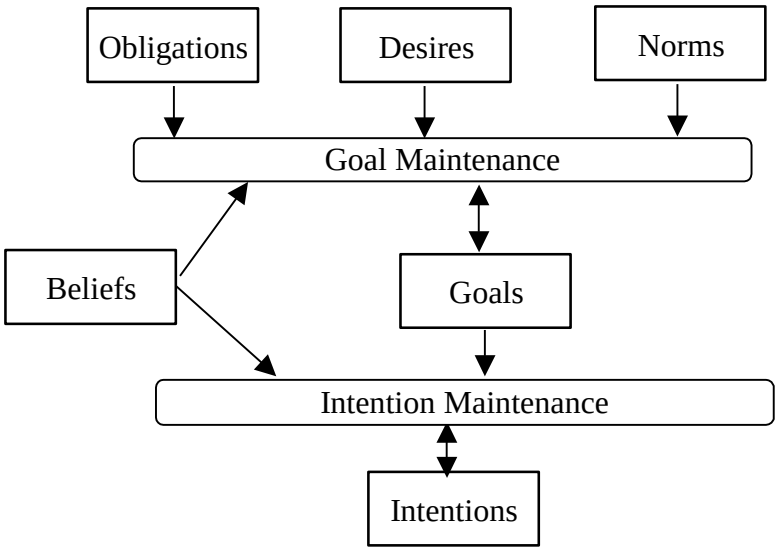


Figure 4.1. Le modèle B-DOING

Dans ce modèle, les obligations et les normes représentent des contraintes établies (à posteriori) parce qu'elles ont été déjà créées par l'agent. Les obligations sont les contraintes sur le comportement de l'agent issues de ses interactions passées, tandis que les normes ont été déjà adoptées par l'agent.

Malgré sa généralité et son pouvoir de générer tout type d'autonomie, ce modèle ne raisonne pas sur des contraintes à priori – il ne présente pas le raisonnement de l'agent avant ou pendant les interactions avec d'autres agents, ni sa prise de décision d'adopter ou pas une norme. Pour résumer, le modèle B-DOING permet la création d'agents qui peuvent avoir de l'autonomie sociale, mais il ne donne aucune indication sur la manière de raisonner sur les implications de certaines décisions sur l'autonomie sociale.

(3) L'architecture DESIRE modifiée

L'architecture DESIRE (Brazier *et al.*, 1999) considère un agent composé de plusieurs composants de haut-niveau qui peuvent à leur tour être constitués de plusieurs composants de bas-niveau. Le concepteur de l'agent a la possibilité de créer des agents avec seulement les fonctionnalités désirées en rajoutant ou supprimant des composants. En plus, l'approche basée composants facilite la modification du type d'agents : pour créer un agent capable d'agir dans un nouveau domaine, des composants peuvent être remplacés avec d'autres, plus adaptés, sans modifier le reste de l'agent. Les agents créés avec l'architecture DESIRE sont des agents sociaux qui peuvent interagir avec d'autres agents, mais qui ne prennent pas en compte de concepts institutionnels.

Pour résoudre ce problème, deux nouveaux composants ont été proposés et introduits dans l'architecture existante (Castelfranchi *et al*., 2000) : le composant de maintenance de l'information sur la société (Maintenance of Society Information) et le composant de gestion des normes (Norm Management). Le premier, un composant de haut-niveau, a le rôle d'assurer le fait que l'agent est conscient des normes existantes – il représente l'interface entre l'agent et l'organisation à laquelle il appartient. Le deuxième composant, de bas-niveau, représente la partie du raisonnement de l'agent responsable de l'adoption des normes. Ceci est le composant qui décide si l'agent adopte ou non une norme, norme qui une fois adoptée peut générer des buts pour l'agent.

Les buts de l'agent peuvent être de provenance institutionnelle (générés par les normes adoptées), interpersonnelle (générés par les interactions avec d'autres agents) ou propre (générés par ses désirs), mais aucun but n'a la garantie d'être poursuivi. Des composants de raisonnement sont utilisés, pour choisir quels seront les buts à poursuivre. Ils assurent ainsi l'autonomie de l'agent par rapport aux normes ou aux autres agents. De plus, l'existence du module de Norm Management fait que l'agent a une forme d'autonomie par rapport aux normes, tout simplement parce qu'il peut ne pas poursuivre les buts normatifs ou ne pas adopter une norme. Malgré le fait que les auteurs ne décrivent pas comment ce module peut prendre en considération les implications d'adopter ou pas une norme (implications sociales, mais aussi sur son autonomie), à notre connaissance cette modification de l'architecture DESIRE est la première architecture qui s'intéresse à ce problème.

(4) L'architecture NoA

L'architecture NoA proposée par Kollingbaum (Kollingbaum 2005) est une architecture pour un agent qui choisit et exécute un plan parmi un ensemble de plans possibles en fonction de ses croyances sur l'état du monde et des normes actives. Un langage NoA est proposé pour permettre la spécification des normes, celles-ci peuvent être des obligations, permissions ou prohibitions et concernent des actions ou des états du monde et qui ont des conditions d'activation et désactivation. Comme la Figure 4.2 le

montre, plusieurs modules dans l'architecture s'intéressent à la gestion des normes, tels que le module d'adoption de normes (Norm Adoption), le module d'activation de normes (Norm Activation) ou le module de filtrage à base de normes (Norm Filter). Un agent basé sur cette architecture choisit ses plans pour satisfaire les obligations actives, les plans (et les actions) choisis passent ensuite par un filtre des prohibitions et permissions qui vérifie la conformité aux normes.

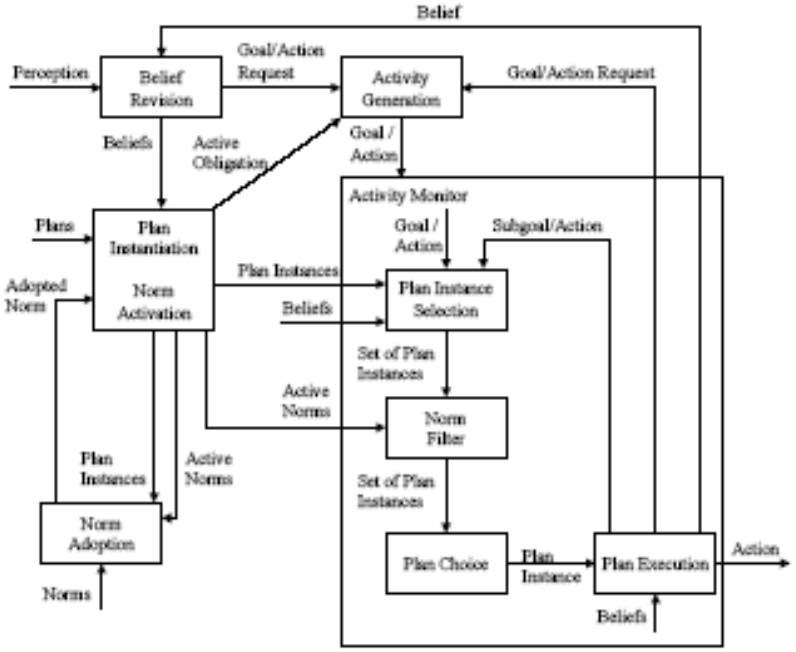


Figure 4.2. L'architecture NoA

En ce qui concerne l'autonomie par rapport aux normes adoptées, les agents NoA ne peuvent pas être autonomes. Une fois une norme adoptée, elle sera toujours respectée – si c'est une obligation, elle générera des plans qui la satisferont et si c'est une permission ou une prohibition, elle filtrera et rejettera les plans ou les actions en contradiction. Il est donc nécessaire pour le bon fonctionnement d'un agent de s'assurer que les normes qu'il a adoptées sont consistantes et qu'elles ne peuvent pas générer de contradictions. Pour cela, un agent NoA a de l'autonomie dans l'adoption des normes – il peut refuser l'adoption d'une norme. Nous discuterons plus tard le raisonnement effectué par un tel agent pour décider d'adopter ou pas une norme, pour l'instant nous remarquons seulement qu'une telle décision est nécessaire pour éviter les conflits entre les normes.

Synthèse

Ce passage en revue de plusieurs architectures d'agents proposées dans la littérature illustre comment ces architectures ont évolué pour aborder de nouveaux problèmes. Notamment en lien avec notre travail, différents modules sont apparus dans les architectures, modules qui donnent aux agents différents types d'autonomie, dans les interactions avec d'autres agents, mais aussi dans des organisations. Nous avons illustré comment des agents de plus en plus complexes sont capables de nouveaux types de raisonnement, tels que poursuivre ou pas un but généré par une norme ou adopter ou pas une norme. Par la suite nous synthétisons ce passage en revue en mettant en évidence les différents types de raisonnement qui existent dans les architectures présentées, pour analyser ensuite plus attentivement ceux qui nous intéressent dans cette thèse.

4.2. Types de raisonnement

Dans (Wooldridge 1999), Wooldridge décrit simplement le fonctionnement d'un agent en considérant que l'agent choisit une action à effectuer en fonction de sa perception sur l'extérieur (environnement, autres agents, etc.) et son état interne. Ce choix de comportement a de multiples facettes, comme l'illustre les multiples architectures d'agents existantes. Malgré leur finalité commune qui est de choisir une action à effectuer, les décisions qu'un agent prend sont diverses, allant de choisir un but à poursuivre, un plan à effectuer, adopter ou pas une norme ou un but d'un autre agent, à obéir ou pas à une norme, etc. Plusieurs catégories de raisonnement effectué par un agent semblent donc exister ; pour mieux décrire ces types de raisonnement nous allons nous appuyer sur la description du même système utilisé dans le Chapitre 1.

Ce système, basé sur celui proposé par (Bourne *et al.*, 2000), contient des agents qui se déplacent sur une grille pour trouver et exécuter des tâches. Ces tâches peuvent être simples (ST) et effectuées par un seul agent ou coopératives (CT) et effectuées par l'effort conjoint de plusieurs agents. Les agents ont donc besoin d'interagir pour former des accords pour s'aider les uns les autres à effectuer des CT. Les agents peuvent appartenir à une société ou être en dehors de la société – les agents qui en font partie sont le sujet d'une norme qui leur interdit de violer les accords avec d'autres membres de la société. Un agent qui n'appartient pas à la société peut donc rompre son accord avec un autre agent, mais d'autres agents peuvent aussi rompre leurs accords avec lui sans violer aucune norme.

Considérons maintenant les raisonnements effectués par un agent dans ce système. Nous pouvons classer ces raisonnements en quatre catégories, que nous appelons *a priori*, *pendant*, *a posteriori* et *planification* (ordonnancement).

Raisonnement a priori

Il représente le raisonnement effectué *avant* de prendre une décision qui modifie complètement le contexte (social) d'un agent sur les implications de cette décision. Dans le cadre du système considéré, c'est le raisonnement d'entrer ou non dans la société, c.-à-d., notamment d'adopter ou pas la norme, de ne pas rompre ses accords. Un agent doit évaluer les conséquences de cette décision avant de la prendre : s'il entre dans la société son comportement sera limité par la norme, mais il sera aussi protégé par la norme. Un autre exemple de ce type de raisonnement est le choix des partenaires d'interaction avant d'interagir avec eux. Quand un agent trouve une CT, il doit choisir avec qui interagir pour arriver à un accord de coopération pour exécuter la tâche. Plusieurs facteurs peuvent influencer ce choix, tels que la disponibilité des autres ou leur réputation. Comme nous le verrons par la suite, un des facteurs qui influence ce raisonnement est l'autonomie en délibération gagnée ou perdue par un agent. Un autre exemple de ce type de raisonnement, présent dans la littérature, est le raisonnement effectué dans la formation des coalitions (Sichman, 1995). Dans ce cas, un agent prend la décision de former une coalition avec d'autres agents, ce qui modifiera les interactions qu'il aura avec des agents membres de la coalition ou les buts qu'il poursuivra.

Raisonnement pendant

Il représente le raisonnement effectué dans le cadre des décisions prises suite au raisonnement *a priori*. Si un exemple de raisonnement *a priori* est le choix d'un partenaire d'interaction, ce raisonnement est celui effectué pendant l'interaction avec le partenaire choisi. Dans le cadre du système considéré, c'est le raisonnement utilisé pendant la négociation des caractéristiques d'un accord (p.ex. : échéance d'exécution d'une tâche, récompense donnée aux participants, etc.). L'exemple le plus connu de ce type de raisonnement est celui des stratégies de négociation utilisées par les agents dans le cadre du commerce électronique (Langer 2002). Ce type de raisonnement est en général très dépendant du problème et du type d'agent considéré.

Raisonnement a posteriori

Il représente le raisonnement effectué sur le résultat des raisonnements précédents. Dans le cadre du système considéré, le résultat d'une interaction (des décisions prises pendant la négociation) peut être la formation d'un accord pour l'exécution d'une CT. Une décision doit être maintenant prise de satisfaire ou pas l'accord, en prenant en compte si l'agent est le sujet d'une norme qui lui interdit de rompre ses accords. L'agent doit donc choisir entre poursuivre le but généré par l'accord (exécuter la CT) ou violer l'accord et donc la norme (s'il fait partie de la société). Comme nous le verrons par la suite, ce raisonnement s'effectue en général sur le fait d'être ou de ne pas être dans un comportement autonome et dépend du type d'agent.

Raisonnement de planification/ordonnancement

Il représente le raisonnement de base d'un agent. Si pour les agents réactifs il est codé en dur, pour les agents délibératifs il est utilisé pour trouver un plan qui satisfait le but poursuivi par un agent ou qui ordonne les intentions créées en fonction de leur durée ou échéance pour pouvoir les effectuer. Dans le cadre du système considéré, c'est la planification effectuée par un agent pour se déplacer sur la grille et exécuter les tâches qu'il veut exécuter, tout en prenant en compte leur durée d'exécution, échéances et les accords avec d'autres agents dans le cas des CT. Ce raisonnement est loin d'être simple, surtout dans le cas général, où un agent doit coordonner son comportement avec d'autres agents. Parmi les diverses approches proposées dans la littérature pour ce type de raisonnement, nous citons ici les algorithmes GPGP (et leurs modèles associés) (Lesser *et al.*, 2001), cf. Chapitre 3.

Synthèse

Le système considéré comme exemple, malgré sa simplicité, nous permet de mettre en évidence plusieurs types de raisonnement qui, dans une forme ou une autre, se retrouvent dans la plupart des architectures ou modèles d'agents existants. De nombreux travaux et de larges communautés s'intéressent au raisonnement pendant les interactions et au raisonnement en planification/ordonnancement. Ceci est expliqué par deux domaines d'application très importants pour les systèmes multi-agents, le commerce électronique (voir par exemple la série des workshops AMEC[6]) et les applications d'agents coopérants qui doivent coordonner leurs efforts pour un but commun (des applications militaires ou de sauvetage – voir par exemple les compétitions

[6] Workshops « Agent Mediated Electronic Commerce » (AMEC) – workshops annuels depuis 199?

RoboRescue[7]), domaines d'application dans lesquels ces deux types de raisonnement sont essentiels. Les deux autres types de raisonnement, *a priori* et *a posteriori*, peuvent être liés à la notion d'autonomie et représentent les types de raisonnement auxquels ce manuscrit s'intéresse – ils seront donc analysés en détail par la suite.

4.3. Raisonnement *a priori* et *a posteriori*

Les contraintes imposées à un agent par d'autres agents ou par les structures institutionnelles limitent la prise de décision de l'agent, mais constituent aussi l'objet de ces décisions. Le raisonnement d'un agent est donc *effectué sur* et *influencé par* des contraintes qui peuvent être des deux types, interpersonnelles et institutionnelles. Dans cette section nous analysons les approches proposées dans la littérature pour les raisonnements a priori et a posteriori sur des contraintes interpersonnelles et institutionnelles respectivement, en commençant avec le raisonnement a posteriori qui est plus souvent rencontré dans la littérature.

4.3.1. Raisonnement *a posteriori* interpersonnel et institutionnel

Le raisonnement *a posteriori* est un raisonnement effectué sur des contraintes comportementales : *une fois la contrainte établie* – suite à une interaction ou issue d'une norme adoptée – un agent doit décider s'il se plie ou non à la contrainte, s'il suit ou non le cours d'action dicté par la contrainte. Le raisonnement *a posteriori* interpersonnel représente le raisonnement effectué après une interaction pour décider de poursuivre ou non le résultat de l'interaction. Si, par exemple, nous considérons que le résultat d'une interaction est la création d'un engagement social par un agent, ce raisonnement est effectué pour décider si l'agent satisfait ou non l'engagement créé. Le raisonnement *a posteriori* institutionnel représente le raisonnement effectué par un agent pour décider d'obéir ou non à une norme ou à une autre contrainte institutionnelle. Les deux aspects du raisonnement *a posteriori*, interpersonnel et institutionnel, peuvent parfois être mélangés, comme dans le cas d'une norme adoptée qui spécifie l'obligation de satisfaire les engagements sociaux. Pour cette raison, les modèles de raisonnement que nous analysons ici ne peuvent pas être toujours différenciés en raisonnement interpersonnel et institutionnel. Il existe deux grandes catégories d'approches utilisées dans ce genre de raisonnement, celles basées sur l'utilité et celles symboliques.

(1) Raisonnement a posteriori basé sur l'utilité

Une des approches classiques pour le raisonnement d'agents est le raisonnement basé sur l'utilité, raisonnement très fréquent dans la communauté multi-agents à cause de sa simplicité : un agent calcule l'utilité de satisfaire et de ne pas satisfaire les contraintes interpersonnelles existantes et choisit le comportement avec la plus grande utilité. Parmi la littérature importante concernant ce type de raisonnement, nous analysons ici l'approche de Glass et Grosz (Glass and Grosz 2000), qui utilisent ce raisonnement, identifient ses limitations et proposent une approche légèrement différente pour compenser ces limitations.

[7] RoboRescue competition – www.roborescue.org

Considérons, par exemple, le cas d'un agent qui vient de créer un engagement social et qui doit décider de poursuivre ou pas le but – objet de l'engagement. Il calcule ainsi l'utilité d'avoir le but satisfait, en prenant en compte les autres buts qu'il ne pourra pas satisfaire à cause des conflits avec ce but et les éventuelles sanctions associées à la satisfaction ou à la violation de l'engagement. Ce calcul d'utilité peut s'avérer très complexe, surtout si l'agent prend en compte un historique des situations passées dans lesquelles lui ou d'autres agents ont décidé de ne pas poursuivre les engagements créés. Finalement, une fois le calcul d'utilité effectué, la décision est facile à prendre – l'option avec la plus grande utilité est choisie et les autres sont ignorées.

Glass et Grosz argumentent que peu importent les facteurs pris en considération dans un raisonnement basé sur l'utilité, ce raisonnement est trop simpliste – nous ne pouvons réduire la décision de poursuivre ou pas une contrainte interpersonnelle à une simple utilité, à un simple nombre. Comme exemple, ils considèrent le cas d'un agent qui décide de satisfaire une telle contrainte non parce que cela lui est utile, mais parce qu'il est motivé par le besoin de « s'intégrer socialement ». Pour donner une mesure quantitative de cette intégration sociale, ils utilisent des « brownie points » qui sont gagnés par un agent qui agit socialement, c.-à-d., qui satisfait les contraintes sociales. Ces points peuvent être ainsi considérés comme un capital social acquis par un agent, capital qui peut constituer une fin en elle-même ou qui peut être utilisé pour aider l'agent dans ses interactions futures.

Nous partageons le point de vue de cette approche et nous considérons que ce n'est pas toujours une bonne solution de réduire le composant social du raisonnement d'un agent à une simple partie de l'utilité qu'il associe à un comportement. De plus, cette approche permet d'identifier plusieurs types d'agent en fonction du poids associé à l'utilité individuelle et aux points sociaux dans la prise de décision. Nous pouvons ainsi identifier des agents individualistes (self-centered) qui raisonnent en se basant seulement sur leur utilité, des agents sociaux ou altruistes qui ont comme objectif de maximiser les « brownie points » ou bien des agents mixtes qui utilisent les deux critères dans leur raisonnement.

(2) Raisonnements a posteriori symboliques

Comme nous l'avons précisé, plusieurs modèles de contraintes institutionnelles comportementales existent, les engagements sociaux n'étant qu'un d'entre eux. Par exemple, dans le modèle B-DOING (Dignum *et al.*, 2001) les contraintes comportementales sont exprimées sous la forme d'obligations (contraintes interpersonnelles) et de normes (contraintes institutionnelles). Ce modèle utilise un module de gestion des buts, module responsable avec le raisonnement *a posteriori* de la satisfaction de ces contraintes. La décision de poursuivre ou pas une de ces contraintes est effectuée par ce module qui rajoute ou supprime dans la liste des buts de l'agent le but généré par une telle contrainte. Les auteurs du modèle ne précisent pas comment ce raisonnement est effectué, mais ils proposent une logique associée au modèle, donc probablement un raisonnement symbolique (ou au moins basé sur une représentation symbolique).

Un autre modèle relativement similaire au modèle B-DOING est le modèle et architecture BOID (Broersen *et al.*, 2001). Ce modèle considère un agent composé de quatre modules contenant ses croyances, obligations, désirs et intentions. Les obligations dans ce cadre représentent les contraintes comportementales interpersonnelles et institutionnelles et le raisonnement de l'agent est un raisonnement pour la résolution des conflits entre ces contraintes et les désirs, croyances et intentions

propres. Ce raisonnement est aussi symbolique – des opérateurs logiques sont définis et associés à chaque module – mais une relation de préférences est aussi utilisée pour définir des priorités. Le modèle BOID permet au concepteur d'agents de définir ainsi plusieurs types d'agents en définissant les priorités associées à chaque module. Par exemple, les agents sociaux ont le module des obligations plus prioritaire que le module des désirs, situation inversée pour les agents individualistes.

Synthèse

Le raisonnement *a posteriori* d'un agent est effectué pour décider d'obéir ou de satisfaire à une contrainte comportementale. Nous l'appelons *a posteriori* parce que cette contrainte est une contrainte *établie* par l'agent suite à une interaction passée avec un autre agent ou une organisation, donc le raisonnement est effectué *après* avoir établi une contrainte. Conforme au Chapitre 1, les contraintes établies ou comportementales sont liées à la perspective externe sur autonomie, perspective qui s'intéresse au comportement autonome d'un agent. Plus précisément, un agent a un comportement autonome s'il ne respecte pas une contrainte établie. Ceci revient au fait qu'*un agent effectue un raisonnement a posteriori pour décider d'être ou ne pas être autonome par rapport à une contrainte*. Comme nous l'avons vu, cette décision dépend fortement du type d'agent considéré, les agents individualistes et les agents sociaux utilisant des critères différents pour évaluer les conséquences de leur comportement autonome.

4.3.2. Raisonnement *a priori*

Le raisonnement *a priori* est effectué par un agent avant de prendre une décision d'établir une contrainte. Il porte sur les conséquences possibles ou probables de cette décision. Son objectif est généralement d'évaluer comment les contraintes qui influencent la prise de décision de l'agent se modifient en fonction de la décision prise par l'agent. Si ces contraintes se réfèrent aux interactions entre agents nous parlons d'un *raisonnement a priori interpersonnel*. Un exemple de ce type de raisonnement est le choix de partenaires d'interaction – c'est un raisonnement *a priori* parce qu'il est effectué avant d'interagir et s'intéresse à l'interaction dans laquelle l'agent aura le plus de chances de la conclure avec succès, c.-à-d., une interaction dans laquelle il sera pas contraint, mais ses partenaires si. Le *raisonnement a priori institutionnel* est effectué avant de prendre une décision d'adopter ou pas une norme, de jouer ou pas un rôle ou d'entrer/sortir ou pas d'une organisation, etc. Toutes ces décisions modifient le contexte social de l'agent en créant ou supprimant des contraintes imposées sur sa prise de décision. Ces deux types de raisonnement seront analysés par la suite.

4.3.2.1. Raisonnement *a priori* interpersonnel

Ce raisonnement est généralement effectué avant une interaction pour choisir les partenaires d'interaction. Le choix des partenaires est important parce que les relations existantes entre les agents peuvent influencer la prise de décisions pendant l'interaction – par exemple, pour un agent il peut s'avérer plus facile de négocier avec un agent qu'avec un autre. Ce choix de partenaire est également important dans la perspective post-interaction – par exemple un agent peut considérer que l'accord créé suite à une interaction aura plus de chances d'être satisfait par un agent que par un autre.

Généralement, un agent interagit avec d'autres agents parce qu'il ne peut pas satisfaire ses buts tout seul et il a besoin de l'aide des autres agents. Une approche existante sur le raisonnement *a priori*

interpersonnel est donc de lier ce raisonnement au besoin d'interaction qui pousse l'agent à interagir. Comme nous l'avons présenté dans le Chapitre 3, Sichman (Sichman 1994) identifie des relations de dépendance entre les agents. Les agents interagissent ensuite pour minimiser leurs dépendances, c.-à-d., pour obtenir l'aide des agents dont ils dépendent. En plus de proposer une formalisation de ces relations de dépendance, Sichman propose aussi un mécanisme pour identifier les situations de dépendance d'un agent et de choisir des partenaires d'interactions à l'aide de ces situations. Un agent calcule quelles sont ses situations de dépendance, mais aussi celles d'autres agents peut effectuer un raisonnement *a priori* interpersonnel. L'agent a ainsi l'intérêt d'interagir avec les agents dont il dépend – agents qui peuvent l'aider à satisfaire ses buts. De plus, si un agent dépend à son tour de lui (dépendance mutuelle ou réciproque), l'agent aura plus de chances d'interagir avec succès – un échange d'aide pouvant être mis en place. Ce raisonnement est limité par les connaissances incomplètes des agents sur l'environnement et sur les autres agents, mais il est néanmoins très utile pour les agents qui peuvent ainsi augmenter le nombre des buts satisfait en effectuant des interactions pertinentes. Un raisonnement similaire est proposé par (Ashri *et al*., 2003) qui identifient plusieurs types de relations similaires aux dépendances et qui proposent des agents qui raisonnent sur le management de ces relations.

Une autre approche existante pour le raisonnement *a priori* interpersonnel est celle de (Munroe *et al.*, 2005), qui s'intéresse à la sélection des partenaires de négociation à l'aide de l'autonomie « motivationnelle ». Des agents négociateurs doivent choisir un partenaire de négociation parmi plusieurs agents disponibles. Ce choix est effectué en se basant sur un historique de négociations passées et sur les motivations des agents. Un agent négocie pour satisfaire des buts qui à leur tour satisfont ses motivations (qui peuvent avoir des priorités associées). Un raisonnement basé sur l'utilité identifie les partenaires possibles avec qui les négociations précédentes ont satisfait le plus de motivations – ces partenaires de négociation sont ensuite préférés aux autres. Cette approche s'intéresse à l'autonomie « motivationnelle » parce qu'elle continue le travail de (Luck and d'Inverno 2001), où l'existence des motivations est considérée comme une garantie pour l'autonomie des agents.

D'autres approches pour le choix des partenaires d'interaction ne s'intéressent pas à améliorer la prise de décision pendant l'interaction, mais à améliorer la probabilité de former des accords qui seront ensuite poursuivis par les autres agents. Notamment, plusieurs travaux s'intéressent aux notions de confiance ou de réputation, notions qui sont utilisées dans ce type de raisonnement. Nous citons ici les travaux de Sabater (Sabater 2002), qui identifie aussi des relations existantes entre les agents d'un système et qui propose des modèles de confiance et de réputation dans ces agents. En ce basant sur ces relations et modèles, un agent peut raisonner sur la probabilité qu'un autre agent remplisse sa partie d'un potentiel accord et donc choisir des partenaires d'interaction auxquels il fait confiance.

Synthèse

Ces approches présentées peuvent être divisées en deux catégories : celles qui s'intéressent au choix des partenaires qui permettront à un agent de créer les meilleurs accords possibles – comme celle de Sichman (basée sur les relations de dépendance) ou de Monroe *et al*. (basée sur les motivations) et celles qui s'intéressent au choix des partenaires qui sont les plus probables à satisfaire les accord créés – comme celle de Sabater (basée sur la confiance et la réputation). Nous voulons souligner que la première catégorie peut être facilement associée à la notion d'autonomie : comme nous l'avons discuté

dans le Chapitre 1, l'autonomie peut être vue comme une indépendance (Castelfranchi 2001) ou comme la présence des motivations (Luck & d'Inverno 2001). Cette relation entre le raisonnement *a priori* interpersonnel et l'autonomie sociale peut être expliquée par l'objectif de ce raisonnement. Un agent raisonne et choisit un autre agent pour interagir avec lui ; comme pendant l'interaction il doit prendre des décisions, il cherche à interagir avec un agent par rapport à qui il a le plus d'autonomie, c.-à-d., il est libre de prendre les décisions qu'il veut sans être influencé par l'autre agent. Ou, si c'est possible, il cherche à interagir avec un agent qui n'a pas d'autonomie par rapport à lui, donc qui est limité dans sa prise de décision pendant l'interaction.

4.3.2.2. Raisonnement *a priori* institutionnel

Le raisonnement *a priori* institutionnel est effectué par un agent pour évaluer comment des contraintes institutionnelles sont créées ou modifiées suite à ces décisions. Comme nous l'avons vu auparavant, ces contraintes peuvent être de divers types, comme l'existence d'une norme, le fait que l'agent joue un rôle ou appartient à une organisation, etc. Ce raisonnement est donc utilisé par un agent pour décider d'adopter ou pas une norme, de jouer ou pas un rôle (ou de renoncer à un rôle qu'il joue déjà), d'entrer ou pas dans une organisation, etc. Ce raisonnement est particulièrement important pour un agent parce que la décision prise a une grande influence pour la suite : des nouvelles contraintes apparaissent et sont imposées à l'agent qui sera donc limité dans sa prise de décision ou son comportement (p.ex. il devra obéir les règles de l'organisation dans laquelle il est entrée).

Ce raisonnement *a priori* institutionnel a deux facettes. La première est de reconnaître les normes existantes et de les accepter en tant que telles (les adopter), tandis que la deuxième est plutôt de choisir parmi plusieurs cadres normatifs. La première est analysée par exemple par Conte et Castelfranchi dans (Conte & Castelfranchi 1999) qui s'intéressent au processus de raisonnement par lequel un agent reconnaît une norme comme étant une norme, l'adopte et est conscient que dans un certain contexte la norme est active et doit être appliquée. Ce raisonnement est particulièrement intéressant dans le cadre des normes émergentes, où les normes ne sont pas spécifiées explicitement, mais les agents observent le comportement des autres et déduisent l'existence des normes qu'ils adoptent ensuite.

Un problème similaire est adressé par Kollingbaum (Kollingbaum 2005) dans le cadre de l'architecture NoA que nous avons décrit auparavant. Même si les agents NoA n'ont pas un raisonnement institutionnel *a posteriori* (ils ne décident pas d'obéir ou pas aux normes adoptées), ils ont un raisonnement institutionnel *a priori* parce qu'ils décident d'adopter ou pas une norme. Ce raisonnement est effectué avec le seul but d'assurer la consistance de l'ensemble de normes adoptées – une nouvelle norme est adoptée seulement si elle n'est pas en contradiction avec les normes déjà adoptées auparavant. Ce problème d'assurer la consistance des normes est très important, mais nous considérons qu'il est surtout de la compétence de l'organisation (ou l'organisme imposant des normes) de s'assurer que des normes conflictuelles ne sont pas imposées au même agent et non. De plus, dans ce travail nous prenons le point de vue d'un agent qui raisonne donc sur des normes explicitement représentées (les seules qu'il connaît) et nous partons de l'hypothèse quelles sont aussi consistantes[8].

[8] Même si cette thèse ne s'intéresse pas aux problèmes tels que comment reconnaître des normes implicites, comment estimer s'il y a des conflits entre les normes, comment concevoir un système dans lequel les agents

La deuxième facette de ce raisonnement *a priori* institutionnel ne s'intéresse pas à la reconnaissance des normes ou des situations normatives, ni à la résolution des conflits entre les normes. Considérons maintenant des rôles et des structures organisationnelles bien définis, chaque rôle étant associé à un ensemble de normes. Le raisonnement *a priori* auquel nous faisons référence est le raisonnement utilisé par un agent afin de choisir parmi plusieurs rôles à sa disposition. Par exemple un agent qui ne joue aucun rôle a le choix entre plusieurs rôles possibles ou un agent qui joue déjà un ou plusieurs rôle doit décider de ne plus le(s) jouer. Ce problème peut être étendu au cas où un agent devrait choisir d'entrer ou pas dans une organisation (et jouer des rôles dedans) ou l'agent veut quitter l'organisation. Nous considérons que ce raisonnement est particulièrement important, surtout dans le cadre des systèmes ouverts, où les agents sont censés entrer à tout moment – ils devraient être capables d'évaluer avant d'entrer les implications de cette entrée et donc de prendre une décision pertinente.

Malheureusement, malgré son importance, peu de travaux s'intéressent à ce type de raisonnement. A notre connaissance, seule la thèse de Lopez y Lopez (Lopez y Lopez 2003) adresse directement le problème de comment un agent raisonne sur le fait d'entrer ou pas dans une organisation. Ce travail étend le travail de Luck et d'Inverno (Luck and d'Inverno, 2001) qui proposent une formalisation d'un système multi-agent dans laquelle les agents autonomes sont dotés des motivations qui représentent le moteur de leur raisonnement : le comportement des agents a comme objectif de satisfaire ces motivations. Lopez y Lopez propose ainsi l'utilisation des normes pour limiter l'autonomie des agents et assurer donc un comportement cohérent du système – ces normes sont définies formellement en précisant parmi d'autres choses les agents qui sont concernés, les buts que ces agents doivent satisfaire et le contexte dans lequel les normes sont actives. Ensuite, l'auteur décrit le raisonnement effectué par un agent qui veut entrer dans une organisation : il identifie les normes qui lui seront associées (parce qu'il jouera un ou plusieurs rôles) et analyse ces normes par rapport à ces motivations. Un agent cherche à entrer dans une organisation où les normes qu'il aura lui demandent de satisfaire des buts qui aident à la satisfaction de ses motivations et ne lui demandent pas la satisfaction de buts contraires à ses motivations. Autrement dit, le raisonnement *a priori* institutionnel d'un agent consiste à choisir le rôle/organisation/normes qui lui permet (ou au moins qui ne l'empêche pas) de satisfaire ses désirs (motivations). Un inconvénient de ce raisonnement réside dans le contexte d'activation de normes – les normes peuvent n'être actives que dans un contexte, donc il est généralement impossible pour un agent de savoir *a priori* quelles seront les normes actives et donc si ses motivations seront satisfaites ou non par elles.

A notre avis, ce raisonnement est aussi un peu simpliste, dans le sens où nous considérons qu'il y a d'autres raisons pour un agent d'entrer dans une organisation autres que de satisfaire ses désirs concrets et d'autres critères peuvent être utilisés pour analyser un rôle avant de décider de le jouer ou pas. Nous pensons notamment à la notion d'autonomie – si le raisonnement *a priori* interpersonnel semble utiliser cette notion, pourquoi pas celui institutionnel aussi ? Ceci est motivé par le fait que le raisonnement *a priori* est un raisonnement effectué sur l'évolution des contraintes imposées à l'agent, contraintes qui limiteront sa liberté de décision, c.-à-d., son autonomie.

---

peuvent résoudre ces conflits, etc., ceux-ci restent des problèmes de recherche intéressants et importants à résoudre dans le but de la conception des organisations ouvertes et dynamiques.

4.4. Synthèse

Dans ce chapitre nous avons analysé une partie des architectures et modèles d'agents proposés dans la littérature, en mettant en évidence comment ils ont évolué pour pouvoir résoudre des problèmes de plus en plus complexes tels que comment gérer les interactions avec d'autres agents (agent sociaux) ou comment adopter des nouvelles normes et se conformer à elles (agents normatifs). Cette analyse d'architectures nous a permis d'identifier plusieurs catégories de raisonnement effectué par les agents, raisonnements que nous avons appelés *a priori*, pendant, *a posteriori* et planification. Par exemple, dans le cadre des interactions entre agents, ces raisonnements sont effectués respectivement avant d'interagir (pour choisir un partenaire de négociation), pendant l'interaction (décisions prises pendant une négociation), après une telle interaction (pour poursuivre ou pas l'accord formé dans l'interaction) et, si avoir décidé de poursuivre l'accord, la formation et exécution d'un plan qui satisfait cet accord.

Les types de raisonnement que nous retenons pour ce travail et qui sont en lien avec la notion d'autonomie sont les raisonnements *a priori* et *a posteriori*. Ce dernier s'intéresse à la décision d'obéir ou pas une contrainte établie (comportementale), tandis que le premier s'intéresse à prendre des décisions qui minimisent les contraintes potentielles sur le processus de décision d'un agent. Comme ces contraintes sont issues des interactions entre agents (interpersonnelles) ou des structures normatives (institutionnelles), nous pouvons donc identifier des raisonnements *a priori* et *a posteriori* interpersonnels et institutionnels, comme dans le Tableau 4.1..

Tableau 4.1. Types de raisonnement

<table>
<tr><th>Raisonnement</th><th>interpersonnel</th><th>institutionnel</th></tr>
<tr><td rowspan="2">a priori</td><td rowspan="2">« avec qui interagir ? »

réseaux de dépendance (Sichman)
autonomie motivationnelle (Munroe)
confiance/réputation (Sabater)</td><td>« reconnaître les normes »

(Conte & Castelfranchi), (Kollingbaum)</td></tr>
<tr><td>« jouer ou non un rôle ? »

(Lopez y Lopez)</td></tr>
<tr><td rowspan="2">a posteriori</td><td>« remplir ou non mes engagements ? »

utilité + brownie points
(Glass & Grosz)</td><td>« obéir ou non aux normes ? »</td></tr>
<tr><td colspan="2">modified DESIRE, modèles BOID et B-DOING</td></tr>
</table>

Plusieurs approches existent pour effectuer le raisonnement *a posteriori*, raisonnement généralement effectué pour répondre à des questions de type « remplir ou pas mes engagements ? » ou « obéir ou pas à mes normes ? ».Les modèles de raisonnement existants sont généralement conçus pour répondre à des problèmes précis, mais ils peuvent être généralisés en essayant de répondre à une question qui généralise les deux précédentes : *« être ou ne pas être autonome ? »*. Cette généralisation des deux questions est motivée par les définitions et exemples d'autonomie que nous avons présentés dans le Chapitre 1.

En ce qui concerne le raisonnement *a priori*, plusieurs approches sont proposées pour répondre à la question « avec qui interagir ? » (raisonnement interpersonnel), approches qui sont liées plus ou moins directement à la notion d'autonomie – l'exception de Sabater est expliquée par le fait qu'il utilise un autre critère pour répondre à cette question. Le raisonnement *a priori* institutionnel s'intéresse à « comment reconnaître les normes existantes ? », mais aussi à « comment choisir entre plusieurs rôles/organisations/normes possibles ». Nous considérons que cette dernière question, malgré le manque d'intérêt accordé dans la communauté multi-agents, a une importance particulière dans le cadre des systèmes ouverts et nous nous y intéressons en détail dans la troisième partie de cette thèse.

Synthèse

Cette thèse s'intéresse au raisonnement social des agents dans des systèmes ouverts. Son objectif est de proposer un modèle de raisonnement à base d'autonomie, modèle qui permettra aux agents par exemple de décider s'ils veulent entrer dans un système ou non. Dans le dernier chapitre nous avons analysé quelques architectures et modèles d'agents proposés dans la littérature, ce qui nous a permis d'identifier plusieurs types de raisonnement effectués par les agents, raisonnements que nous avons appelé respectivement *a priori*, pendant, *a posteriori* et planification. Par exemple, dans le cadre des interactions entre agents, ces raisonnements sont effectués respectivement avant d'interagir (pour choisir un partenaire de négociation), pendant l'interaction (décisions prises pendant une négociation), après une telle interaction (pour poursuivre ou pas l'accord formé dans l'interaction) et, si avoir décidé de poursuivre l'accord, la formation et exécution d'un plan qui satisfait cet accord. Comme un autre exemple, le raisonnement *a priori* institutionnel s'intéresse à la décision qu'un agent prend avant d'entrer dans une organisation (système) : qu'est-ce qu'il va gagner ou perdre en entrant, lequel parmi les rôles qu'il peut jouer est le plus avantageux pour lui, etc. ?

Nous avons argumenté que les raisonnements *a priori* et *a posteriori* sont liés à la notion d'autonomie de décision et que même si d'autres approches sont possibles et/ou utilisées, il semble intéressant de baser ces types de raisonnements sur un seul concept, l'autonomie. Ce concept est utile pour le raisonnement *a priori* institutionnel (p.ex. : choisir comme partenaire d'interaction celui avec qui l'agent a le plus d'autonomie, pour augmenter les chances de former un accord qui lui convient) ou dans le raisonnement *a posteriori* qui constitue la décision d'un agent d'être ou non dans un comportement autonome.

Il existe peu de travaux qui s'intéressent au *raisonnement a priori institutionnel*, notamment à cause de la difficulté d'évaluer les gains ou les pertes associés au fait de jouer un rôle, vu que généralement les spécifications d'un rôle sont dépendantes de contexte. Notre approche à base d'autonomie sera donc certainement bénéfique pour ce type de raisonnement : un agent pourra évaluer les conséquences que jouer un rôle aura sur son autonomie sociale de décision et prendre ainsi des décisions plus informées. Cependant, il existe souvent des confusions sur la signification de la notion d'autonomie et des nombreuses définitions qui y sont associées – il faudra donc définir formellement l'autonomie des agents afin qu'elle puisse être utilisée dans leurs raisonnements.

Nous avons présenté dans le Chapitre 1 quelques propriétés de l'autonomie, ce qui nous a permis d'identifier plusieurs types d'autonomie dans les systèmes multi-agents. Comme nous voulons utiliser l'autonomie pour améliorer la prise de décision des agents dans une société, nous nous sommes intéressés de plus près à l'*autonomie sociale de décision*. Suite aux différentes définitions de ce concept proposées par la communauté multi-agents, nous avons identifié deux points de vue différents sur l'autonomie : *avoir de l'autonomie* représente la liberté de décision d'un agent, tandis que *être autonome* décrit le comportement d'un agent qui ne respecte pas les contraintes qui lui sont imposés. L'autonomie représente ainsi la réponse d'un agent aux contraintes sociales qui lui sont imposées par d'autres agents ou par l'organisation à laquelle il appartient.

Si nous voulons obtenir des agents capables de raisonner en utilisant cette notion, c.-à-d., de choisir un cours d'action en fonction de comment ils modifient leur autonomie, nous avons besoin de quantifier et qualifier l'autonomie des agents. Nous pouvons le faire à l'aide des contraintes qui la limitent. Ces contraintes sont issues du besoin de coordination qui existe dans le système. Conforme Chapitre 1, elles peuvent être classées en des contraintes interpersonnelles (provenant d'autres agents) et institutionnelles (provenant des mécanismes organisationnels) ou bien dans des contraintes décisionnelles (potentielles) et comportementales (établies) si elles limitent la prise de décision des agents ou leur comportement. Afin de créer des agents capables de raisonnements a posteriori et surtout a priori dans des systèmes ouverts et donc afin de modéliser la notion d'autonomie, nous avons besoin de modéliser les différents types de contraintes de manière à pouvoir calculer le degré d'autonomie d'un agent dans différentes hypothèses. La relation entre les types de contraintes et les types d'autonomie présentée dans le chapitre 1 est rappelée dans le tableau ci-dessous:

Tableau 4.2. Types de contraintes et d'autonomie

Types de contraintes	**interpersonnelle**	**institutionnelle**
décisionnelle (potentielle, inhérente)	**perspective interne : avoir de l'autonomie (dans sa prise de décision)**	
	ex: la décision d'un agent d'adopter ou pas le but d'un autre agent est influencée par ce deuxième	ex: la décision d'un agent d'adopter une norme est influencée par le rôle qu'il joue ex: la décision d'un agent de poursuivre un but, est influencée par une norme qu'il a
comportementale (établie)	**perspective externe : être autonome (dans son comportement)**	
	ex: un agent s'est engagé envers un autre de satisfaire un de ses buts – s'il viole son engagement il est dans un comportement autonome	ex: un agent a adopté une norme qui l'oblige de satisfaire un but – s'il ne le satisfait pas (viole la norme), il est dans un comportement autonome

En ce qui concerne les contraintes interpersonnelles – provenant des interactions avec d'autres agents, nous avons présenté quelques modèles qui peuvent les représenter. La théorie de la dépendance décrit les raisons qui poussent les agents à interagir les uns avec les autres : ils ne peuvent pas satisfaire tout seuls leurs buts et ils ont besoin d'aide. Une telle dépendance limite la prise de décision d'un agent et ainsi son autonomie. Une extension à cette théorie est la théorie du pouvoir social qui va plus loin, en modélisant aussi les pouvoirs qu'un agent a sur un autre (ce qui limite aussi l'autonomie du deuxième agent) et en s'intéressant aussi aux relations qui existent entre les agents dans un contexte institutionnel et non seulement interpersonnel.

Si ces deux approches permettent la représentation des contraintes qui poussent les agents à interagir, le paradigme des engagements sociaux est souvent utilisé pour décrire les contraintes qui apparaissent après les interactions, quand un agent devient engagé vers un autre pour satisfaire un but (ou exécuter une action, etc.). Ces contraintes limitent le comportement d'un agent qui doit satisfaire le but pour

lequel il est engagé, sinon il est considéré autonome et souvent subit des conséquences (sanctions). Nous avons également décrit des meta-engagements, appelés des politiques sociales, qui ont une représentation identique aux engagements sociaux, mais qui représentent un concept similaire à celui de norme. Comme nous avons pu le voir, les normes sont souvent utilisées par les modèles organisationnels pour limiter le comportement des agents qui jouent des rôles. Les rôles décrivent le comportement attendu des agents qui les jouent, en termes des buts à satisfaire et normes à obéir et des hiérarchies de rôles sont souvent mises en place. Le fait de jouer un rôle (et ainsi d'être le sujet des normes, d'avoir une place dans une hiérarchie, etc.) impose des contraintes à un agent, contraintes qui peuvent limiter sa prise de décision (p.ex. : adopter un but chaque fois que son supérieur dans la hiérarchie le demande) ou son comportement (p.ex. : il est interdit d'effectuer une action).

Les problèmes générés par cette diversité de sources de contraintes sociales imposées à un agent sont augmentés par le fait qu'il n'existe pas une représentation unique de ces contraintes. Le tableau ci-dessous présente les modèles de représentation de divers types de contraintes que nous trouvons particulièrement intéressants pour ce travail.

Tableau 4.3. Représentations des contraintes

<table>
<tr><th>Types de contraintes</th><th>interpersonnelles</th><th>institutionnelles</th></tr>
<tr><td rowspan="2">décisionnelles</td><td>théorie de la dépendance</td><td>hierarchies de rôles</td></tr>
<tr><td colspan="2">politiques sociales - normes
théorie du pouvoir social</td></tr>
<tr><td>comportementales</td><td>engagements sociaux</td><td>rôles
normes associées aux rôles
contrats pour jouer des rôles</td></tr>
</table>

Nous considérons que, afin de pouvoir raisonner sur l'autonomie et donc de baser ce concept sur les contraintes sociales qui le limitent, nous avons besoin d'une représentation plus uniforme de ces contraintes. Les liens qui existent entre les modèles présentés ci-dessus et leur pouvoir de représentation nous indiquent ainsi quelques pistes à suivre vers une représentation plus uniforme. Par exemple, les engagements sociaux sont similaires à la notion de contrat utilisée souvent dans les institutions électroniques et ils peuvent également représenter des politiques sociales qui sont similaires aux normes. De plus, même si elle décrit principalement des concepts interpersonnels, la théorie du pouvoir social peut être aussi utilisée dans un contexte institutionnel.

Dans la partie suivante nous décrivons le modèle proposé pour représenter d'une manière plus uniforme ces contraintes, ce qui nous permettra de définir formellement l'autonomie des agents. Ensuite, nous utiliserons ce modèle pour montrer comment un raisonnement à base d'autonomie peut être utilisé par les agents et quel est l'intérêt d'un tel raisonnement.

II. Modèle

Introduction

L'objectif de notre travail est d'enrichir le raisonnement d'un agent dans un système ouvert en lui permettant d'utiliser la notion d'autonomie dans certains types de raisonnement, comme par exemple avant de prendre une décision vis-à-vis d'une contrainte (*a priori*) ou après l'avoir prise (*a posteriori*). Nous envisageons des agents qui utilisent un tel raisonnement dans les interactions avec d'autres agents (contexte interpersonnel), mais aussi dans leurs interactions avec une organisation (contexte institutionnel). Par exemple, un agent pourra évaluer les conséquences que jouer un rôle implique sur son autonomie et prendre ainsi des décisions plus informées sur le choix d'un rôle à jouer ou d'entrer ou non dans une organisation (système). Cependant, il existe souvent des confusions sur la signification de la notion d'autonomie et cette notion manque de définition formelle et utilisable dans un tel raisonnement. Dans le Chapitre 1 nous avons argumenté la relation qui existe entre la notion d'autonomie et les contraintes interpersonnelles et institutionnelles qui limitent la prise de décision et le comportement des agents.

Pour permettre aux agents de raisonner en utilisant la notion d'autonomie, nous devons d'abord définir formellement cette notion : notre approche consiste à baser cette définition sur ces contraintes imposées aux agents. Pour être plus précis, afin de pouvoir raisonner sur (l'évolution de) son autonomie, un agent doit être capable de représenter les contraintes qui lui sont imposées de l'extérieur. La grande diversité des types de contraintes, ainsi que les différents modèles proposés dans la littérature nous mènent à chercher une représentation uniforme de ces contraintes : dans cette partie nous présentons le modèle que nous proposons pour représenter les contraintes qui limitent l'autonomie d'un agent, tandis que la partie suivante s'intéressera au raisonnement à base de ce modèle.

Avant de présenter le modèle formel qui nous permet de définir la notion d'autonomie, il nous semble très important de souligner une caractéristique de ce modèle : il s'agit dans ce travail d'*un modèle interne aux agents*. Nous ne proposons ainsi ni un modèle d'interactions entre agents, ni un modèle organisationnel ou institutionnel – autrement dit, nous ne proposons pas une approche pour concevoir un système multi-agents. Nous considérons que de tels systèmes multi-agents existent et la coordination dans ces systèmes est atteinte d'une manière ou d'une autre (voir la partie précédente pour des différentes approches existantes dans la littérature). De plus, nous ne proposons pas non plus de concevoir un agent capable de fonctionner correctement dans de tels systèmes : en règle générale la conception des agents suit de près la conception du système. Par exemple, si dans un système multi-agents la communication se fait à base des actes de langages, les agents dans ce système sont conçus pour comprendre ces actes et pouvoir communiquer entre eux.

Au travers du modèle que nous proposons, nous représentons du point de vue d'un agent les mécanismes de coordination employés dans un système ou, plus précisément, l'effet de ces mécanismes sur la prise de décision ou le comportement de l'agent. Par exemple, considérons une organisation multi-agents définie à l'aide du modèle MOISE+ (Hubner 2003) et un agent qui joue un rôle dans cette organisation. Bien évidemment, l'agent doit être capable de comprendre la structure

organisationnelle et ce qui est attendu de lui étant donné le rôle qu'il joue. Mais il est important pour l'agent d'évaluer la manière dont sa prise de décision sera limitée ou contrainte par le rôle qu'il joue, d'évaluer, avant même de jouer ou non un rôle, si ce rôle sera avantageux pour ses propres buts ou non. Le modèle que nous proposons a comme objectif de répondre à ce type de questions. Par ailleurs, le modèle doit fonctionner même si les rôles sont définis à l'aide d'un autre modèle organisationnel, par exemple (Dastani *et al.* 2003). En fait, ce modèle interne à l'agent ne décrit pas un rôle, mais les implications qui en découlent lorsque l'agent le joue.

Comme nous l'avons argumenté dans le Chapitre 1, l'existence des mécanismes de coordination dans un système multi-agents implique des contraintes qui limitent la prise de décision et le comportement des agents. Le modèle proposé dans cette partie représente ces contraintes et, comme elles sont directement liées à la notion d'autonomie, ce modèle nous permet de définir également cette notion. Même si nous ne proposons pas un modèle de coordination, le modèle interne que nous proposons utilise quelques-unes des approches les plus utilisées dans la littérature. Le chapitre suivant (Chapitre 5) décrit les bases de notre modèle : les notions fondamentales utilisées, telles que but, action, permission, etc. et la modélisation proposée pour les interactions entre agents. Comme nous l'avons précisé dans le Chapitre 3, le paradigme des engagements sociaux est souvent utilisé pour représenter le résultat des interactions : nous incorporons dans notre modèle ce paradigme et nous montrons comment les implications des interactions qui nous intéressent peuvent être représentées avec ce modèle par un agent.

Nous nous intéressons ensuite à la représentation des concepts institutionnels tels que ceux de rôle ou norme (introduits dans le Chapitre 2), concepts qui représentent la source de nombreuses contraintes sur le comportement des agents. Nous proposons ainsi dans notre modèle une représentation des implications de l'existence de ces concepts, représentation qui utilise le même paradigme des engagements sociaux et des méta-engagements (politiques sociales). Nous définissons ainsi à l'aide de ces concepts les contraintes que le fait de jouer un rôle peut imposer sur le comportement et sur la prise de décision d'un agent. Nous concluons le Chapitre 6 en décrivant comment des organisations peuvent être "vues" par un agent à travers notre modèle.

Comme présenté dans le Chapitre 3, plusieurs théories, telles que la théorie de la dépendance et celle du pouvoir social, s'intéressent aux contraintes interpersonnelles et abordent les contraintes institutionnelles. Dans le Chapitre 7 nous intégrons ces deux théories dans notre modèle, en les étendant et en proposant une définition formelle pour tous les concepts utilisés. Nous serons ainsi capables de définir les pouvoirs qu'un agent a de satisfaire seul ses buts et, le cas échéant, les relations de dépendances qui apparaissent entre un agent et d'autres agents ou entre un agent et des rôles. Nous montrerons ensuite comment ces relations de dépendance et les engagements et politiques sociales utilisées auparavant peuvent être considérées du même point de vue à travers la notion de pouvoir social.

Nous obtenons ainsi un modèle formel qui, en utilisant principalement des engagements sociaux et des relations de pouvoir social, permet à un agent de représenter d'une manière uniforme les divers types de contraintes qui lui sont imposées. Comme dans ce travail nous considérons l'autonomie comme un concept résultant directement de la présence ou de l'absence de ces contraintes, ce modèle débouche ensuite sur une définition formelle de l'autonomie d'un agent.

5. Fondements du modèle

Dans ce chapitre, nous décrivons les notions qui constituent la base de notre modèle, notions qui seront utilisées dans les chapitres suivants pour définir les contraintes qui limitent la prise de décision et le comportement des agents. Nous commençons ainsi par décrire les éléments constitutifs des agents, telles que les actions, buts, ressources ou plans. Même si dans ce manuscrit nous ne nous intéressons pas spécialement au processus de planification, dans ce chapitre nous insistons un peu plus sur la notion de plan ; comme nous le verrons dans le Chapitre 7, les relations qui existent entre les éléments d'un plan qu'un agent a pour satisfaire un but ont une grande importance sur les capacités de l'agent de satisfaire ses buts et ainsi sur ses relations avec d'autres agents.

Après avoir identifié les notions de base qui caractérisent les agents dans notre modèle, nous allons nous intéresser aux interactions entre agents. Comme nous l'avons précisé dans le Chapitre 3, les engagements sociaux sont souvent utilisés pour décrire le résultat des interactions entre agents. Dans ce chapitre nous introduisons le modèle d'engagements sociaux que nous utilisons dans notre travail. Nous présentons ensuite comment ces engagements peuvent être gérés dans un système en définissant les opérations qui les manipulent. Dans le Chapitre 1 nous avons précisé qu'un type d'interaction entre agents est particulièrement important du point de vue de l'autonomie sociale : la délégation/adoption des buts entre agents. Dans ce chapitre nous montrerons comment ce type d'interaction peut être réduit et représenté à l'aide d'engagements sociaux, ce qui nous offre une abstraction très utile des interactions entre agents. Ainsi, nous représentons les contraintes issues des interactions indépendamment du modèle d'interaction entre agents, qu'il soit à base des actes de langage ou autre.

5.1. Notions de base

La notation dans ce manuscrit utilise la logique des prédicats du premier ordre, bien que, comme nous le verrons plus tard, nous serons parfois obligés d'utiliser des prédicats d'ordre supérieur pour décrire les notions utilisées. Nous voulons préciser que la décidabilité de cette logique ne nous intéresse pas : nous ne l'utilisons que pour représenter d'une manière claire et précise les notions utilisées et non pour prouver des propriétés du modèle proposé. Comme souvent dans la littérature (voir par exemple les travaux de Sichman sur la théorie de la dépendance présentée dans le Chapitre 3), nous décrivons le comportement et le raisonnement des agents en utilisant les notions d'action, de ressource, de but et de plan. Par la suite nous décrivons ces notions de base et comment elles sont reliées aux agents.

(1) Actions, ressources, buts et plans

Dans ce manuscrit nous identifions avec des majuscules des ensembles, tels que l'ensemble de toutes les actions (*Act*), l'ensemble de toutes les ressources (*Res*), l'ensemble de tous les buts (*G*) et l'ensemble de tous les plans (*Pl*) existants et possibles dans le système. Nous notons les éléments de ces ensembles respectivement par les minuscules *a*, *r, g* et *pl*, avec des indices permettant de les différencier :

$$\text{Act}=\{a_1,a_2,...\}, \text{Res}=\{r_1,r_2,...\}, \text{G}=\{g_1,g_2,...\}, \text{Pl}=\{pl_1,pl_2,...\} \quad [5.1]$$

Des caractéristiques telles que la durée (d'une action), une échéaace ou date limite – deadline – (d'un but), etc., peuvent être associées à la plupart de ces notions. Les notations présentées ci-dessus ne concernent que le nom de ces notions, tandis que la représentation de leurs caractéristiques sera présentée dans les sections suivantes.

Nous regroupons ces quatre concepts de base, action, ressource, but et plan, sous l'appellation générique de *tâche* : nous notons par T l'ensemble de toutes les tâches possibles, ensemble représentant la réunion des ensembles d'actions, de ressources, de buts et de plans possibles. Nous notons les tâches par λ, avec des indices permettant de les différencier si nécessaire :

$$\mathrm{T}=\{\lambda_1,\lambda_2,...\}=\mathrm{Act}\cup\mathrm{Res}\cup\mathrm{G}\cup\mathrm{Pl} \qquad [5.2]$$

(2) Agents

Dans la suite de ce manuscrit, nous notons les agents par des lettres situées en fin de l'alphabet (ex. x, y,…), sauf dans le cas des noms concrets utilisés dans les exemples, et par Ag l'ensemble de tous les agents. Un agent sait effectuer un nombre d'actions, possède des ressources, poursuit des buts et connaît un ensemble de plans. Ces éléments propres à un agent x sont représentés respectivement par les ensembles (potentiellement vides) Act_x, Res_x, G_x et Pl_x :

$$\forall x\in \mathrm{Ag}:\mathrm{Act}_x=\{a_1,...\}\subseteq\mathrm{Act},\ \mathrm{Res}_x=\{r_1,...\}\subseteq\mathrm{Res},\ \mathrm{G}_x=\{g_1,...\}\subseteq\mathrm{G},\ \mathrm{Pl}_x=\{pl_1,...\}\subseteq\mathrm{Pl} \qquad [5.3]$$

Comme nous le verrons dans le chapitre suivant, le fait d'appartenir à une organisation et de jouer un rôle donne aux agents des permissions d'exécuter des actions, d'utiliser des ressources, de poursuivre et satisfaire des buts ou d'effectuer des plans. Même si nous ne nous intéressons pas pour l'instant à comment ces permissions sont obtenues, nous décrivons ici la notation utilisée : l'ensemble des permissions qu'un agent x a est noté par $Perm$. Cet ensemble contient des éléments des ensembles d'actions, ressources, buts et plans possibles, donc des tâches :

$$\forall x\in\mathrm{Ag}:\mathrm{Perm}_x\subseteq\mathrm{T} \qquad [5.4]$$

Nous soulignons que l'ensemble de permissions qu'un agent a ne correspond pas forcement à l'union des ensembles de ses actions, ressources, buts et plans. Il est en effet possible qu'un agent sache exécuter une action sans qu'il ait la permission de l'exécuter ou qu'il ait la permission d'utiliser une ressource sans qu'il la possède. Nous verrons dans le Chapitre 7 l'importance de ces aspects.

Nous essayons de garder notre modèle le plus générique possible en ne le liant pas à un modèle d'agent précis (p.ex. BDI). Cependant, nous avons besoin d'exprimer les croyances d'un agent : pour cela nous utilisons le prédicat *believes(agent,expression)*, prédicat qui dénote le fait qu'un agent croit qu'une expression logique est vraie. La signification de cette croyance est la même avec la sémantique des croyances du modèle BDI (voir Chapitre 4). Il est notamment possible qu'une croyance soit fausse, c.-à-d., que l'expression que l'agent croit vraie soit en effet fausse.

(3) Plans

Comme nous l'avons vu dans le Chapitre 4, une partie du raisonnement d'un agent consiste à, une fois un but choisi, trouver et effectuer un plan qui satisfait ce but. Même si ce manuscrit ne s'intéresse pas directement à ce processus de planification et d'ordonnancement, nous nous intéressons à la notion de plan qui peut avoir des implications intéressantes sur les pouvoirs d'un agent, comme nous le verrons plus tard. Dans le Chapitre 3 nous avons présenté le framework de coordination entre agents GPGP, qui consiste en une famille d'algorithmes de planification distribuée et qui est conçu pour utiliser la représentation des plans TAEMS. Ceci représente le plan (possiblement partiel) pour la satisfaction d'un but sous la forme d'une structure de tâches, où une tâche représente une action, un but ou une ressource. Un plan est donc composé de sous-buts à atteindre, d'actions à exécuter et de ressources à utiliser. Pour chaque sous-but il existe un ou plusieurs plans qui sont à leur tour formés par des sous-buts, etc. Cette récursivité s'arrête à un niveau élémentaire formé uniquement par des actions et ressources. De plus, des relations peuvent exister entre les différents éléments d'un plan (c.-à-d., des tâches), des relations telles que *enables, hinders*, etc.), relations détaillées dans le Chapitre 3.

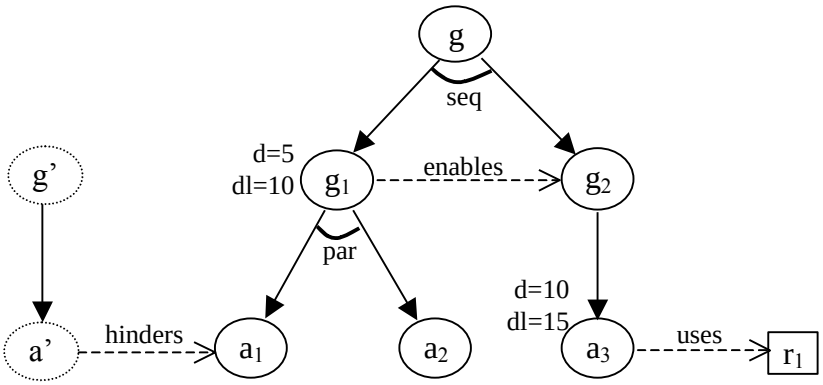


Figure 5.1. Plan représenté sous la forme d'une structure de tâches TAEMS

La Figure 5.1 décrit un plan représenté à l'aide du modèle TAEMS, plan pour la satisfaction d'un but *g* par la satisfaction de deux sous-buts g_1 et g_2 l'un après l'autre. La satisfaction de g_1 prend 5 unités de temps (d), a une date limite de 10 (dl). Elle sera atteinte en exécutant en parallèle (par) les deux actions a_1 et a_2. La satisfaction de g_2 n'est possible qu'après (seq) celle de g_1. Elle nécessite l'exécution de l'action a_3. Celle-ci a une durée de 10, une date limite de 15 et nécessite l'utilisation de la ressource r_1 (relation *uses* sur la figure). L'exécution de l'action a_1 est impossible si l'action *a'* est exécutée avant elle (cf. relation *hinders* sur la figure). Bien que *g'* et *a'* n'appartiennent pas au plan (cf. ligne en pointillés sur la figure), ils apparaissent dans la structure de tâches puisqu'ils influencent les éléments du plan.

Du point de vue des objectifs de ce manuscrit, la manière dont un agent planifie et représente ses plans ne nous intéresse pas – vue la grande puissance de représentation du modèle TAEMS nous pouvons considérer que ce modèle est utilisé dans le raisonnement de planification d'un agent. Nous ne nous intéressons pas non plus à l'ordonnancement des tâches dans un plan, mais à la manière dont les éléments du plan peuvent limiter le raisonnement ou le comportement d'un agent. Nous représentons ainsi un plan comme un ensemble constitué du but à satisfaire et des ensembles de sous-buts, d'actions et de ressources nécessaires :

$$pl_g = \{g, SG_g, Act_g, Res_g\}, \text{ où } g \in G, SG_g \subseteq G, Act_g \subseteq Act, Res_g \subseteq Res \quad [5.5]$$

Cette notation représente simplement quelles sont les tâches constituant un plan et n'est pas utilisable par un agent pour décider de l'action à exécuter en premier dans la mesure où toute notion de séquentialité disparaît. La notation pl_g désigne un plan pour satisfaire le but g, c'est-à-dire un plan qui contient l'élément g. Pour simplifier nos formules, nous utilisons la notation $\alpha_i \in pl_g$ au lieu d'écrire $g_i \in SG_g \in pl_g$ (resp. a_i , r_i), pour dénoter qu'une tâche (action, ressource, sous-but) appartient à un plan.

Cette représentation d'un plan sous la forme d'un ensemble permet d'identifier les éléments d'un plan, mais elle ignore les relations entre ces éléments. A partir du modèle TAEMS, nous retenons trois types de relations qui nous intéressent particulièrement : l'exécution d'une tâche *rend possible (enables)* l'exécution d'une autre, l'exécution d'une tâche *empêche (hinders)* l'exécution d'une autre et le *conflit* existant entre l'exécution de deux tâches au sein d'un même plan (*conflictS*).

$$enables(\alpha_1, \alpha_2), hinders(\alpha_1, \alpha_2), conflict(pl_g, \alpha_1, \alpha_2) \quad [5.6]$$

Ces trois relations entre tâches au sein d'un plan sont représentées à l'aide de trois prédicats :

- Le prédicat *enables* décrit le fait que l'exécution d'une tâche (l'exécution d'une action, la satisfaction d'un sous-but ou l'utilisation d'une ressource) est possible seulement si une autre tâche (appartenant ou non au même plan) a été exécutée avec succès avant. La tâche qui rend possible l'exécution d'une autre représente une action ou un but ($\alpha_1 \in Act \cup G$).

- Le prédicat *hinders* décrit le fait que l'exécution d'une tâche est possible seulement si une autre tâche (appartenant ou non au même plan) n'a pas été exécutée avec succès avant. La tâche qui empêche l'exécution d'une autre représente une action ou un but ($\alpha_1 \in Act \cup G$).

- Le prédicat *conflict* décrit le fait que dans un plan il existe un conflit entre les exécutions de deux de ses tâches. Ce conflit peut être provoqué par exemple par le fait que deux actions doivent être effectuées en parallèle ou qu'à cause de leurs durées et dates limites, deux actions ne peuvent pas être exécutées par le même agent. Les deux tâches qui se retrouvent en conflit représentent des actions ou des buts ($\alpha_1, \alpha_2 \in Act \cup G$).

Exemple 5.1

Considérons maintenant un exemple pour mieux illustrer la représentation des plans, buts, actions et ressources d'un agent. Cet exemple, que nous allons développer au cours des prochains chapitres, concerne le plan qu'un agent a pour satisfaire le but g « gagner de l'argent ». Une représentation graphique de type TAEMS de ce plan est présentée dans la Figure 5.2. Ce plan consiste à négocier les caractéristiques d'un travail et l'exécuter (action *exec_job*). La négociation concerne l'exécution parallèle de deux actions neg_1 et neg_2, actions correspondant à l'envoi et à la réception des messages par deux agents, chacun exécutant ainsi une des deux actions. L'exécution d'un travail (*exec_job*) peut être annulée par un autre agent en annulant le contrat créé (action *cancel*).

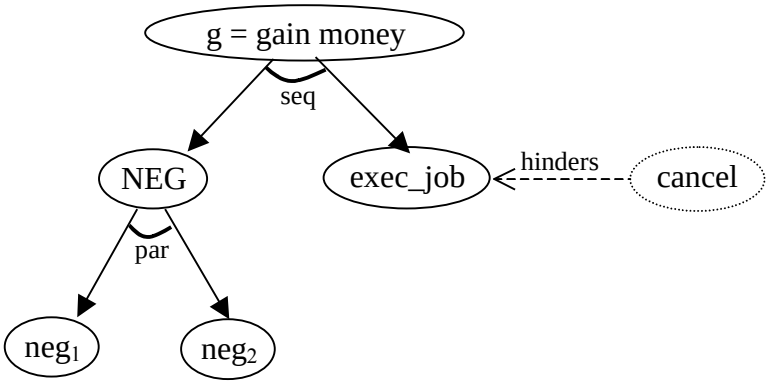


Figure 5.2. Exemple : plan pour gagner de l'argent

Considérons ce plan pour la satisfaction du but *g*, sa représentation sous la forme d'un ensemble de tâches avec les relations existant entre elles est présentée ci-dessous. Ce plan consiste dans un sous-but *neg*, une action *exec_job* n'impliquant aucune ressource avec l'exécution de l'action *exec_job* qui peut être empêchée par l'exécution d'une action *cancel* :

$$pl_g=\{g, SG_g, Act_g, Res_g\}, \text{où } SG_g=\{neg\}, Act_g=\{exec\_job\}, Res_g=\{\ \}$$
$$hinders(cancel, exec\_job)$$

En ce qui concerne la satisfaction du sous-but *neg*, elle peut être atteinte par l'exécution d'un plan qui ne contient aucun sous-but, deux actions et aucune ressource, un conflit existant entre les deux actions à cause du besoin de les effectuer en parallèle :

$$pl_{neg}=\{neg, SG_{neg}, Act_{neg}, Res_{neg}\}, \text{où } SG_{neg}=\{\ \}, Act_{neg}=\{neg_1, neg_2\}, Res_{neg}=\{\ \}$$
$$conflict(pl_{neg}, neg_1, neg_2)$$

Avant de présenter d'autres notions de base utilisées dans notre modèle, nous soulignons l'importance du fait que nous ne voulons pas concevoir un agent capable de planifier, agir, poursuivre des buts, etc. Nous considérons qu'un tel agent existe et qu'il est ainsi capable d'avoir une représentation interne des plans (que ce soit en utilisant TAEMS ou autre), d'identifier les éléments d'un plan et les relations entre ces éléments, etc. Ce que nous proposons dans ce travail est l'existence à l'intérieur d'un agent d'un module différent qui sera responsable de calculer l'autonomie de l'agent dans différentes situations. Pour ceci, ce module – le modèle proposé dans cette partie – n'a besoin que d'une représentation superficielle des tâches et de leurs caractéristiques : nous sommes conscients que la représentation des plans présentée ci-dessus ne suffit pas pour un processus de planification, mais elle nous suffit pour identifier des contraintes qui limitent l'autonomie d'un agent, comme nous le verrons dans le Chapitre 7.

(4) Etat des tâches

Pour raisonner sur l'existence des différents types de contraintes qui lui sont imposées, un agent ne doit pas seulement être capable de se représenter quelles sont les tâches existantes, mais aussi quel est leur état à un moment donné. Par exemple, dans notre approche il est important pour un agent de

représenter non seulement qu'il connaît un plan pour satisfaire un but, mais aussi si ce plan a été exécuté avec succès ou non ou si le but a été satisfait ou non.

Pour représenter l'état des différentes tâches existantes, nous utilisons quatre prédicats correspondant aux quatre types de tâche, ressource, action, but et plan. Ces prédicats décrivent ainsi l'utilisation d'une ressource, l'exécution d'une action, la satisfaction d'un but ou l'exécution d'un plan :

$$\begin{aligned} &used(r_i), \text{ où } r_i \in \text{Res}, \\ &done(a_i), \text{ où } a_i \in \text{Act}, \\ &achieved(g_i), \text{ où } g_i \in \text{G}, \\ &executed(pl_i), \text{ où } pl_i \in \text{Pl} \\ &done(\lambda_i), \text{ où } \lambda_i \in T \end{aligned} \qquad [5.7]$$

En général, un agent a besoin de représenter l'état d'une tâche en précisant également les caractéristiques pertinentes de celle-ci, comme par exemple le moment où un but a été satisfait. Parmi les différentes caractéristiques des tâches que nous utilisons dans ce modèle, comme la durée d'une action ou la date limite de la satisfaction d'un but, certaines caractéristiques peuvent être importantes dans le raisonnement d'un agent, tandis que d'autres peuvent être ignorées. Ces caractéristiques diffèrent en fonction de la tâche analysée mais surtout en fonction du cas d'application envisagé. Il nous est donc impossible de définir dans un modèle générique quelles seront les caractéristiques des actions, ressources, buts et plans que nous représentons à chaque fois. L'exemple ci-dessous présente la représentation de l'état d'une action *pay* :

Exemple 5.2

$$done(< pay, x, y, 100, t \leq 5 >)$$

Ce prédicat décrit le fait qu'une action de paiement a été effectuée : l'agent *x* a payé à l'agent *y* une somme de *100* unités monétaires. La date limite de cette action est le moment *5* (unités temporelles) – l'action est considérée ayant été effectuée avec succès seulement si elle a été effectuée avant ce moment.

L'exemple ci-dessus illustre le besoin de représenter des conditions temporelles, spécialement pour décrire la date limite des tâches. Nous sommes conscients que pour avoir une représentation générique des conditions temporelles, nous devrons utiliser un formalisme tel qu'une logique temporelle. Cependant, dans notre modèle, une représentation simple nous suffit : nous avons principalement besoin de décrire une date limite. Nous utilisons ainsi la notation $t<dl$ – la tâche doit être exécutée avant que le moment *dl* (date limite) soit atteint – ou bien $t_i < t < t_f$ – pour représenter une condition valide entre les deux instants t_i et t_f.

(5) Evaluation de l'état des tâches

Dans notre modèle, un agent ne doit pas seulement représenter les tâches existantes et leur état, il doit être capable d'associer une valeur de vérité à cet état. Autrement dit, nous représentons l'exécution d'une tâche à l'aide d'un prédicat, mais un agent doit être capable d'évaluer si ce prédicat est vrai ou faux, c.-à-d., si la tâche a été exécutée (avant son terme limite) ou non. Des problèmes d'évaluation

d'un tel prédicat par un agent proviennent des connaissances limitées que l'agent a sur son monde environnant – l'agent doit être omniscient pour savoir précisément si une action quelconque a été effectuée ou non par un autre agent dans le passé.

Une autre source de problèmes d'évaluation de l'état d'une tâche par un agent concerne les conditions temporelles présentes. Même si nous supposons qu'un agent est omniscient, il est impossible pour l'agent de dire si une tâche a été exécutée avant sa date limite si la tâche est en cours d'exécution et que la date limite n'est pas encore atteinte. Nous considérons qu'une distinction doit être faite entre une tâche qui n'a pas été exécutée du tout et une tâche qui n'a pas encore été exécutée mais dont la date limite n'a pas encore été atteinte.

Ceci revient au fait que, dans notre modèle, l'évaluation par un agent d'un prédicat (ou une formule logique en général) résulte en des valeurs de vérité *vrai* ou *faux*, mais aussi *inconnu.* De plus, le résultat de cette évaluation dépend de l'agent qui l'effectue et de l'instant où elle est faite. Nous notons ainsi par *eval* la fonction qui décrit le résultat d'évaluation d'une formule logique par un agent *x* à un moment de temps *t* :

$$eval(x,t,\alpha) \in \{\text{true}, \text{false}, \text{unk}\}, \text{où } \alpha \text{ est une formule logique} \qquad [5.8]$$

Les formules logiques qu'un agent peut évaluer dans notre modèle concernent principalement des conditions temporelles et des formules bien formées de la logique des prédicats que nous utilisons. Nous notons ainsi par *EF* (Evaluable Formulae), l'ensemble des formules qui peuvent être évaluées par un agent – toutes les formules présentées dans ce manuscrit appartiennent à cet ensemble.

Comme notre objectif est de proposer un modèle qui sera utilisé par un agent, dans la suite de ce manuscrit nous n'utiliserons plus le paramètre *x* dans la fonction d'évaluation des formules : l'agent qui utilise ce modèle est celui qui évalue les formules. Ce modèle est conçu pour être utilisé constamment par l'agent : le monde extérieur change sans cesse, donc les contraintes représentées à l'aide de ce modèle aussi. Nous n'utiliserons donc plus le paramètre *t* dans la fonction d'évaluation des formules et nous sous-entendons que l'évaluation décrite est celle faite au moment d'utilisation. De plus, cette fonction d'évaluation apparaîtra par la suite seulement dans les cas où il est important de souligner le fait que l'agent évalue une formule logique – nous éviterons ainsi d'écrire *eval(α)* pour chaque formule logique décrite dans ce manuscrit.

Synthèse

Dans les sections précédentes nous avons décrit plusieurs notions de base qui constituent les briques de notre modèle. Il s'agit notamment des concepts et notations qui se retrouveront dans la plupart des formules que nous allons introduire par la suite : leur sens et les raisons de leur utilisation deviendront ainsi plus claires dans les sections suivantes.

5.2. Interactions entre agents

Dans la littérature (cf. Chapitre 3), les engagements sociaux sont souvent utilisés pour donner une sémantique à la communication : en envoyant un message, un agent s'engage envers le destinataire du message sur la validité du contenu du message. Autrement dit, le processus d'interaction et sa

sémantique sont décrits à travers ce paradigme. Nous ne nous intéressons pas ici à cette utilisation des engagements sociaux, mais à leur utilisation pour représenter les contraintes découlant de l'interaction entre agents. A la suite de l'interaction avec un autre agent, un agent peut devenir engagé envers l'autre pour l'objet de l'interaction. L'agent sera donc contraint par cet engagement social. Dans cette section nous décrivons le modèle utilisé dans notre travail pour les engagements sociaux, quels sont les différents états possibles d'un engagement et comment ils sont gérés.

5.2.1. Modèle d'engagement social

Comme nous l'avons vu dans le Chapitre 3, plusieurs définitions ont été proposées dans la littérature pour le concept d'engagement social. Les différences entre ces définitions sont en général issues de l'utilisation envisagée : les engagements utilisés dans la communication entre agents ont des propriétés parfois différentes de ceux utilisés pour décrire les résultats d'interactions. Dans ce manuscrit nous prenons des définitions existantes les éléments qui nous semblent les plus pertinents pour décrire des contraintes interpersonnelles comportementales, c.-à-d., les résultats d'interactions entre agents.

(1) Définition d'un engagement social

Comme la plupart des définitions existantes (voir Chapitre 3), nous définissons un engagement social comme une collection de plusieurs caractéristiques essentielles. Un engagement social est défini par les deux agents débiteur et créditeur concernés – *debtor* et *creditor*, l'agent témoin de l'engagement – *agent institutionnel –ia-*, l'*objet* –object- et la *condition de validité –condition-* de cet engagement. La signification de ce concept est que l'agent débiteur est engagé envers l'agent créditeur devant l'agent institutionnel à réaliser l'objet de l'engagement tant que la condition de validité reste vraie.

$$sc=<debtor,creditor,ia,object,condition> \quad [5.9]$$

Au delà des deux agents directement concernés par l'engagement, le débiteur et le créditeur, notre définition inclue le témoin de l'engagement, témoin que nous appelons *agent institutionnel* (ia). Les caractéristiques de ce témoin, ainsi qu'une explication pour le nom que nous avons choisi seront présentées à la fin de cette section.

L'objet d'un engagement est une formule logique : le débiteur est donc engagé de s'assurer de la véracité de cette formule. Dans notre modèle, les engagements sociaux concernent souvent des tâches à effectuer, où une tâche peut être une action, une ressource, un but ou un plan (voir la section précédente). Dans ces cas, l'objet de l'engagement est un des prédicats définis dans la formule 5.7, c.-à-d., l'agent est engagé à arriver dans un état du monde dans lequel le prédicat est vrai, donc la tâche a été exécutée avec succès. Nous verrons plus tard d'autres expressions logiques qui peuvent être utilisés comme objets des engagements.

La condition de validité d'un engagement social dans notre modèle est une autre formule logique : si celle-ci est vraie l'engagement est considéré *valide*, si elle est fausse l'engagement est considéré *invalide*. Autrement dit, l'agent débiteur reste engagé envers l'agent créditeur tant que cette formule est vraie. La signification de cette validité d'un engagement porte sur le comportement de l'agent débiteur : tant que son engagement est valide, son comportement est censé être orienté vers la

satisfaction de l'objet (p.ex. d'effectuer une action), tandis que si son engagement n'est plus valide il ne doit plus satisfaire l'objet de l'engagement. Généralement, cette condition est temporelle et elle prend la forme d'une date limite comme celles que nous avons introduites comme caractéristique des tâches dans la section précédente. Ce modèle d'engagements assez simple nous permet par exemple d'exprimer le fait que la date limite d'un engagement (condition de validité) est différente de la date limite de son objet (caractéristique de la tâche – objet de l'engagement). Dans le chapitre suivant nous verrons des exemples d'autres conditions de validité non-temporelles.

Un engagement social dans notre modèle est défini donc par les caractéristiques présentées ci-dessus. Nous notons par SC l'ensemble de tous les engagements sociaux possibles. Nous notons aussi les engagements sociaux par *sc*, avec des indices permettant de les différencier si nécessaire :

$$\mathrm{SC} = \left\{ sc =< x, y, z, \alpha, \beta > \middle| \forall x, y, z \in Ag, \forall \alpha, \beta \in EF \right\} \qquad [5.10]$$

(2) Etat d'un engagement social

En fonction de la véracité de sa condition de validité, un engagement social peut être valide ou invalide, ce qui représentent deux des états dans lesquels il peut se trouver. Généralement un agent ne s'intéresse qu'à ses engagements valides, ceux qui sont invalides ne représentant pas des contraintes sur son comportement, même s'ils peuvent (re)devenir valides à un moment dans le futur. Si un agent est engagé envers un autre à assurer que la formule logique qui constitue l'objet de l'engagement soit vraie (par exemple en effectuant une tâche), alors il existe généralement deux possibilités. Soit il le fait avant que la condition de validité de l'engagement soit expirée, soit il ne le fait pas. Dans le premier cas l'engagement est considéré être dans l'état *accompli (fulfilled),* tandis que dans le deuxième cas il est *violé (violated)*. Les formules ci-dessous présentent la définition des ensembles d'engagements valides, accomplis et violés respectivement (*ActSC, FulSC, VioSC*) :

$$\begin{aligned} \mathrm{ActSC} &= \left\{ sc =< x, y, z, \alpha, \beta > \in SC \;\middle|\; eval(\beta) = true \right\} \\ \mathrm{FulSC} &= \left\{ sc =< x, y, z, \alpha, \beta > \in ActSC \;\middle|\; eval(\alpha) = true \right\} \\ \mathrm{VioSC} &= \left\{ sc =< x, y, z, \alpha, \beta > \in ActSC \;\middle|\; eval(\alpha) = false \right\} \end{aligned} \qquad [5.11]$$

Nous voulons souligner qu'un engagement social accompli ou violé doit être d'abord valide : si la condition de validité d'un engagement a expiré, il n'y a plus aucun sens de parler de la satisfaction de l'engagement. Cependant, les engagements valides sont soit accomplis, soit violés. Du fait que la fonction d'évaluation d'un agent peut aussi avoir la valeur *unknown*, il existe des engagements sociaux valides qui ne sont ni accomplis ni violés. A ce moment, vu qu'il n'y a pas d'évaluation possible de l'état de l'objet de l'engagement, nous ne pouvons rien dire sur la satisfaction de cet engagement.

(3) Tableaux d'engagements et agents institutionnels

Même si nous pouvons définir l'ensemble de tous les engagements sociaux possibles ou valides, ces ensembles ne sont pas très utiles en pratique. Un agent est plutôt intéressé par les engagements qu'il a envers d'autres et éventuellement par les engagements d'autres agents. En général, les engagements qui intéressent les agents sont ceux qui existent, pas ceux qui sont possibles. Comme les travaux similaires

(cf. Chapitre 3), nous utilisons le concept de *tableau d'engagements* (*commitment store* en anglais), tableau qui regroupe les engagements existants dans un système. Dans notre modèle, ce tableau d'engagements est géré par un agent institutionnel et il contient ainsi tous les engagements créés dont l'agent institutionnel est le témoin. Le rôle de l'agent institutionnel est donc de gérer un tableau d'engagements.

Comme discuté dans le Chapitre 3, le rôle de cet agent témoin est très important, dans la mesure où il transforme un simple accord entre deux agents dans un engagement social : cet agent connaît les engagements créés et ces engagements sont utilisables dans le raisonnement social d'un agent grâce au tableau d'engagements qu'il gère. Nous ne voulons pas proposer une approche pour concevoir un système à base d'engagements sociaux, cependant nous pouvons préciser ici quelques caractéristiques d'un tel système. Un ou plusieurs agents institutionnels peuvent être présents dans le système, chacun gérant un tableau d'engagements. Une synchronisation doit être faite entre ces agents institutionnels afin de préserver la cohérence des engagements. De plus, les agents institutionnels doivent être omniscients (ou au moins informés par les autres agents) afin de pouvoir ajouter ou supprimer des engagements sociaux ou de modifier leur état au fur et à mesure que le temps passe et que des tâches sont effectuées. Comme nous le verrons dans le chapitre suivant, nous pouvons aussi considérer qu'une responsabilité des agents institutionnels est de prendre les mesures nécessaires une fois qu'un engagement social devient accompli ou violé – ceci est fait dans le cadre des institutions électroniques, d'où le nom pour ces agents.

Dans notre travail, un agent utilise le modèle que nous proposons dans son raisonnement interne. Dans ce modèle, l'agent représente le fait qu'il existe des engagements sociaux et qu'ils sont stockés dans un tableau d'engagements géré par un agent institutionnel. L'existence d'un tel agent dans le système et la manière dont il arrive à résoudre les problèmes mentionnés ci-dessus, ne concernent pas l'agent que nous analysons. De son point de vue, chaque fois qu'il crée un engagement social, celui-ci apparaît dans le tableau d'engagement et chaque fois que l'état de l'engagement change ou que l'engagement est annulé, ce changement est également présent dans le tableau.

Nous voyons apparaître ici deux opérations sur les engagements sociaux, la création et l'annulation, opérations que nous analysons par la suite. Mais avant ceci, nous voulons préciser que dans ce manuscrit nous notons les agents institutionnels par *ia* (avec des indices si nécessaire), l'ensemble de tous les agents institutionnels existants par IA ($IA \subseteq Ag$) et par SC_{ia} le tableau d'engagements géré par l'agent institutionnel *ia*, donc les engagements dont il est le témoin :

$$IA = \{ia \in Ag \mid \exists sc = < x, y, ia, \alpha, \beta > \in SC\} \qquad [5.12]$$
$$SC_{ia} \subseteq SC, \forall sc = < x, y, z, \alpha, \beta > \in SC_{ia}, z = ia$$

(4) Création et annulation d'un engagement social

Dans les formules 5.12 ci-dessus nous n'avons pas donné une définition complète du tableau d'engagements SC_{ia} géré par l'agent institutionnel *ia*. Ce tableau d'engagements contient les engagements sociaux créés avec l'agent institutionnel comme témoin. Chaque fois qu'un tel engagement est crée, il est ajouté au tableau et chaque fois qu'il est annulé, il en est retiré. La création et l'annulation d'un engagement social représentent des résultats des interactions entre des agents. Par

exemple, une négociation dans un marché électronique peut se conclure en la création d'un engagement social d'un agent envers un autre et souvent en la création d'un engagement dans l'autre sens (un agent qui s'engage à fournir un service et l'autre qui s'engage à payer pour ce service). De même, deux agents peuvent négocier ou tout simplement communiquer pour rompre un accord existant, c.-à-d., annuler un engagement social.

Comme notre modèle a comme objectif de représenter des contraintes décisionnelles ou comportementales, il ne représente pas ces interactions entre agents. Les opérations de création et d'annulation d'un engagement social ne sont donc pas présentes dans notre modèle, mais seulement leur résultat : la présence ou absence de l'engagement du tableau d'engagements. Dans la suite de ce manuscrit nous considérons que les engagements sociaux qui existent (qui ont été créés), tout en sachant que l'ensemble de ces engagements varie constamment suite aux interactions entre agents. Cette approche peut être reformulée en disant qu'un agent ne raisonne pas sur un engagement appartenant à l'ensemble de tous les engagements possibles (*SC*), mais seulement un engagement appartenant à un des tableaux d'engagements existants (SC_{ia}).

Exemple 5.3

La formule ci-dessous décrit un engagement social de l'agent *x* envers l'agent *y* pour effectuer l'action de *payer à l'agent z la somme de 100 avant le moment 10*. L'engagement est valide jusqu'au moment *20* et, comme il a été créé, il se retrouve dans le tableau d'engagements géré par l'agent *ia* qui est le témoin de l'engagement :

$$< x, y, ia, done(pay, x, z, 100, t \leq 10), t \leq 20 > \in SC_{ia}$$

5.2.2. Interactions entre agents et politiques sociales

Les engagements sociaux représentent des contraintes sur le comportement des agents, contraintes issues des interactions entre les agents et sur lesquelles nous voulons que les agents raisonnent. Même si nous ne nous intéressons pas dans ce manuscrit directement à la manière dont les agents interagissent (échange des messages, syntaxe ou sémantique de ces messages, etc.), nous avons besoin de représenter le fait qu'un engagement social est créé suite à une interaction.

La forme d'interaction entre deux agents qui nous intéresse dans ce travail et qui est reliée à la notion d'autonomie est représentée par la délégation/adoption des buts (actions, plans, etc.), comme nous l'avons vu dans le Chapitre 1. Dans le cadre de ce type d'interaction, un agent demande (en utilisant un performatif de type FIPA, un jeu de dialogue, etc.) à un autre de satisfaire un but (exécuter une action, etc.) à sa place. Il essaie donc de déléguer à un autre agent un de ses buts. L'autre agent peut accepter ou refuser l'adoption de ce but. La signification des engagements sociaux est que si un agent accepte d'adopter un but d'un autre, il s'engage envers l'autre pour la satisfaction du but. L'adoption ne représente donc rien de plus que d'effectuer l'opération de création d'un engagement social, engagement social qui se retrouvera ainsi dans un tableau d'engagements. Ainsi, dans notre modèle, nous considérons qu'un agent a adopté l'exécution d'une tâche pour un autre agent si un engagement social correspondant existe dans le tableau d'engagements géré par l'agent institutionnel.

En ce qui concerne la délégation de l'exécution d'une tâche (satisfaction d'un but, exécution d'une action ou d'un plan, etc.), elle a besoin d'une représentation à part. Nous représentons l'interaction (négociation, etc.) entre agents ayant comme objectif la délégation de l'exécution d'une tâche d'un agent à un autre à l'aide du prédicat *request*. Ce prédicat a comme paramètres l'agent demandeur (celui qui délègue), l'agent receveur (celui qui adoptera ou non) et l'expression logique concernant l'exécution de la tâche déléguée, avec ses caractéristiques potentielles :

$$request(x, y, done(\lambda)) \qquad [5.13]$$

Exemple 5.4

La formule ci-dessous décrit la délégation de l'exécution de l'action *exec_job* avec la date limite *10*, délégation faite par l'agent *x* à l'agent *y* :

$$request(x, y, done(< exec\_job, x, t \leq 10 >)\ )$$

Si l'agent y décide d'accepter la délégation, il a adopté l'exécution de l'action et il s'est donc engagé envers *x* pour cette exécution :

$$\exists sc =< x, y, ia,\ done(< exec\_job, x, t \leq 10 >),\ t \leq 10 > \in SC_{ia}$$

Politiques sociales

Comme nous l'avons discuté dans le Chapitre 1, l'autonomie d'agents fait que le résultat de cette délégation est incertain. Il est possible que suite à cette demande un engagement social soit créé, c.-à-d., une adoption soit effectuée, mais ce n'est pas toujours le cas. Des règles doivent être donc mises en place dans un système pour définir dans quelles conditions un agent est obligé d'accepter une délégation, c.-à-d., quand un prédicat *request* implique la création d'un engagement social. Dans la littérature (voir Chapitre 3), ces règles sont appelées des *politiques sociales*, qui représentent des normes interpersonnelles (voir Chapitre 2). Ces politiques sociales sont représentées comme des méta-engagements, c.-à-d., des engagements sociaux qui ont comme objet une expression logique contentant des engagements sociaux :

$$SPol = \left\{ < x, y, ia, \alpha, \beta > \in SC \;\middle|\; \begin{array}{l} \forall x, y \in Ag, \forall ia \in IA, \forall \alpha, \beta \in EF, \\ \alpha = \forall z \in Ag\ \ \forall \lambda \in T : request(z, x, done(\lambda, \beta)) \Rightarrow \\ \qquad\qquad \exists sc =< x, z, ia, done(\lambda), \beta > \in SC_{ia} \end{array} \right\} \qquad [5.14]$$

L'ensemble d'engagements sociaux décrit ci-dessus représente toutes les politiques sociales qui peuvent exister concernant les délégations d'exécution des tâches, les *request*. Une telle politique sociale représente ni plus ni moins que l'engagement social d'un agent *x* envers un agent *y* devant un agent institutionnel *ia* d'adopter toute exécution de tâche déléguée par un agent *z*, c.-à-d., si *z* demande à *x* d'effectuer une tâche pour lui, alors *x* s'engage à l'effectuer. De point de vue formel, une politique sociale est un engagement social : elle est créée devant un agent institutionnel et elle est stockée dans un tableau d'engagements avec les autres engagements. Une politique sociale peut être donc accomplie ou violée et elle a des conditions de validité, comme tout engagement social.

L'utilisation des politiques sociales nous permet de faire la différence entre la violation d'une politique et la violation d'un engagement. L'engagement de *x* vers *z* est violé si *x* n'effectue pas la tâche-objet de l'engagement, par exemple parce qu'il ne veut pas ou qu'il essaie mais n'y arrive pas. Mais, en tout cas, *x* avait précisé sa volonté d'effectuer la tâche-objet en créant l'engagement. La politique sociale est violée si x *refuse* la création de l'engagement quand *z* le demande, c.-à-d., *x* communique qu'il ne veut pas effectuer la tâche-objet. La politique sociale violée est de *x* envers *y* (et non envers *z*) et donc le comportement autonome de *x* peut être traité différemment s'il viole une politique sociale ou un engagement social créé suite à cette politique.

Ces règles de création d'engagements que nous exprimons sous forme de politiques sociales sont des tentatives pour restreindre le comportement autonome des agents, en leur précisant la réponse permise dans une interaction et quand il faut s'engager à effectuer une tâche. Ceci coïncide avec le but d'utilisation des rôles et structures organisationnelles, qui essaient de décrire le comportement désiré des agents, comme nous le verrons dans le chapitre suivant.

5.3. Synthèse

Dans ce chapitre nous avons défini le cadre de base sur lequel nous allons nous appuyer dans la suite de ce document, notamment les notions qui caractérisent le comportement des agents : les actions, les ressources, les buts et les plans. En nous appuyant sur un modèle de représentation des plans, nous avons identifié les caractéristiques d'un plan qui seront utiles dans notre modèle, utilité qui sera décrite dans le Chapitre 7. Dans ce chapitre nous nous sommes également intéressés à la modélisation des interactions entre agents. Comme nous l'avons précisé auparavant, nous avons choisi le paradigme d'engagements sociaux pour décrire le résultat des interactions entre agents. Nous avons ainsi proposé un modèle d'engagements sociaux inspiré des modèles existants dans la littérature (voir Chapitre 3) et nous avons décrit comment fonctionne un système à base d'engagements sociaux. Un tableau d'engagement est géré par un agent institutionnel qui vérifie le comportement d'agents, modifie l'état de leurs engagements et prend les mesures nécessaires. Nous voulons souligner que, même si le modèle d'engagements sociaux proposé peut être utilisé pour concevoir un système multi-agents dans lequel les interactions se font via des engagements sociaux, ceci ne constitue pas notre objectif. Dans notre contexte, un agent utilise ce modèle pour représenter les interactions, même si en réalité l'approche utilisée est différente.

Les engagements sociaux peuvent ainsi représenter des contraintes comportementales interpersonnelles, c.-à-d., des contraintes issues des interactions entre les agents. Nous avons choisi de modéliser le type d'interactions lié à la notion d'autonomie, la délégation/adoption des buts entre agents, comme des opérations sur des engagements sociaux, ce qui nous permet ainsi de traiter les relations interpersonnelles entre les agents à travers un paradigme unique : les engagements sociaux. Si jusqu'à maintenant nous nous sommes intéressés à définir les briques de base de notre modèle, notamment les notions qui caractérisent le comportement individuel d'un agent et les interactions entre agents, par la suite nous décrirons l'aspect institutionnel de notre modèle : comment nous modélisons les organisations et leur influence sur la prise de décision et le comportement des agents.

6. Engagements sociaux institutionnels

Après avoir décrit dans le chapitre précédent les notions de base de notre approche et le modèle d'engagements sociaux que nous utilisons pour décrire les interactions entre agents, dans ce chapitre nous nous intéressons à la modélisation des aspects institutionnels. Nous commençons par rappeler les notions le plus souvent utilisées dans le cadre de la coordination institutionnelle, notions présentées dans le Chapitre 2, telles que les rôles ou les normes. Pour modéliser les contraintes institutionnelles imposées aux agents appartenant à une organisation par le fait de jouer des rôles dans celle-ci, nous utilisons dans ce manuscrit le paradigme des engagements sociaux. Nous justifions ce choix et nous décrivons comment les engagements et les méta-engagements (politiques sociales) peuvent décrire le comportement attendu d'un agent qui joue un rôle et les normes qui lui sont imposées.

Vu que dans ce travail nous nous intéressons particulièrement au raisonnement effectué par un agent pour décider de jouer ou non un rôle, nous décrivons dans ce chapitre comment nous modélisons à l'aide d'engagements sociaux le fait qu'un agent joue un rôle et comment des contraintes institutionnelles sont issues de ce fait. Nous intéressant à des agents autonomes qui sont capables de désobéir à ces contraintes (et subir les conséquences qui en découlent), nous analysons ensuite comment une organisation à base d'engagements sociaux telle que nous proposons, gère la violation des engagements et impose des sanctions. Nous ne proposons pas la création d'une telle organisation, mais tout simplement le fait qu'un agent représente une organisation de cette manière. Nous finissons ce chapitre en soulignant l'originalité et l'importance pour notre travail de la modélisation à base d'engagements sociaux d'une organisation d'agents.

6.1. Contraintes institutionnelles

Dans le Chapitre 2 de ce manuscrit nous avons analysé les notions utilisées pour définir les mécanismes et structures de coordination institutionnelle. La notion de *rôle* est centrale dans le cadre de cette coordination : les rôles sont utilisés pour décrire le comportement attendu des agents appartenant à une organisation. D'une manière générale, une organisation fonctionne en définissant des rôles et en spécifiant le comportement que tout agent jouant un rôle doit manifester : si les agents obéissent les spécifications de leurs rôles, les buts globaux de l'organisation seront satisfaits. Une organisation divise ainsi ses buts globaux dans des sous-buts qu'elle associe aux rôles : tout agent qui joue un rôle doit satisfaire ces buts que nous appelons les *buts du rôle*.

Dans le cas général, une organisation essaie d'aider ses membres à satisfaire les buts de leurs rôles, et cette aide peut prendre plusieurs formes. Souvent, des structures organisationnelles telles que les hiérarchies de rôles sont utilisées : ces hiérarchies définissent des *liens d'autorité* entre les rôles. Un agent jouant un rôle peut ainsi déléguer une partie de ses buts (ou plus généralement de ses tâches) à un autre agent sur lequel il a de l'autorité parce qu'un lien d'autorité existe entre les rôles qu'ils jouent. Le deuxième agent ne peut qu'obéir et adopter la tâche déléguée par le premier. Dans le cas contraire il n'obéit pas aux spécifications de son rôle. Une autre possibilité d'aider les agents à satisfaire les buts de leurs rôles est de mettre à leur disposition des *ressources* qu'ils pourront utiliser dans le cadre de la satisfaction de ces buts. Cette mise à disposition de ressources peut prendre la forme d'un don de

ressource (p.ex. : de l'argent) ou d'une *permission* donnée aux agents. Dans le dernier cas une organisation peut spécifier les tâches (ressources, actions, buts, plans) que les agents jouant des rôles ont le droit d'effectuer. Comme nous le verrons dans le chapitre suivant, ces permissions ont une grande importance dans le calcul des pouvoirs des agents. Une autre possibilité d'aider les agents à satisfaire les buts de leurs rôles est de leur apprendre comment effectuer de nouvelles tâches, notamment des actions ou des plans. Les agents pourront ainsi utiliser les connaissances acquises dans la satisfaction des buts des rôles, mais aussi dans la satisfaction de leurs propres buts.

La permission est un cas particulier de norme, normes qui peuvent également prendre la forme d'interdictions ou d'obligations. Les interdictions sont souvent considérées complémentaires aux permissions : certains modèles spécifient explicitement les permissions associées à un rôle et considèrent que tout ce qui n'est pas explicitement permis est interdit, tandis que d'autres font l'inverse. Les *obligations* associées aux rôles ont comme objectif de limiter le comportement des agents jouant les rôles afin d'assurer une coordination des activités locales. Comme nous avons pu le voir dans le Chapitre 2, ces normes sont souvent dépendantes du contexte : quand un agent joue un rôle, il adopte les normes associées à son rôle, normes qu'il doit activer dans des contextes spécifiques et qui spécifient des tâches qu'il doit obligatoirement effectuer. Un cas particulier de ce contexte d'activation de norme est « quand l'organisation le demande » : souvent les organisations utilisent des normes qui spécifient que, quand elles le jugent nécessaire, un agent sera obligé d'effectuer une tâche. Ceci correspond à une mise à la disposition d'une organisation d'un pouvoir d'un agent (cf. Chapitre 3). Comme nous le verrons dans le chapitre suivant, ceci a une grande influence sur l'évolution de l'autonomie des agents.

Pour résumer, nous pouvons considérer que les spécifications institutionnelles associées à un agent par le fait qu'il joue un rôle dans une organisation peuvent être divisées en plusieurs catégories :

- les ressources qu'il reçoit ;
- les actions et plans qu'il apprend à exécuter ;
- les permissions d'effectuer des taches, telles que la permission d'exécuter une action ou d'utiliser un plan ;
- les buts qu'il doit satisfaire – les buts de son rôle ;
- les liens d'autorité qui le lient à d'autres agents jouant des rôles – ces liens peuvent être orientés dans les deux sens : il peut avoir de l'autorité sur des agents et d'autres peuvent avoir de l'autorité sur lui ;
- les obligations qu'il doit remplir, obligations qui souvent sont dépendantes du contexte.

Par la suite nous allons proposer une représentation à base d'engagements sociaux (ou de méta-engagements – politiques sociales) des trois derniers types de spécifications institutionnelles, tandis que les trois premiers types seront analysés dans la section 6.3.2 et puis dans 7.2.2. Mais avant ceci, nous voulons présenter quelques notations utilisées par la suite.

Une organisation représente un concept abstrait et nous avons besoin de représenter les relations entre un agent et une organisation. Nous utilisons ainsi le terme d'*agent organisationnel* (*oa*) qui est l'agent qui représente l'organisation dans les interactions avec d'autres agents et qui appartient à l'ensemble de tous les agents organisationnels existants (OA). Dans les travaux similaires sur les institutions

électroniques, souvent l'agent organisationnel (qui représente l'organisation) et l'agent institutionnel (qui gère le résultat des interactions entre agents) sont les mêmes (*oa=ia*), mais ceci est un choix de réalisation d'un système et nous préférons garder une représentation plus générique en faisant la distinction entre eux.

Nous notons par *R* majuscule des rôles (en utilisant parfois des indices) et par $Roles_{oa}$ l'ensemble de tous les rôles dans l'organisation représentée par l'agent *oa*. Nous notons le fait qu'un agent *x* joue un rôle *R* par *x:R*. Nous donnons une signification plus précise de cette notation plus tard (section 6.3) :

$$x : R, \text{où } x \in Ag,\ R \in Roles_{oa}, oa \in OA \subseteq Ag \qquad [6.1]$$

6.2. Représentation à base d'engagements et politiques sociales

Comme nous l'avons précisé dans le Chapitre 2, certains auteurs, p.ex. (Boella and van der Torre 2004a), considèrent que les spécifications d'un rôle peuvent être représentées à l'aide de la notion de *contrat*. Vu que la notion de contrat n'est qu'un cas particulier d'engagement social dans un contexte institutionnel, nous utiliserons le paradigme d'engagements sociaux pour représenter les contraintes institutionnelles issues du fait de jouer des rôles. Nous arriverons ainsi à une représentation plus uniforme (qui utilise les mêmes concepts) des contraintes interpersonnelles et institutionnelles.

(1) Les buts d'un rôle

Quand il joue un rôle dans une organisation, un agent doit satisfaire des buts pour l'organisation. Dans notre vision, ceci est équivalent au fait qu'un *agent est engagé envers l'organisation pour satisfaire ces buts*. Nous représentons avec un engagement social le fait qu'un agent jouant un rôle doit satisfaire un but de l'organisation. L'*engagement pour un but d'un rôle* est donc fait entre un agent qui joue le rôle, envers un agent organisationnel, devant un agent institutionnel et pour le but que l'agent doit satisfaire. Nous voulons souligner le fait que, vu le modèle d'engagement social que nous utilisons, nous pouvons représenter l'engagement d'un agent envers l'organisation pour tout type de tâche (action, etc.) qu'un agent doit effectuer, même si dans la littérature seulement les buts des rôles sont généralement pris en compte. La condition de validité de cet engagement est représentée par le fait que l'agent joue le rôle : s'il ne joue plus ce rôle, il n'est plus engagé à satisfaire le but. Bien évidemment, le but peut avoir un terme limite pour être satisfait, mais c'est le terme limite de la tâche-objet de l'engagement, et non la condition de validité de l'engagement :

$$\begin{aligned} &rolecomm = < x, oa, ia, \alpha, x : R >, \\ &\text{où } x \in Ag,\ R \in Roles_{oa}, oa \in OA, ia \in IA, \alpha \in EF \\ &\text{en général } \alpha = done(\lambda), \lambda \in T \text{ or } \alpha = achieved(g), g \in G \end{aligned} \qquad [6.2]$$

Exemple 6.1

Prenons par exemple le cas d'un agent *ca* qui joue le rôle *contractor* (que nous notons *RContr*) dans une organisation qui représente une place de marché. Cette organisation est représentée par un agent *ma* (agent marché), agent qui représente également l'agent institutionnel des engagements sociaux existants dans cette organisation. A travers des exemples suivants nous détaillerons plus

cette organisation. Les spécifications du rôle *RContr* sont de négocier avec d'autres agents pour former des contrats d'exécution d'un job. Ces contrats sont naturellement représentés comme des engagements sociaux entre deux agents pour l'action *exec_job* :

$$sc_1 = < x, ca, ma, done(exec\_job, x, t \le 10),\ 5 \le t \le 15 >$$

Dans cette organisation, le rôle *RContr* a le but associé de négocier, donc de satisfaire la tâche *neg* (voir l'Exemple 5.1). Un agent *ca* qui joue ce rôle est ainsi engagé envers l'agent représentant l'organisation pour satisfaire ce but :

$$rc_1 = < ca, ma, ma, achieved(neg), ca : RContr >$$

(2) Les obligations d'un rôle

Il est souvent impossible de spécifier dès la phase de conception tous les buts attribués à un rôle. Un degré de flexibilité est nécessaire. Les organisations utilisent des obligations dépendantes du contexte pour spécifier le comportement désiré d'un agent jouant un rôle dans un certain contexte. L'agent n'est plus engagé à satisfaire un but précis quand il joue le rôle, mais un mécanisme est nécessaire pour préciser qu'il doit s'engager dans un contexte. Nous représentons ce nouveau type de limite imposée à un agent à l'aide d'une *politique sociale pour une obligation d'un rôle*. L'agent est donc engagé envers l'organisation, devant un agent institutionnel, de satisfaire une formule logique – l'objet de la politique sociale. Cet objet prend généralement la forme d'une implication : si une formule logique est vraie (le contexte), alors l'agent doit créer un engagement social un agent jouant le rôle *R* est engagé de créer un engagement social pour effectuer une tâche quelconque.

Dans notre modèle, le contexte d'une obligation devient actif quand l'organisation le précise explicitement. Autrement dit, la norme spécifie que, quand l'organisation lui délègue une tâche, l'agent doit s'engager pour l'effectuer. Dans le chapitre précédent nous avons représenté la délégation des tâches à l'aide d'un prédicat *request* qui peut résulter dans la création d'un engagement social. Nous appelons l'ensemble des obligations possibles *ROPol* (*Role Obligation Policies*) et il est constitué des obligations représentées sous la forme des politiques sociales :

$$ROPol_{oa} = \left\{ < x, oa, ia, \alpha, x : R > \in SC_{ia} \;\middle|\; \begin{array}{l} \forall x \in Ag, \forall oa \in OA, \forall ia \in IA, \forall R \in Roles_{oa}, \\ \alpha = \forall \lambda \in T : request(oa, x, done(\lambda, \beta)) \Rightarrow \\ \qquad \exists sc = < x, oa, ia, done(\lambda), \beta > \in SC_{ia} \end{array} \right\} \qquad [6.3]$$

L'ensemble défini ci-dessus contient des politiques sociales qui sont stockées dans le tableau d'engagements géré par l'agent institutionnel et qui sont valides tant que l'agent joue son rôle. Elles spécifient que chaque demande de création d'engagement que l'organisation (l'agent organisationnel) fait est suivie par la création de l'engagement par l'agent. C'est-à-dire, quand l'organisation décide qu'un agent jouant un rôle doit effectuer une tâche (généralement il s'agit de la satisfaction d'un but), l'agent doit s'engager pour l'effectuer. L'ensemble $ROPol_{oa}$ contient donc toutes les obligations dans l'organisation représentée par l'agent *oa*. Bien évidemment, dans le cas réel les normes existantes ne représentent qu'un sous-ensemble de $ROPol_{oa}$, il est très restrictif pour un agent de devoir s'engager

pour n'importe quelle tâche déléguée par l'organisation et donc les tâches pour lesquelles un agent est engagé envers l'organisation doivent être spécifiées explicitement.

Comme les politiques sociales présentées ci-dessus représentent des obligations envers une organisation et comme l'objet de ces obligations est en général la satisfaction d'un but, nous utilisons un raccourci de notation en utilisant le prédicat *obliged_to* qui signifie tout simplement qu'il existe une obligation pour un agent concernant un but, obligation associée à un rôle qu'il joue dans une organisation :

$$
\begin{aligned}
obliged\_to(x, oa, g) \equiv & \exists R \in Roles_{oa}\ x : R \wedge \exists sp \in ROPol_{oa,} \\
& sp = < x, oa, ia, \\
& \quad request(oa, x, achieved(g, \beta)) \Rightarrow \\
& \qquad \exists < x, oa, ia, achieved(g), \beta > \in SC_{ia} \\
& \quad , x : R > \\
\text{où } & x \in Ag,\ oa \in OA, g \in G, ia \in IA, \beta \in EF
\end{aligned}
\qquad [6.4]
$$

Exemple 6.1 (cont.)

Pour donner un exemple de ce type de politique sociale pour l'obligation d'un rôle, nous continuons l'exemple précédent : un agent *ca* joue le rôle *RContr* dans une organisation (place de marché) représentée par l'agent *ma*. Ce rôle a le but de négocier, ce qui résulte dans la formation d'un contrat (représenté comme un engagement social - exemple précédent). Ce rôle a aussi l'obligation de payer une taxe à l'organisation chaque fois que l'organisation lui demande. Un agent qui joue ce rôle est donc engagé d'exécuter la tâche *pay* (action de payer) dans un certain contexte – quand un *request* est fait :

$$
\begin{aligned}
& < ca, oa, ia,\ request(oa, ca, done(< pay, ca,\ 100, \beta >)) \Rightarrow \\
& \qquad \exists < ca, oa, ia, done(< pay, x, 100, \beta >), \beta > \in SC_{ia},\ \ ca : RContr >
\end{aligned}
$$

(3) Les relations d'autorité entre les rôles

Le dernier type de contraintes institutionnelles imposées aux agents jouant des rôles sont les liens d'autorité qui définissent les hiérarchies de rôles présentes dans les structures organisationnelles. Les rôles sont liés par des relations d'autorité, dans lesquelles le rôle « supérieur » a autorité sur le rôle « inférieur » pour l'exécution d'une tâche, c.-à-d., chaque fois qu'un agent jouant le premier rôle délègue une tâche à un agent jouant le deuxième, le deuxième l'adopte.

Ces relations d'autorité sont facilement représentables en utilisant les politiques sociales : un agent jouant un rôle est engagé, envers l'organisation (l'agent organisationnel), devant un agent institutionnel, pour créer un engagement d'effectuer une tâche chaque fois qu'un agent jouant un autre rôle lui demande. Cette politique sociale d'autorité entre rôles est valide tant que l'agent joue le rôle. Nous appelons l'ensemble des relations d'autorité possibles *RAPol* (*Role Authority Policies*) et il est constitué des relations d'autorité représentées sous la forme des politiques sociales :

$$RAPol_{oa} = \left\{ < x, oa, ia, \alpha, x : R_1 > \in SC_{ia} \;\middle|\; \begin{array}{l} \forall x \in Ag, \forall oa \in OA, \forall ia \in IA, \forall R_1, R_2 \in Roles_{oa}, \\ \alpha = \forall y \in Ag, \forall \lambda \in T \;\; y : R_2 \wedge \\ \qquad request(y, x, done(\lambda, \beta)) \Rightarrow \\ \qquad\qquad \exists sc = < x, y, ia, done(\lambda), \beta > \in SC_{ia} \end{array} \right\} \quad [6.5]$$

L'ensemble défini ci-dessus contient des politiques sociales qui sont stockées dans le tableau d'engagements géré par l'agent institutionnel et qui sont valides tant qu'un agent joue son rôle. Elles spécifient que chaque délégation faite par un agent jouant un autre rôle est suivie par la création de l'engagement par l'agent. C'est-à-dire, quand un agent supérieur dans l'hiérarchie décide qu'un agent jouant un rôle doit effectuer une tâche (généralement il s'agit de la satisfaction d'un but), l'agent doit s'engager pour l'effectuer. L'ensemble $RAPol_{oa}$ contient donc toutes les relations d'autorité dans l'organisation représentée par l'agent *oa*. Bien évidemment, dans le cas réel les relations d'autorité existantes ne représentent qu'un sous-ensemble de $RAPol_{oa}$, il est très restrictif pour un agent de devoir s'engager pour n'importe quelle tâche déléguée par n'importe qui dans l'organisation et donc les tâches pour lesquelles un rôle est sous l'autorité d'un autre doivent être spécifiées explicitement.

Nous voulons souligner la similarité qui existe entre la formule ci-dessus représentant l'autorité que le rôle R_2 a sur le rôle R_1 et la Formule 6.3 représentant une obligation d'un agent d'obéir les demandes de l'organisation, c.-à-d., l'autorité que l'organisation a sur un agent jouant le rôle. Même si les représentations des deux concepts sont similaires, nous considérons qu'il existe une différence conceptuelle entre les obligations d'un agent (et les politiques sociales associées), obligations qui dans leur nature peuvent être même interpersonnelles, et la hiérarchie de rôles (et les politiques sociales associées), hiérarchie qui est un concept purement institutionnel.

Comme les politiques sociales présentées ci-dessus représentent des liens d'autorité entre agents jouant des rôles dans une organisation et comme l'objet de ces autorités est en général la satisfaction d'un but, nous utilisons un raccourci de notation en utilisant le prédicat *authority* qui signifie tout simplement qu'il existe une relation d'autorité entre deux agents concernant un but, relation associée aux rôles qu'ils jouent dans une organisation :

$$\begin{array}{l} authority(y, x, g, oa) \equiv \exists R_1, R_2 \in Roles_{oa} \;\; x : R_1 \wedge \;\; y : R_2 \wedge \;\; \exists sp \in RAPol_{oa,} \\ \qquad sp = < x, oa, ia, \\ \qquad\qquad request(y, x, achieved(g, \beta)) \Rightarrow \\ \qquad\qquad\qquad \exists < x, y, ia, achieved(g), \beta > \in SC_{ia} \\ \qquad\qquad , x : R_1 > \end{array} \quad [6.6]$$

où $x, y \in Ag,\ oa \in OA, g \in G, ia \in IA, \beta \in EF$

Exemple 6.1 (cont.)

Nous étendons l'exemple précédent dans lequel les agents négocient pour former des contrats en considérant que la négociation est faite à l'aide d'un broker. Nous considérons ainsi un autre rôle, *RBroker*, joué par un agent *ba* (agent broker). Dans cette organisation il existe une relation d'autorité entre le rôle *RContr* et ce rôle : chaque fois qu'un agent jouant le rôle *RContr* délègue

une tâche de négociation (le but *neg*) à un agent jouant le rôle *RBroker*, ce dernier doit adopter et satisfaire ce but. Ceci est représenté comme une politique sociale qu'un agent qui joue le rôle *RBroker* a envers l'agent organisationnel *ma*, politique qui spécifie qu'il doit s'engager pour la satisfaction du but *neg* quand il lui est demandé :

$$< ba, oa, ia,\ \forall ca \in Ag \quad ca : RContr \wedge request(ca, ba, achieved(< neg, ba, \beta >)) \Rightarrow$$
$$\exists < ba, ca, ia, achieved(< neg, ba, \beta >), \beta > \in SC_{ia},\ \ ba : RBro\,\text{ker} >$$

6.3. Définition et attribution de rôle

Dans la section précédente nous avons montré comment les engagements et politiques sociaux (meta-engagements) peuvent exprimer les contraintes imposées aux agents qui jouent des rôles. Ces engagements définissent donc les caractéristiques d'un rôle et ils peuvent être utilisés pour définir le concept de rôle. Les rôles sont utilisés pour décrire le comportement désiré des agents au sein d'organisations : ce qu'ils doivent faire (exprimé à l'aide d'engagements et politiques sociaux) et ce qu'ils ont à la disposition pour faire (les ressources et permissions reçues).

6.3.1. Définition de rôle à l'aide d'engagements sociaux

Nous définissons donc un rôle comme un tuple formé par le nom du rôle et ses spécifications telles que nous les avons énumérées dans la section 6.1 : les ensembles de ressources et de permissions reçues par l'agent qui le joue et les ensembles d'engagements pour satisfaire des buts, de politiques sociales pour les obligations et pour les relations d'autorité qui concernent l'agent qui le joue :

$$\langle R, \text{Res}_r, \text{Know}_r, \text{Perm}_r, \text{RGComm}_r, \text{ROPol}_r, \text{RAPol}_r \rangle$$
$$\text{où } R \in Roles_{oa}, oa \in OA \subseteq Ag \qquad [6.7]$$

Nous notons par Res_r l'ensemble de ressources reçues par un agent qui joue un rôle R – l'indice utilisé correspond au nom du rôle, par $Know_r$ l'ensemble de connaissances reçues par un agent qui joue un rôle R et par $Perm_r$ l'ensemble de permissions reçues par l'agent. L'ensemble de ressources du rôle est un sous-ensemble de l'ensemble Res de toutes les ressources du système, les éléments de l'ensemble de connaissances peuvent appartenir aux ensembles d'actions et plans possibles (l'agent apprend comment exécuter des actions et des plans), tandis que les éléments de l'ensemble de permissions peuvent appartenir aux ensembles de ressources, actions, buts ou plans possibles (les permissions peuvent cibler tout type de tâche), comme précisé dans la section 5.1. Les éléments de l'ensemble $RGComm_r$ sont des engagements sociaux pour les buts du rôle qui ont la forme décrite dans la Formule 6.2, les éléments de l'ensemble $ROPol_r$ sont des politiques sociales pour des obligations du rôle qui ont la forme décrite dans la Formule 6.3 (et ainsi $ROPol_r \subseteq ROPol_{oa}$) et les éléments de l'ensemble $RAPol_r$ sont des politiques sociales pour les liens d'autorité entre rôles qui ont la forme décrite dans la Formule 6.5 (et ainsi $RAPol_r \subseteq RAPol_{oa}$). Nous voulons souligner que les liens d'autorité contenus dans ce dernier ensemble lient le rôle dans les deux sens : avec des rôles sur lesquels il a de l'autorité et avec des rôles qui ont de l'autorité sur lui.

6.3.2. Jouer un rôle

Un agent qui décide de jouer un rôle dans une organisation doit intégrer les caractéristiques du rôle qu'il joue. Comme d'autres approches (p.ex. : (Vázquez-Salceda 2004) que nous avons présentées dans le Chapitre 2, nous considérons que le fait de jouer un rôle peut être vu comme un contrat entre un agent et l'organisation. Cependant, dans notre modèle nous utilisons la notion d'engagement social qui est plus générale que celle de contrat : nous considérons ainsi qu'un agent *s'engage envers l'organisation pour son rôle*, c.-à-d., quand un agent commence à jouer un rôle, un engagement spécial est créé entre l'agent et l'organisation. L'agent s'engage à avoir un comportement conforme aux caractéristiques du rôle : de respecter les permissions qu'il a et de remplir les engagements et politiques sociaux qui en découlent. Cet *engagement de rôle* a les mêmes propriétés qu'un engagement social classique (nombre de paramètres et leur signification) :

$$RoleComm = \left\{ < x, oa, ia, \alpha, true > \in SC_{ia} \;\middle|\; \begin{array}{l} \forall x \in Ag, \forall oa \in OA, \forall ia \in IA, \\ \alpha =< R, \mathrm{Res_r}, \mathrm{Know_r}, \mathrm{Perm_r}, \mathrm{RGComm_r}, \mathrm{ROPol_r}, \mathrm{RAPol_r} > \end{array} \right\}$$

[6.8]

L'ensemble décrit ci-dessus contient les engagements de rôle possibles, engagements qui ont quelques caractéristiques intéressantes. L'objet d'un tel engagement concerne une formule logique contenant un seul atome : le tuple contenant les caractéristiques du rôle à jouer (représenté comme dans la Formule 6.7). La condition de validité d'un tel engagement est toujours vraie, ce qui fait que cet engagement reste toujours valide une fois qu'il a été créé jusqu'au moment où il sera annulé (l'agent quitte son rôle). Bien évidemment, il est possible d'envisager des engagements de rôle qui expirent, mais ceci n'est pas nécessaire pour notre modèle. De plus, vu que l'objet de l'engagement est un atome, cet engagement ne peut pas être violé – en fait, c'est seulement son existence qui compte dans notre modèle, pas son état. Pour mieux comprendre pourquoi cette représentation simplifiée d'un engagement nous est utile, analysons ce qui se passe avec un agent qui joue un rôle.

Considérons le cas d'un agent *x* qui veut jouer un rôle *R* dans une organisation représentée par un agent organisationnel *oa*. Il interagit avec l'agent organisationnel et comme résultat, l'engagement de rôle est créé et stocké dans un tableau d'engagements géré par l'agent institutionnel *ia*. La notation que nous avons utilisé auparavant pour signifier qu'un agent joue un rôle, *x:R*, est définie comme le fait que l'engagement de rôle existe et donc il est valide :

$$\begin{array}{l} x : R \;\equiv\; \exists rc \in RoleComm, \\ rc =< x, oa, ia, < R, \mathrm{Res_r}, \mathrm{Know_r}, \mathrm{Perm_r}, \mathrm{RGComm_r}, \mathrm{ROPol_r}, \mathrm{RAPol_r} >, true > \\ \text{où } x \in \mathrm{Ag}, oa \in \mathrm{OA}, ia \in \mathrm{IA}, R \in \mathrm{Roles_{oa}} \end{array}$$

[6.9]

Tous les engagements et politiques sociales qui caractérisent le rôle (et qui sont spécifiés dans les ensembles présents dans le tuple définissant le rôle) sont automatiquement créés quand un agent s'engage pour jouer un rôle. Comme nous l'avons précisé auparavant (Section 6.2), ces engagements ont comme condition de validité le fait qu'un agent joue un rôle. Ceci fait que les contraintes institutionnelles imposées à un agent deviennent valides si et seulement si l'agent joue un rôle, et pas

autrement. A tout moment, l'engagement de rôle d'un agent peut être annulé, soit à la demande de l'agent qui ne veut plus jouer le rôle, soit comme sanction pour une violation, etc. Les engagements et politiques sociaux qui caractérisent le rôle deviennent automatiquement invalides parce que leur condition de validité (*x:R*) devient fausse.

En ce qui concerne les ressources, actions, plans et permissions d'un agent, ils évoluent aussi quand il joue ou ne joue plus un rôle en fonction des ressources, connaissances et permissions associées à ce rôle. Nous définissons ainsi les ensembles de ressources qu'un agent possède, actions et plans qu'il sait exécuter et permissions qu'il a en fonctions des ensembles de ressources, connaissances et permissions de ses rôles :

$$\begin{aligned} &Res_x = Res_0 \cup \bigcup_{x:R_i} Res_{r_i}, \quad Act_x = Act_0 \cup \bigcup_{x:R_i} Know_{r_i} \cap Act \\ &Pl_x = Pl_0 \cup \bigcup_{x:R_i} Know_{r_i} \cap Pl, \quad Perm_x = \bigcup_{x:R_i} Perm_{r_i} \end{aligned} \qquad [6.10]$$

Dans notre modèle nous notons par Res_0 (resp. Act_0, Pl_0) les ressources (resp. actions, plans) qu'un agent a indépendamment des rôles qu'il joue et nous considérons que toute permission qu'il a est issue d'un de ses rôles, c.-à-d., il n'existe pas d'autres permissions que celles institutionnelles. D'une manière similaire, dans notre modèle, quand un agent ne joue plus un rôle (l'expression correspondante *x:R* devient fausse), il perd automatiquement les ressources et permissions qu'il avait reçues pour jouer ce rôle et il « oublie » les actions et plans qu'il a appris quand il a commencé à jouer le rôle. Même si nous pouvons facilement étendre notre modèle pour prendre en compte des situations comme par exemple le fait qu'un agent garde des ressources de son rôle même après avoir arrêté de le jouer, ceci n'est pas important dans notre travail et nous préférons ainsi garder notre modèle le plus simple possible.

Synthèse

Dans ces deux dernières sections nous avons montré comment les engagements et les politiques sociales peuvent être utilisés pour représenter les spécifications comportementales associées aux rôles dans une organisation et aussi comment le fait de jouer un rôle peut être analysé à l'aide du même paradigme. Nous considérons que cette représentation est très intéressante, surtout si nous prenons en compte le fait qu'à la base les engagements sociaux sont utilisés pour décrire le résultat des interactions (pas forcement institutionnelles) entre agents. La modélisation proposée ci-dessus pour les spécifications comportementales a été volontairement gardée la plus simple possible, nous sommes conscients du fait qu'elle doit devenir plus puissante afin de baser la conception d'un système organisationnel sur elle. Même si nous considérons une telle conception intéressante, elle ne constitue pas notre objectif dans ce travail ; notre objectif est de permettre à un agent de représenter les contraintes que son appartenance à une organisation introduit sur son comportement et sa prise de décision. Pour cet objectif, le modèle à base d'engagements sociaux présenté ci-dessus nous suffit, comme nous le verrons dans le chapitre suivant.

Avant de montrer dans le chapitre suivant comment l'existence des engagements et politiques sociaux associés aux concepts institutionnels influence les pouvoirs sociaux des agents et finalement leur autonomie, nous voulons souligner un aspect intéressant du modèle introduit dans les sections

précédentes. Peu importe leur type (engagement social, politique sociale, engagement de rôle) ou origine (les interactions d'un agent avec d'autres agents ou le rôle qu'il joue dans une organisation), les engagements sociaux d'un agent sont stockés dans un tableau d'engagements géré par un agent institutionnel. Nous avons montré comment ce tableau évolue suite aux interactions entre agents ou suite au fait qu'un agent commence à jouer, joue ou arrête de jouer un rôle. Cependant, nous avons délibérément ignoré les attributions de cet agent dans une institution électronique, ce que nous détaillerons par la suite.

6.4. Renforcement d'engagements par l'agent institutionnel

Tout engagement social doit être créé devant un troisième agent, différent des agents créditeur et débiteur. Cet agent, le témoin de l'engagement, donne la dimension sociale d'un engagement qui n'est pas seulement un accord entre deux agents, mais un accord *public*, c.-à-d., un accord connu par au moins un autre agent. Pour cette raison, l'agent institutionnel garde une liste des engagements existants – le tableau d'engagements.

6.4.1. Caractéristiques d'un agent institutionnel

Dans les modèles d'institutions électroniques existantes (cf. Chapitre 2), à part être le témoin d'un engagement, l'agent institutionnel doit aussi renforcer cet engagement, c.-à-d., s'assurer que les agents remplissent leurs engagements et que les sanctions pertinentes soient imposées selon le cas. C'est en fait la raison pour laquelle cet agent est appelé agent *institutionnel* : il contient les mécanismes caractéristiques des institutions qui renforcent l'obéissance aux normes. Chaque fois qu'un engagement est accompli (*fulfilled*), une sanction positive (une récompense) peut être donnée et chaque fois qu'un engagement est violé (*violated*), une sanction négative (une pénalité) peut être imposée. Nous voulons souligner que ces sanctions ne sont pas obligatoires : le concepteur de l'institution (et de l'agent institutionnel) décide si une sanction devrait être utilisée ou non pour chaque type de situation possible.

En général un agent institutionnel qui renforce l'obéissance aux normes doit être capable de :

- détecter les situations quand un engagement social doit changer d'état et de faire le changement dans le tableau d'engagements
- choisir la sanction pertinente : pas de sanction, sanction positive (récompense) ou sanction négative (pénalité)
- imposer des sanctions aux agents, c.-à-d., s'assurer que l'agent en cause reçoit sa récompense ou paie la pénalité.

Comme les engagements sociaux pour lesquels il est le témoin sont stockés dans un tableau d'engagements qu'il gère, l'agent institutionnel a accès à ces engagements et peut modifier leur état. Pour détecter les situations quand il doit changer l'état d'un engagement, il doit surveiller l'environnement, les actions des agents (pour vérifier qu'ils ont effectué les tâches pour lesquelles ils sont engagés) et les interactions entre les agents (pour vérifier par exemple que les relations d'autorité sont respectées), autrement dit il doit être omniscient.

Le choix de sanctions à imposer à un agent suite à la modification de l'état d'un engagement n'est pas facile à faire. Comme nous l'avons discuté dans le Chapitre 2, les travaux similaires, p.ex. (Pasquier *et al.* 2005), utilisent souvent l'hypothèse de la responsabilité stricte (*strict liability*) : le montant de la sanction imposée à un agent doit représenter la valeur de la tâche pour laquelle il était engagé. Par exemple, un agent qui a rempli un engagement pour exécuter une tâche avec la valeur 10, recevra une récompense de 10. S'il viole cet engagement, il paiera une pénalité de 10. Cependant, il est généralement difficile d'évaluer la « valeur » de toutes les tâches possibles et il est parfois envisageable de ne pas imposer des sanctions pour certains engagements.

Nous voulons également souligner le fait qu'il ne faut pas considérer tout échec de satisfaire un engagement comme une violation intentionnelle qui doit être punie. Prenons par exemple le cas d'un engagement social représentant une obligation d'un agent. Si l'agent n'arrive pas à effectuer la tâche – objet de l'engagement avant la date limite, l'engagement aura l'état *violated*. Même si l'engagement est violé, l'agent institutionnel doit décider si l'obligation a été violée ou pas. Par exemple, si l'agent a essayé d'effectuer la tâche mais il n'a pas réussi, l'obligation peut être considérée comme non-violée (c'était un accident) et donc l'agent ne paiera pas une sanction.

Dans les organisations artificielles il est souvent difficile d'estimer si un agent a eu ou pas l'intention de violer une norme, ceci étant une raison d'utiliser l'hypothèse de la responsabilité stricte. Tout échec d'accomplir un engagement (d'obéir une obligation) est considéré comme violation et une sanction *est imposée* à l'agent et *payée* par celui-ci. Un autre problème apparaît maintenant : comment imposer des sanctions aux agents autonomes, c.-à-d., aux agents qui peuvent refuser de les payer ?

6.4.2. Comment imposer des sanctions aux agents autonomes ?

Les travaux sur les institutions d'agents présentés dans le Chapitre 2 utilisent parfois un module qui gère les engagements sociaux et qui doit être intégré par les agents : si un agent veut appartenir à un système à base d'engagements sociaux, il doit intégrer ce module. L'agent représente et utilise ses engagements à travers ce module qui ne viole pas les engagements. Si l'état d'un engagement fait qu'une sanction doit être imposée à l'agent, ce module, intégré dans l'agent, s'assure que l'agent la paie. Cette solution est utilisée souvent parce qu'elle garantit qu'un agent ne peut pas refuser une sanction qui lui est imposée, mais elle a l'inconvénient de demander aux agents d'intégrer un module externe. Une autre solution pour assurer le fait que les agents autonomes ne refusent pas leurs sanctions est d'imposer des sanctions sociales (voir Chapitre 2). Par exemple, la sanction pour la violation d'un engagement peut être de baisser la réputation de l'agent violateur. Comme la réputation est une notion sociale et sa valeur est une croyance répartie entre les agents du système, un agent ne peut pas contrôler ou refuser une telle sanction.

Une autre possibilité est d'utiliser des *engagements en cascade* où une violation d'un engagement constitue une condition de validité d'un engagement de niveau supérieur. Quand un engagement est violé, un engagement de payer une pénalité devient automatiquement valide. Si un agent viole cet engagement aussi (il ne paie pas), un autre engagement devient valide, pour payer par exemple une somme plus large. Ce système peut continuer sur plusieurs niveaux d'engagements jusqu'à un niveau ou la pénalité pour la violation est l'exclusion du système (ou de l'organisation/institution).

Dans notre modèle nous n'essayons pas de concevoir une institution électronique et nous ne nous intéressons donc pas directement aux problèmes liés aux agents institutionnels mentionnés ci-dessus. Cependant, notre objectif est d'avoir des agents capables de raisonner sur des contraintes sociales que nous avons représentées sous la forme d'engagements sociaux. Ceci implique qu'un agent doit représenter à l'aide de notre modèle non seulement ces contraintes, mais aussi les implications de leur violation. Nous considérons ainsi que les agents utilisant notre modèle utilisent l'hypothèse de la responsabilité stricte. Autrement dit, un agent qui raisonne à l'aide du modèle que nous proposons fait l'hypothèse que toute violation d'engagement social sera détectée et punie. Quant à la forme de la punition, elle dépend du cas d'utilisation et donc du système institutionnel dans lequel l'agent se trouve. Nous verrons dans la partie suivante comment le raisonnement des agents sur leur comportement potentiellement autonome (violation des engagements) prend en compte les sanctions.

6.5. Synthèse

Dans le chapitre précédent nous avons proposé un modèle d'engagements sociaux qui nous permet de représenter les contraintes interpersonnelles issues des interactions entre agents. Nous avons utilisé ce modèle d'engagements et politiques sociales pour décrire également les spécifications des rôles dans une organisation et la signification de jouer un rôle pour un agent. Nous avons aussi décrit comment ce modèle nous permet de rendre explicite comment les contraintes imposées à un agent se modifient quand l'agent commence à jouer, joue ou arrête de jouer un rôle et nous avons décrit comment des sanctions peuvent être imposées aux agents autonomes qui ne remplissent pas leurs engagements.

Nous nous sommes inspirés dans cette approche des travaux similaires présentés dans le Chapitre 2, travaux qui proposent la notion de contrat pour modéliser le comportement attendu d'un agent qui joue un rôle dans une organisation (p.ex. : (Vázquez-Salceda 2004) ou (Boella & van der Torre 2004a). Cependant, en utilisant les engagements sociaux, notre approche est plus générique : les engagements sociaux, bien que similaires aux contrats, représentent une notion plus générale – ils ne sont pas essentiellement institutionnels comme les contrats, mais interpersonnels (liés aux interactions entre agents).

Le même concept nous permet ainsi de représenter les contraintes interpersonnelles (issues des interactions entre agents – comme nous l'avons vu dans le chapitre précédent) *et* institutionnelles (imposées par le fait que les agents jouent des rôles). Comme nous l'avons discuté dans la partie précédente (notamment dans le Chapitre 3 et dans la synthèse), les engagements sociaux représentent des contraintes comportementales (qui limitent le comportement des agents), tandis que les politiques sociales représentent des contraintes décisionnelles (qui limitent la prise de décision des agents). Dans le chapitre suivant nous allons nous intéresser à d'autres sources de contraintes décisionnelles à l'aide des théories de la dépendance et du pouvoir social. Nous allons les synthétiser en montrant comment tous ces modèles se complètent réciproquement en utilisant la notion d'autonomie.

7. Pouvoir social et autonomie

Comme nous l'avons présenté dans le Chapitre 4, un modèle de raisonnement *a priori* interpersonnel a été proposé par Sichman (Sichman 1995), raisonnement qui utilise des situations de dépendance. La notion de dépendance est aussi utilisée dans le cadre de la théorie du pouvoir social (voir par exemple (Castelfranchi 2002)), théorie qui peut également fournir les bases d'un raisonnement social, comme nous l'avons discuté dans le Chapitre 4. Dans ce chapitre nous nous intéressons aux notions de dépendance et pouvoir social pour montrer comment elles peuvent être utilisées pour définir la notion d'autonomie qui représente la notion centrale de cette thèse.

Nous nous intéressons ici à la description de ce qu'un agent est capable de faire tout seul, sans l'aide d'autres agents, par exemple s'il peut satisfaire un but (ou en général d'exécuter une tâche) tout seul ou s'il a le droit de le faire. Après avoir modélisé ces *pouvoirs individuels* d'un agent, nous nous intéressons aux cas où ces pouvoirs manquent : comment un agent peut satisfaire ses buts avec l'aide d'autres agents. Nous identifions et définissons ainsi plusieurs types de *relations de dépendance*, relations qui lient un agent à un autre ou un agent à un rôle et nous soulignons des situations de dépendance dans lesquelles un agent peut se retrouver. Nous montrons ensuite comment les agents peuvent acquérir de nouveaux pouvoirs par l'intermédiaire d'autres agents et comment les relations de dépendance génèrent des relations de *pouvoir social* entre agents. D'autres relations de pouvoir sont aussi identifiées, notamment le pouvoir issu des contraintes institutionnelles.

A travers plusieurs paradigmes intégrés dans notre modèle (engagements sociaux, relations de dépendance, pouvoir social), nous arrivons ainsi à représenter d'une manière uniforme tous les types de contraintes qui limitent le comportement et la prise de décision d'un agent, qu'elles soient interpersonnelles ou institutionnelles. Comme nous l'avons discuté dans le Chapitre 1, ces contraintes définissent l'autonomie sociale en délibération d'un agent : à la fin de ce chapitre nous serons donc en mesure de définir formellement ce concept et montrer comment il peut être calculé.

7.1. Pouvoirs individuels

Comme nous l'avons précisé dans le Chapitre 3, nous considérons que la notion de pouvoir n'est pas intrinsèquement sociale : elle exprime des relations sociales intéressantes, mais à la base le pouvoir se réfère à la relation entre un agent, ses buts et ses capacités et ressources. Dans cette section nous nous intéressons aux pouvoirs *individuels* d'un agent, c.-à-d. aux pouvoirs qu'un agent a en faisant abstraction d'autres agents, comme par exemple le pouvoir d'un agent de satisfaire un de ses buts. Dans notre modèle, nous faisons la différence entre la capacité *d'exécution* (ce qu'un agent peut « physiquement ») et la capacité déontique (ce qu'un agent « a le droit de faire ») des agents. La première dénote le fait qu'un agent a la capacité « physique » et le savoir-faire pour exécuter une action, satisfaire un but, etc. La deuxième dénote le fait qu'un agent a les permissions nécessaires pour le faire. Ces deux types de capacités sont à la base des pouvoirs individuels d'un agent que nous présentons dans la suite.

7.1.1. Pouvoir d'exécution – *can*

Le pouvoir d'exécution d'un agent représente la capacité d'un agent à effectuer une tâche (action, ressource, but ou plan). Nous notons le *pouvoir d'exécution* d'un agent avec le prédicat *can(agent, tâche)*. La signification de ce prédicat dépend du type de tâche, comme nous le verrons par la suite.

(1) Pouvoir d'exécution d'une ressource

Un agent a *le pouvoir d'exécution d'une ressource* s'il possède cette ressource : le pouvoir d'exécution dans ce cas représente la capacité d'un agent à utiliser une ressource – dans notre modèle il a cette capacité s'il possède la ressource :

$$can(x,r) \equiv r \in \mathrm{Res}_x \text{ , où } x \in Ag,\ r \in Res \qquad [7.1.a]$$

(2) Pouvoir d'exécution d'une action

Un agent a le *pouvoir d'exécution d'une action* s'il sait exécuter l'action : le pouvoir d'exécution dans ce cas représente la capacité d'un agent à exécuter une action – dans notre modèle l'agent a cette capacité si l'action appartient à son répertoire :

$$can(x,a) \equiv a \in \mathrm{Act}_x \text{ , où } x \in Ag,\ a \in Act \qquad [7.1.b]$$

(3) Pouvoir d'exécution d'un plan

Un agent a le *pouvoir d'exécution d'un plan* s'il a le pouvoir d'exécution des éléments du plan et s'il est capable d'éviter les conflits inhérents au plan : le pouvoir d'exécution dans ce cas représente la capacité d'un agent à exécuter tous les éléments d'un plan. Pour cela il doit avoir le pouvoir d'exécution de chaque élément *et* de toutes les pré-conditions de ces éléments. De plus, un plan dans lequel existent des conflits entre ses éléments (p.ex. : des actions qui doivent être exécutées en parallèle, etc.) ne peut pas être exécuté par un agent tout seul :

$$can(x,pl) \equiv (\forall \alpha \in pl\ can(x,\alpha)) \wedge (\forall \alpha \in pl\ \exists \alpha' \in \mathrm{T}\ enables(\alpha',\alpha) \Rightarrow can(x,\alpha')) \wedge (\forall \alpha,\alpha' \in pl\ \neg conflict(pl,\alpha,\alpha')), \qquad [7.1.c]$$

où $x \in Ag$, $pl \in Pl$

Nous voulons souligner que dans la définition du pouvoir d'exécution d'un plan, il n'est pas nécessaire pour un agent de connaître le plan : cette définition s'intéresse seulement à la capacité d'un agent à exécuter un plan et non à ses connaissances. Cependant, cet aspect devient important dans le cas du pouvoir d'exécution d'un but.

(4) Pouvoir d'exécution d'un but

Un agent a le *pouvoir d'exécution d'un but* s'il connaît un plan qui satisfait le but et s'il a le pouvoir d'exécution de ce plan. Dans ce cas aussi, cette définition s'intéresse seulement à la capacité d'un agent à satisfaire un but et non si l'agent a ou non ce but :

$$can(x,g) \equiv \exists pl_g \in \mathrm{Pl}_x \; g \in pl_g \wedge can(x,pl_g), \text{ où } x \in Ag,\ g \in G \qquad [7.1.d]$$

Ces quatre formules définissent le pouvoir d'exécution des agents pour tout type de tâche, c.-à-d. la capacité des agents à utiliser des ressources, à exécuter des actions ou des plans ou à satisfaire des buts.

Exemple 7.1

Considérons maintenant comme exemple le plan décrit dans l'Exemple 5.1, plan qui satisfait un but *g* par la satisfaction d'un sous-but *neg* suivie de l'exécution d'une action *exec_job*. En ce qui concerne le but *neg*, le seul plan qui le satisfait consiste en l'exécution en parallèle de deux actions neg_1 et neg_2. Un conflit existe ainsi au sein de ce plan : aucun agent ne peut exécuter les deux actions de négociation tout seul. Conformément à la Formule 7.1.c, aucun agent n'a le pouvoir d'exécution de ce plan et, comme ces plans sont les seuls à exister pour la satisfaction des buts *g* et *neg*, aucun agent n'a le pouvoir d'exécution de ces buts. L'expression ci-dessous est donc vraie :

$$\forall x \in \mathrm{Ag} \quad \neg can(x,pl_{neg}) \wedge \neg can(x,pl_g) \wedge \neg can(x,neg) \wedge \neg can(x,g)$$

7.1.2. Pouvoir déontique – *may*

Afin d'effectuer une tâche, un agent doit être « physiquement » capable de le faire, mais il doit aussi avoir la permission (le droit) de le faire. Nous notons le *pouvoir déontique* d'un agent avec le prédicat *may(agent,tâche)* et la signification de ce prédicat dépend du type de tâche, comme nous le verrons par la suite.

(1) Pouvoir déontique d'une ressource

Un agent a le *pouvoir déontique d'une ressource* s'il a la permission d'utiliser cette ressource – comme nous l'avons précisé auparavant, l'ensemble $Perm_x$ contient toutes les tâches qu'un agent *x* a la permission d'effectuer :

$$may(x,r) \equiv r \in \mathrm{Perm}_x \text{ , où } x \in Ag,\ r \in Res \qquad [7.2.a]$$

(2) Pouvoir déontique d'une action

Un agent a le *pouvoir déontique d'une action* s'il a la permission de l'exécuter – dans notre modèle ceci est équivalent au fait que l'action se retrouve parmi les éléments de l'ensemble de permissions de l'agent :

$$may(x,a) \equiv a \in Perm_x \text{ , où } x \in Ag,\ a \in Act \qquad [7.2.b]$$

(3) Pouvoir déontique d'un plan

Un agent a le *pouvoir déontique d'un plan* s'il a la permission de l'exécuter et s'il a la permission d'exécuter les éléments du plan. Ceci signifie que dans notre modèle, il ne suffit pas pour un agent

d'avoir les permissions nécessaires pour les tâches qui forment un plan, mais aussi la permission d'utiliser le plan. En ce qui concerne les conflits entre ces éléments ou leurs pré-conditions, ces aspects ne font pas partie de la définition du pouvoir déontique : ils influencent la capacité d'un agent d'exécuter le plan (son pouvoir d'exécution), mais ils n'influencent pas les droits qu'il a :

$$may(x,pl) \equiv pl \in \mathrm{Perm}_x \wedge \forall \alpha \in pl\ may(x,\alpha)\ , \text{ où } x \in Ag,\ pl \in Pl \qquad [7.2.c]$$

Nous voulons souligner que, de la même manière que pour le pouvoir d'exécution, il n'est pas nécessaire pour un agent de connaître le plan pour avoir le pouvoir déontique de l'exécuter.

(4) Pouvoir déontique d'un but

Un agent a le *pouvoir déontique d'un but* s'il a la permission de poursuivre ce but et s'il connaît un plan qui satisfait le but et pour lequel il a le pouvoir déontique. Dans ce cas aussi, cette définition s'intéresse seulement au droit d'un agent de satisfaire un but et non si l'agent a ou pas ce but :

$$may(x,g) \equiv g \in \mathrm{Perm}_x \wedge \left(\exists pl_g \in \mathrm{Pl}_x\ g \in pl_g \wedge may(x,pl_g)\right), \text{ où } x \in Ag,\ g \in G \qquad [7.2.d]$$

Ces quatre formules définissent le pouvoir déontique des agents pour tout type de tâche, c.-à-d. le droit des agents d'utiliser des ressources, d'exécuter des actions ou des plans ou de satisfaire des buts. Comme nous l'avons précisé dans le chapitre précédent, les permissions qu'un agent a sont issues des rôles qu'il joue, même si d'autres possibilités peuvent être envisagées (p.ex. : un agent peut donner des permissions à un autre). Du point de vue des définitions ci-dessus, le sujet qui donne ces permissions et la manière dont elles sont données ne sont pas importants. Ce qui nous intéresse est seulement de savoir si l'agent a ou n'a pas ces permissions.

Exemple 7.1 (cont)

Considérons maintenant comme exemple le plan décrit dans l'Exemple 5.1, plan qui satisfait un but *g* par la satisfaction d'un sous-but *neg* suivie de l'exécution d'une action *exec_job*. En ce qui concerne le but *neg*, le seul plan qui le satisfait concerne l'exécution en parallèle de deux actions neg_1 et neg_2. Un conflit existe ainsi au sein de ce plan : aucun agent ne peut exécuter les deux actions de négociation tout seul. A partir de la Formule 7.1.c, aucun agent n'a le pouvoir d'exécution de ce plan et, comme ces plans sont les seuls à exister pour la satisfaction des buts *g* et *neg*, aucun agent n'a le pouvoir d'exécution de ces buts, l'expression ci-dessous est donc vraie :

$$\forall x \in \mathrm{Ag}\ \ \neg can(x,pl_{\mathrm{neg}}) \wedge \neg can(x,pl_g) \wedge \neg can(x,neg) \wedge \neg can(x,g)$$

Considérons maintenant le cas d'un agent qui a les permissions d'exécuter les actions neg_1 et neg_2 : à partir de la Formule 7.2.c, s'il a aussi la permission d'exécuter le plan qui satisfait le but *neg* et qui est composé de ces deux actions, alors il a le pouvoir déontique pour ce plan. Si cet agent a aussi la permission de poursuivre le but *neg* et il connaît le plan en cause, alors, selon la Formule 7.2.d, il a le pouvoir déontique pour ce but :

$$\forall x \in \mathrm{Ag}\ \ \mathrm{Perm}_x = \{neg_1, neg_2, pl_{\mathrm{neg}}\} \xrightarrow{cf.\ 7.2c} may(x, pl_{\mathrm{neg}})$$
$$\forall x \in \mathrm{Ag}\ \ neg \in \mathrm{Perm}_x \wedge may(x, pl_{\mathrm{neg}}) \xrightarrow{cf.\ 7.2d} may(x, g)$$

Cet exemple montre la complémentarité entre les deux types de pouvoir individuel, le pouvoir d'exécution et le pouvoir déontique. Il est possible pour un agent d'avoir un type de pouvoir sans avoir l'autre et cette situation est fréquente dans les sociétés d'agents où les agents n'ont pas le droit d'effectuer tout ce dont ils sont capables. Par la suite nous nous intéressons au cas où un agent a les deux types de pouvoir pour la même tâche.

7.1.3. Pouvoir individuel – *power of*

Peu importe la nature d'une tâche d'un agent (action ou plan à exécuter, ressource à utiliser, but à satisfaire), nous considérons qu'un agent est vraiment en position de l'effectuer seulement s'il a les deux capacités de base pour cette tâche. Si un agent *x* a à la fois le pouvoir d'exécution et celui déontique pour une tâche, il pourra l'effectuer sans l'aide d'autres agents- nous appelons ceci le *pouvoir individuel* d'un agent et nous le notons avec le prédicat *power_of(agent,tâche)*.

(1) Pouvoir individuel d'une ressource, Pouvoir individuel d'une action

Un agent a le *pouvoir individuel pour une ressource* s'il a le pouvoir d'exécution (il la possède) et il a aussi le pouvoir déontique (il a la permission de l'utiliser) de cette ressource. D'une manière similaire, un agent a le *pouvoir individuel pour une action* s'il a le pouvoir d'exécution (il sait l'exécuter) et il a aussi le pouvoir déontique (il a la permission de l'exécuter) de cette action :

$$\begin{aligned} &power\_of(x,r) \equiv can(x,r) \wedge may(x,r), \\ &power\_of(x,a) \equiv can(x,a) \wedge may(x,a), \\ &\text{où } x \in Ag,\ r \in Res,\ a \in Act \end{aligned} \qquad [7.3.a]$$

(2) Pouvoir individuel d'un plan

En ce qui concerne les plans, la définition du pouvoir individuel d'un agent n'est plus aussi simple. Un agent a le *pouvoir individuel pour un plan* s'il a le pouvoir d'exécution (il sait l'exécuter) et il a aussi le pouvoir déontique (il a la permission de l'exécuter) de ce plan, mais d'autres conditions doivent être également remplies. Pour être vraiment en mesure d'exécuter un plan, un agent doit aussi le connaître et comme des éléments du plan ont parfois des pré-conditions, il doit avoir le droit de les exécuter (la définition du pouvoir d'exécution inclut ces pré-conditions, mais pas celle du pouvoir déontique) :

$$\begin{aligned} power\_of(x,pl) \equiv\ & pl \in \mathrm{Pl}_x \wedge can(x,pl) \wedge \\ & may(x,pl) \wedge (\forall \alpha \in pl\ \exists \alpha' \in \mathrm{T}\ enables(\alpha',\alpha) \Rightarrow may(x,\alpha')), \\ \text{où } & x \in Ag,\ pl \in Pl \end{aligned} \qquad [7.3.b]$$

(3) Pouvoir individuel d'un but

Concernant les buts, nous considérons qu'un agent a le *pouvoir individuel pour un but* s'il connaît un plan qui satisfait le but et s'il a le pouvoir individuel pour ce plan (voir ci-dessus). Il est donc en

mesure d'exécuter le plan qui satisfait le but. De plus, pour être vraiment en mesure de satisfaire un but, nous considérons qu'un agent doit aussi avoir ce but (un agent ne peut pas poursuivre et ainsi satisfaire un but s'il ne l'a pas) et également il doit avoir la permission de le poursuivre :

$$power\_of(x,g) \equiv g \in G_x \cap Perm_x \wedge (\exists pl_g \in Pl_x \; g \in pl_g \wedge power\_of(x,pl_g)), \quad [7.3.c]$$
où $x \in Ag$, $g \in G$

Ces quatre formules définissent le pouvoir individuel des agents (*power_of*) pour tout type de tâche, c.-à-d. la capacité et le droit des agents d'utiliser des ressources, d'exécuter des actions ou leurs plans ou de satisfaire leurs buts. A partir des définitions précédentes, tout agent a maintenant la possibilité de faire la différence entre ce qu'il peut faire (*can*) et ce qu'il a le droit de faire (*may*). Il est surtout capable de mettre en évidence le manque de tels pouvoirs. Par exemple, un agent qui n'a pas de pouvoir d'exécution pour un de ses buts parce qu'il lui manque une ressource cherchera à acquérir cette ressource. Un agent qui n'a pas de pouvoir déontique pour un de ses buts parce qu'il lui manque la permission d'utiliser une ressource cherchera à obtenir cette permission ou essaiera de désobéir à cette interdiction pour utiliser la ressource. Dans la suite nous présentons la notion de dépendance qui est à la base de ce raisonnement.

7.2. Relations de dépendance

Il est généralement improbable qu'un agent soit capable de satisfaire tous ses buts sans l'aide d'autres agents. Il dépend donc d'autres agents pour satisfaire ces buts. Comme nous le verrons par la suite, la dépendance peut être provoquée par un manque de pouvoir individuel, d'exécution ou déontique. Nous considérons important de souligner le fait que les relations de dépendance que nous allons introduire sont objectives : elles existent même si les agents impliqués dans ces relations ne sont pas conscients de leur existence. Puisqu'un agent n'a pas toujours accès aux informations telles que les buts que les autres ont ou les actions qu'ils savent faire, il lui est souvent difficile de déduire toutes les relations de dépendance qui le concernent. Cependant, identifier les situations de dépendance dans lesquelles il se trouve par rapport à d'autres agents peut s'avérer très utile pour un agent, comme nous le montrerons dans la partie suivante. Dans cette section nous présentons comment dans notre modèle il est possible de calculer les relations de dépendance d'un agent envers un autre liées à l'utilisation d'une ressource, l'exécution d'une action ou d'un plan ou la satisfaction d'un but. Nous montrons ensuite comment des relations de dépendance similaires peuvent être identifiées entre un agent et les rôles qu'il peut jouer et nous finirons par présenter les différentes situations de dépendance.

7.2.1. Relations de dépendance entre agents – *depends*

Analysons maintenant les cas où un agent n'a pas le pouvoir individuel pour une tâche (action, ressource, but ou plan). Pour effectuer cette tâche, il risque ainsi d'avoir besoin de l'aide d'un autre agent : d'après Sichman (voir Chapitre 3), une relation de dépendance existe entre les deux agents. Nous notons cette relation de dépendance d'un agent vers un autre pour une tâche avec le prédicat *depends(agent,agent,tâche)* et par la suite nous identifions les situations dans lesquelles cette relation apparaît en fonction du type de tâche en cause.

(1) Dépendance pour l'utilisation d'une ressource

A partir de la Formule 7.3.a, un agent n'a pas le pouvoir individuel d'utiliser une ressource soit parce qu'il ne possède pas la ressource (il n'a pas le pouvoir d'exécution), soit parce qu'il n'a pas le droit de l'utiliser (il n'a pas le pouvoir déontique). Dans le premier cas, il n'a pas la ressource. Il dépend donc d'un autre agent qui a cette ressource et qui pourra ainsi la lui donner ou l'utiliser à sa place. Dans le deuxième cas, il n'a pas le droit d'utiliser la ressource. Il dépend donc d'un agent qui a ce droit : l'agent qui possède la ressource et a le pouvoir individuel pour elle et qui pourra l'utiliser à sa place. Comme précisé ci-dessus, dans notre modèle les agents ne peuvent pas se donner des permissions les uns aux autres donc cet aspect n'est pas pris en compte dans ce calcul de dépendance.

Pour résumer, *un agent dépend d'un autre agent pour une ressource* s'il n'a pas le pouvoir individuel pour cette ressource et que l'autre agent a ce pouvoir, ou si l'autre agent peut au moins compléter le pouvoir qui manque à l'agent (d'exécution ou déontique) :

$$\begin{aligned} depends(x,y,r) \equiv & \neg power\_of(x,r) \wedge power\_of(y,r) \\ & \vee \neg can(x,r) \wedge may(x,r) \wedge can(y,r) \\ & \vee can(x,r) \wedge \neg may(x,r) \wedge may(y,r) \end{aligned} \qquad [7.4.a]$$

où $x,y \in Ag$, x≠y, $r \in Res$

(2) Dépendance pour l'exécution d'une action

A partir de la Formule 7.3.a, un agent n'a pas le pouvoir individuel d'exécuter une action soit parce qu'il ne sait pas l'exécuter (il n'a pas le pouvoir d'exécution), soit parce qu'il n'a pas le droit de le faire (il n'a pas le pouvoir déontique). Dans le premier cas, il ne sait pas exécuter l'action. Il dépend donc d'un autre agent qui sait exécuter cette action et qui a le droit de l'exécuter. Il a donc le pouvoir individuel pour elle et pourra ainsi l'exécuter à la place de l'autre agent. Dans le deuxième cas, il n'a pas le droit d'exécuter l'action. Il dépend donc d'un agent qui a ce droit et qui sait exécuter cette action. Il a donc le pouvoir individuel pour elle et pourra ainsi l'exécuter à la place de l'autre agent. Dans notre modèle, un agent ne peut pas donner des permissions aux autres et il ne peut pas non plus enseigner comment exécuter une action.

Pour résumer, *un agent dépend d'un autre agent pour une action* s'il n'a pas le pouvoir individuel pour cette action et si l'autre agent a ce pouvoir :

$$depends(x,y,a) \equiv \neg power\_of(x,a) \wedge power\_of(y,a) \qquad [7.4.b]$$

où $x \in Ag$, x≠y, $a \in Act$

(3) Dépendance pour l'exécution d'un plan

En ce qui concerne la dépendance d'un agent envers un autre agent pour un plan, plusieurs raisons existent. A partir des Formules 7.3.b, 7.1.c et 7.2.c, un agent a le pouvoir individuel d'un plan si :

- il connaît le plan et a le droit de l'exécuter ;
- il n'existe pas de conflits entre ces éléments
- il a le pouvoir individuel (d'exécution et déontique) pour chaque élément du plan ;
- il a le pouvoir individuel pour les pré-conditions des éléments du plan.

Analysons maintenant les relations de dépendance liées à ces quatre critères qu'un agent doit remplir pour avoir le pouvoir d'un plan. Si un agent ne connaît pas le plan ou s'il n'a pas le droit de l'exécuter, puisque d'autres agents ne peuvent pas lui faire connaître le plan ou lui donner la permission, il dépend d'un agent qui peut exécuter le plan à sa place, c.-à-d. qui a le pouvoir individuel pour le plan :

$$depends\_1(x,y,pl) \equiv \neg power\_of(x,pl) \wedge power\_of(y,pl) \qquad [7.4.c]$$

où $x \in Ag$, x≠y, $pl \in Pl$

Si, dans un plan, il existe un conflit entre deux tâches, ce conflit est généré par l'impossibilité de faire exécuter ces deux tâches par le même agent (p.ex. : deux actions en parallèle ou avec des durées et dates limites incompatibles, etc.). Pour pouvoir exécuter le plan, un agent dépend donc d'un autre qui pourra effectuer une des deux tâches – un agent doit avoir le pouvoir individuel pour une de ces tâches, tandis que l'autre doit avoir le pouvoir individuel pour l'autre. Cette relation entre les deux agents représente la dépendance d'un agent envers un autre pour un plan à travers une tâche et nous la notons avec le prédicat prédicat *depends(agent,agent,tâche,plan)* :

$$depends\_2(x,y,pl) \equiv \exists \alpha,\alpha' \in pl\ conflict(pl,\alpha,\alpha') \wedge power\_of(x,\alpha) \wedge power\_of(y,\alpha') \qquad [7.4.d]$$

où $x \in Ag$, x≠y, $pl \in Pl$

Si, dans un plan, il existe un élément qu'un agent ne peut pas effectuer (il n'a pas le pouvoir individuel pour cet élément), pour pouvoir exécuter le plan l'agent dépend d'un autre qui pourra effectuer cet élément. Autrement dit, si un agent dépend d'un autre pour une ressource, action ou but, il dépend également de cet agent pour tout plan qui contient cette ressource, action ou but. Ceci est aussi valable pour le cas d'une pré-condition d'un élément d'un plan : si un agent dépend d'un autre pour une ressource, action ou but, il dépend également de cet agent pour tout plan qui contient un élément qui nécessite l'exécution comme pré-condition de cette ressource, action ou but :

$$depends\_3(x,y,\alpha,pl) \equiv \exists \alpha \in pl\ depends(x,y,\alpha) \vee \exists \alpha' \in T\ enables(\alpha',\alpha) \wedge depends(x,y,\alpha') \qquad [7.4.e]$$

où $x \in Ag$, x≠y, $pl \in Pl$

Même si elle n'apparaît pas dans la définition du pouvoir individuel d'un agent pour un plan, il existe une autre source pour la dépendance d'un agent vers un autre. Il est possible qu'un agent soit en mesure d'exécuter un plan, mais il n'y arrivera que si un autre agent ne l'empêche pas. Il s'agit ici de la relation *hinders* (voir Chapitre 5) qui décrit le fait que l'exécution d'une action ou la satisfaction d'un but empêche la réalisation d'un élément d'un plan. L'agent dépend ainsi d'un autre agent qui a le pouvoir individuel pour cette action ou but, dépendance qui a comme objet la non-exécution de l'action (notée avec $\neg a$) ou la non-satisfaction du but (notée avec $\neg g$) :

$$\begin{aligned} depends\_4(x,y,\neg a,pl) &\equiv \forall \alpha \in pl \wedge \exists a \in Act\ hinders(a,\alpha) \wedge power\_of(y,a) \\ depends\_5(x,y,\neg g,pl) &\equiv \forall \alpha \in pl \wedge \exists g \in G\ hinders(g,\alpha) \wedge power\_of(y,g) \end{aligned} \qquad [7.4.f]$$

où $x \in Ag$, x≠y, $pl \in Pl$

Les Formules d à f permettent le calcul d'une relation de dépendance entre deux agents pour un plan à travers une tâche. Cette relation peut être utilisée pour calculer la dépendance « plus générale » entre deux agents pour un plan (dépendance qui n'explicite pas la raison de son existence) :

$$depends(x,y,pl) \equiv depends\_1(x,y,pl) \vee depends\_2(x,y,pl) \vee \exists \alpha \in T: depends\_3(x,y,\alpha,pl) \vee \exists a \in Act: depends\_4(x,y,\neg a,pl) \vee \exists g \in G: depends\_5(x,y,\neg g,pl) \quad [7.4.g]$$

où $x,y \in Ag$, x≠y, $pl \in Pl$

(4) Dépendance pour la satisfaction d'un but

A partir de la Formule 7.3.c, un agent n'a pas le pouvoir individuel de satisfaire un de ses buts, soit parce qu'il n'a pas le droit de le poursuivre (il n'a pas le pouvoir déontique), soit parce qu'il ne connaît aucun plan qui satisfait le but pour lequel il a le pouvoir individuel d'exécution. S'il n'a pas le droit de poursuivre le but ou s'il ne connaît aucun plan qui le satisfait, il dépend d'un autre agent qui a le pouvoir individuel pour satisfaire le but et qui pourra le satisfaire à sa place. Si, par contre, il connaît un tel plan mais qu'il n'a pas le pouvoir individuel pour ce plan et qu'il est en relation de dépendance avec un autre agent pour ce plan, alors il dépend aussi de l'autre agent pour son but :

$$depends(x,y,g) \equiv \neg power\_of(x,g) \wedge (power\_of(y,g) \vee \exists pl_g \in Pl_x \; g \in pl_g \wedge depends(x,y,pl_g)) \quad [7.4.h]$$

où $x,y \in Ag$, x≠y, $g \in G_x$

Dans notre modèle, le comportement des agents est orienté vers la satisfaction des buts – les ressources, actions et plans représentant seulement les éléments nécessaires pour cette satisfaction. Ainsi, comme nous le verrons dans la partie suivante, le raisonnement qui se base sur les relations de dépendance s'intéresse surtout aux dépendances pour la satisfaction des buts, les autres types de relation de dépendance n'étant utilisés que pour le calcul de ces dernières. Pour cette raison, la plupart de nos exemples s'intéressent à ce type de relation de dépendance et non aux autres – cependant notre modèle nécessite le calcul de tous les types de relation de dépendance.

Exemple 7.2

Considérons maintenant le plan décrit dans l'Exemple 5.1, plan qui satisfait un but *g* par la satisfaction d'un sous-but *neg* suivie de l'exécution d'une action *exec_job* – l'exécution de cette dernière peut être empêchée par l'exécution de l'action *cancel*. En ce qui concerne le but *neg*, le seul plan qui le satisfait consiste en l'exécution en parallèle de deux actions neg_1 et neg_2. Un conflit existe ainsi au sein de ce plan : aucun agent ne peut exécuter les deux actions de négociation tout seul :

$$pl_g = \{g, SG_g, Act_g, Res_g\}, \text{où } SG_g = \{neg\}, Act_g = \{exec\_job\}, Res_g = \{\}$$
$$pl_{neg} = \{neg, SG_{neg}, Act_{neg}, Res_{neg}\}, \text{où } SG_{neg} = \{\}, Act_{neg} = \{neg_1, neg_2\}, Res_{neg} = \{\}$$
$$conflict(pl_{neg}, neg_1, neg_2), \; hinders(cancel, exec\_job)$$

Considérons maintenant trois agents *x*, *y* et *z* avec des caractéristiques différentes : l'agent *x* sait exécuter les actions neg_1, neg_2 et *exec_job*, l'agent *y* sait exécuter que les deux premières, tandis que l'agent *z* sait exécuter les actions *exec_job* et *cancel*. Aucun agent ne possède de ressources et

tous les trois connaissent les deux plans pl_g et pl_{neg}. Les agents *x* et *y* ont les permissions d'exécuter les deux actions de négociation, d'exécuter les deux plans et de poursuivre les deux buts, tandis que l'agent *z* a seulement la permission d'exécuter l'action *exec_job* :

$$\text{Res}_x=\text{Res}_y=\text{Res}_z=\{\ \},\ \text{Pl}_x=\text{Pl}_y=\text{Pl}_z=\{pl_g,\ pl_{neg}\}$$
$$\text{Act}_x=\{neg_1, neg_2, exec\_job\}.\text{Act}_y=\{neg_1, neg_2\}.\text{Act}_z=\{cancel, exec\_job\}$$
$$\text{Perm}_x=\text{Perm}_y=\{neg_1, neg_2, pl_g, pl_{neg}, g, neg\}.\text{Perm}_z=\{exec\_job\}$$

Considérons le cas de l'agent *x* et du but *neg* : dans le seul plan qu'il connaît qui satisfait ce but il existe un conflit entre les actions neg_1 et neg_2. A partir de la Formule 7.3.b, il n'a pas le pouvoir individuel pour ce plan. Comme il a le pouvoir individuel pour une de ces deux actions et qu'il existe l'agent *y* qui a le pouvoir pour l'autre, à partir de la Formule 7.4.d, l'agent *x* dépend de l'agent *y* pour l'exécution du plan à travers l'action neg_2. Ainsi, appliquant les Formules 7.4.g et 7.4.h, l'agent *x* dépend de l'agent *y* pour la satisfaction du but *neg* :

$$conflict(pl_{\text{neg}},neg_1,neg_2) \xrightarrow{cf.\ 7.4d} depends(x,y,neg_2,pl_{\text{neg}}) \xrightarrow{cf.\ 7.4g\ et\ 7.4h} depends(x,y,neg)$$

En ce qui concerne le but *g*, étant donné le seul plan que l'agent x connaît et qui satisfait ce but, l'agent *x* n'a pas le pouvoir individuel pour ce but parce qu'il n'a pas le pouvoir de satisfaire le sous-but *neg*, ni le pouvoir individuel pour l'action *exec_job* (il manque la permission). Selon la Formule 7.4.e, il dépend de l'agent *z* qui a le pouvoir individuel pour cette action :

$$\neg power\_of(x,exec\_job) \wedge power\_of(z,exec\_job) \xrightarrow{cf.\ 7.4e} depends(x,z,exec\_job,pl_{\text{g}})$$

Nous voulons souligner que l'agent *z* a le pouvoir d'exécution d'une action *cancel* qui peut empêcher l'exécution du plan par l'agent *x*. Heureusement pour *x*, *z* n'a pas le pouvoir déontique d'exécuter cette action. S'il l'avait, à partir de la Formule 7.4.f, l'agent x dépendrait ainsi de lui pour la non-exécution de l'action. Pour résumer, dans la satisfaction du but *g*, l'agent *x* dépend de l'agent *y* (qui peut l'aider à résoudre un conflit entre deux tâches) et dépend aussi de l'agent *z* (qui peut exécuter à sa place une action qu'il n'a pas le droit d'exécuter) :

$$depends(x,y,g) \wedge depends(x,z,g)$$

7.2.2. Relations de dépendance entre agents et rôles – *role_depends*

Nous avons fait quelques hypothèses dans le calcul des relations de dépendance entre deux agents présenté dans la section précédente. Notamment, nous avons considéré qu'un agent peut transférer des ressources à un autre, mais qu'il ne peut ni lui apprendre comment exécuter une action ou un plan, ni lui donner des permissions. Ceci fait que si un agent manque de l'un de ces pouvoirs (p.ex. : il ne sait pas exécuter une action ou il n'a pas le droit de le faire), il dépend d'un autre agent qui peut le faire à sa place, mais il ne peut pas recevoir le pouvoir qui lui manque. Cependant, dans le modèle de rôles introduit dans le chapitre précédent nous avons décrit comment un agent peut recevoir des ressources, des connaissances et des permissions de la part du rôle qu'il joue.

Notre modèle nous permet ainsi de décrire le fait que si un agent manque d'un pouvoir d'effectuer une tâche, il peut demander à un autre agent de l'effectuer à sa place, mais il peut aussi jouer un rôle et acquérir le pouvoir qui lui manque. Il est donc possible de parler de *relation de dépendance entre un agent et un rôle pour une tâche*, relation que nous notons par le prédicat *role_depends(agent,rôle,tâche)*. Le calcul de cette relation de dépendance est basé sur la différence existante entre les pouvoirs d'un agent quand il joue et quand il ne joue pas un rôle. Nous avons ainsi besoin de représenter d'une manière différente un agent qui joue un rôle ($x+R$) et un agent qui ne joue pas un rôle ($x-R$) :

$$\forall x \in \mathrm{Ag}\ \ \forall R \in \bigcup_{oa \in \mathrm{Ag}} \mathrm{Roles}_{oa}:$$

$$\begin{aligned}
&\mathrm{Res}_{x-R} = \mathrm{Res}_0 \cup \bigcup_{x:R_i}^{R_i \neq R} \mathrm{Res}_{r_i}\ ,\ \mathrm{Res}_{x+R} = \mathrm{Res}_{x\text{-}R} \cup \mathrm{Res}_r \\
&\mathrm{Act}_{x-R} = \mathrm{Act}_0 \cup \bigcup_{x:R_i}^{R_i \neq R} \mathrm{Act}_{r_i} \cap \mathrm{Act},\ \mathrm{Act}_{x+R} = \mathrm{Act}_{x\text{-}R} \cup \mathrm{Know}_r \cap \mathrm{Act} \\
&\mathrm{Pl}_{x-R} = \mathrm{Pl}_0 \cup \bigcup_{x:R_i}^{R_i \neq R} \mathrm{Pl}_{r_i} \cap \mathrm{Pl},\ \mathrm{Pl}_{x+R} = \mathrm{Pl}_{x\text{-}R} \cup \mathrm{Know}_r \cap \mathrm{P}l \\
&\mathrm{Perm}_{x-R} = \bigcup_{x:R_i}^{R_i \neq R} \mathrm{Perm}_{r_i}\ ,\ \mathrm{Perm}_{x+R} = \mathrm{Perm}_{x-R} \cup \mathrm{Perm}_r
\end{aligned} \qquad [7.5]$$

Les ressources d'un agent x sans prendre en compte le rôle R joué sont données par les ressources propres à l'agent (Res_0) et la réunion de toutes les ressources provenant des rôles qu'il joue, excepté le rôle R. La même signification reste valable pour actions, plans et permissions, respectivement. Nous voulons souligner que la notation utilisée dans le chapitre précédent, $x:R$, a une signification plus riche (un engagement de rôle existant et valide, etc.), tandis que ces deux notations utilisées ici (x-R et $x+R$) ne représentent que les modifications apportées aux ensembles de ressources, actions, plans et permissions d'un agent (ensembles définis dans la Formule 6.14) par le fait de jouer ou pas un rôle.

(1) Rôle-dépendance pour l'utilisation d'une ressource

Dans le cas d'une ressource, un agent peut recevoir d'un rôle la ressource (physique) et/ou la permission de l'utiliser. Si quand il ne joue pas le rôle il n'a pas le pouvoir individuel pour cette ressource, mais quand il le joue il a ce pouvoir, nous considérons que *l'agent dépend de ce rôle pour l'utilisation de la ressource* (il a besoin de jouer le rôle pour avoir le pouvoir) :

$$\begin{aligned}
&role_depends(x,R,r) \equiv \neg\, power_of(x-R,r) \wedge power_of(x+R,r), \\
&\text{où } x \in Ag, r \in \mathrm{Res},\ R \in \bigcup_{oa \in OA} \mathrm{Roles}_{oa}
\end{aligned} \qquad [7.6.a]$$

(2) Rôle-dépendance pour l'exécution d'une action

D'une manière similaire, un agent peut recevoir d'un rôle le savoir-faire d'une action et/ou la permission de l'exécuter. Si, quand il ne joue pas le rôle, il n'a pas le pouvoir individuel pour cette

action, mais quand il le joue il a ce pouvoir, nous considérons que *l'agent dépend de ce rôle pour l'exécution de l'action* (il a besoin de jouer le rôle pour avoir le pouvoir) :

$$role\_depends(x,R,a) \equiv \neg\, power\_of(x-R,a) \wedge power\_of(x+R,a),$$
$$\text{où } x \in Ag, a \in Act,\ R \in \bigcup_{oa \in OA} \text{Roles}_{oa} \qquad [7.6.b]$$

(3) Rôle-dépendance pour l'exécution d'un plan

En ce qui concerne les plans, le calcul de dépendance d'un rôle est un peu différent. Un agent peut recevoir d'un rôle le savoir-faire d'un plan et/ou la permission de l'exécuter. Cependant, même après avoir reçu ceci, il est possible que l'agent ne soit toujours pas capable d'exécuter le plan tout seul (le pouvoir individuel pour un plan nécessite le pouvoir pour chaque élément, chaque pré-condition, etc., voir la section précédente). Nous considérons *qu'un agent dépend d'un rôle pour l'exécution d'un plan* s'il reçoit de la part du rôle le savoir-faire du plan et la permission de l'exécuter ou bien s'il dépend du rôle pour un élément du plan ou une pré-condition d'un élément :

$$\begin{aligned} role\_depends(x,R,pl) \equiv\ & pl \in \text{Pl}_{x+R} - \text{Pl}_{x-R} \wedge pl \in \text{Perm}_{x+R} - \text{Perm}_{x-R} \\ & \vee \exists \alpha \in pl\ role\_depends(x,R,\alpha) \\ & \vee \forall \alpha \in pl\ \exists \alpha' \in \text{T}\ enables(\alpha',\alpha) \wedge role\_depends(x,R,\alpha') \end{aligned} \qquad [7.6.c]$$
$$\text{où } x \in Ag, pl \in Pl,\ R \in \bigcup_{oa \in OA} \text{Roles}_{oa}$$

(4) Rôle-dépendance pour la satisfaction d'un but

Etant donné un de ses buts, un agent peut manquer le pouvoir individuel pour le satisfaire parce qu'il n'a pas la permission de le poursuivre, parce qu'il ne connaît aucun plan qui le satisfait ou parce qu'il n'a pas le pouvoir individuel pour un tel plan. Un agent peut ainsi recevoir d'un rôle le savoir-faire d'un plan qui satisfait son but et/ou la permission de poursuivre celui-ci. Cependant, même après avoir reçu ceci, il est possible que l'agent ne soit toujours pas capable d'exécuter le plan tout seul (voir ci-dessus). Nous considérons qu'un *agent dépend d'un rôle pour la satisfaction d'un but* s'il reçoit de la part du rôle la permission de le poursuivre ou s'il reçoit le savoir-faire d'un plan qui satisfait le but (pour lequell il n'avait pas ce savoir-faire) ou bien s'il dépend du rôle pour tout plan qui satisfait le but:

$$\begin{aligned} role\_depends(x,R,g) \equiv\ & g \in \text{Perm}_{x+R} - \text{Perm}_{x-R} \\ & \vee \forall pl \in \text{Pl}_{x-R}\ g \notin pl \wedge \exists pl_g \in \text{Pl}_{x+R}\ g \in pl_g \\ & \vee\ \neg power\_of(x,g) \wedge \exists pl_g \in \text{Pl}_x\ g \in pl_g \wedge role\_depends(x,R,pl_g) \end{aligned} \qquad [7.6.d]$$
$$\text{où } x \in Ag, g \in G,\ R \in \bigcup_{oa \in OA} \text{Roles}_{oa}$$

Les formules ci-dessus définissent les relations de dépendance qui existent entre des agents et les rôles qu'ils peuvent jouer. Comme nous le verrons par la suite de ce chapitre et dans le chapitre suivant, elles ont leur importance dans le raisonnement qu'un agent effectue sur les contraintes institutionnelles qui lui sont imposées.

Exemple 7.2 (cont.)

Si nous considérons l'exemple précédent, dans le plan qui satisfait le but *g* il existe une action *exec_job* qui doit être exécutée. L'agent *x* a le pouvoir individuel pour toutes les actions du plan, sauf pour cette action : il a le pouvoir d'exécution pour elle (il sait l'exécuter), mais il n'a pas le pouvoir déontique (il n'a pas la permission). Considérons maintenant un rôle que nous appelons *RExec* dans une organisation – place de marché – représentée par l'agent *ma*. Une des caractéristiques de ce rôle est qu'il donne aux agents qui le jouent la permission d'exécuter l'action *exec_job*.

Dans le cadre de l'exemple précédent, à cause de ce manque de permission, l'agent *x* dépendait d'un autre agent *z* qui était capable d'exécuter l'action à sa place. Maintenant, l'existence de ce rôle fait que l'agent dépend également du rôle *RExec* parce qu'il peut lui fournir la permission qui lui manque (conforme Formule 7.6.b). Ceci fait que, d'après les Formules 7.6.c, 7.6.d, l'agent dépend de ce rôle pour son but *g* :

$$exec\_job \in \mathrm{Act}_x,\ exec\_job \notin \mathrm{Perm}_x,\ RExec \in \mathrm{Roles}_{ma},\ exec\_job \in \mathrm{Perm}_{RExec} \xrightarrow{cf.\ 7.6b} role\_depends(x, RExec, exec\_job) \xrightarrow{cf.\ 7.6c\ et\ 7.6d} role\_depends(x, RExec, g)$$

7.2.3. Situations de dépendance

Comme nous le verrons dans la partie suivante, nous considérons qu'il est utile pour un agent de calculer les relations de dépendance qui le lient aux autres agents, ce qui peut l'aider dans son raisonnement. Cependant, nous suivons l'approche de Sichman (Sichman 1995) qui envisage également le calcul des *situations de dépendance*, calcul qui nécessite l'identification des relations de dépendance des autres agents. Les relations de dépendance telles que décrites dans les sections précédentes sont assez génériques et pas forcément utiles aux agents en tant que telles. Par exemple, un agent peut calculer qu'il dépend d'un autre agent pour une ressource, mais s'il ne veut pas utiliser la ressource (il n'en n'a pas besoin dans la satisfaction de son but courant), il n'est pas très important de connaître ou non cette relation de dépendance. Une situation de dépendance représente ainsi une (ou plusieurs) relations de dépendance qu'un agent a calculées et dont il est conscient. Nous représentons ceci en utilisant le prédicat *believes* qui dénote les croyances d'un agent sur ses dépendances vis-à-vis d'un autre agent : *believes(agent,depends(agent,autre_agent,tâche))*.

Il existe des situations de dépendance particulièrement intéressantes dans lesquelles un agent peut se trouver et qui peuvent influencer son raisonnement, situations que nous divisons en deux catégories : dépendances mutuelles ou réciproques et dépendances d'une organisation.

(1) Dépendances mutuelles et réciproques

Pour calculer s'il se trouve dans une situation de dépendance mutuelle ou réciproque, un agent doit identifier les relations de dépendance que les autres agents ont envers lui. Il doit donc faire des hypothèses sur les caractéristiques de ces agents, telles que leurs buts, plans, actions, ressources ou permissions, mais aussi sur leurs croyances. Même si dans un contexte général il est difficile pour un agent de faire ce type d'hypothèses correctement, dans notre modèle ce calcul est sensiblement

simplifié par le fait qu'une grande partie de ces caractéristiques des agents proviennent des rôles qu'ils jouent. A partir des spécifications organisationnelles qu'il connaît, un agent peut ainsi déduire une partie des caractéristiques des agents et calculer leurs relations de dépendance. Bien évidemment, comme les connaissances qu'un agent a sur les autres sont généralement incomplètes ou fausses, les situations de dépendances identifiées par un agent peuvent être incomplètes ou fausses. Cependant, comme nous le verrons dans la partie suivante, même avec le risque de se tromper, un agent peut sensiblement améliorer son raisonnement en utilisant ces situations.

La situation de *dépendance réciproque* entre deux agents apparaît quand un agent dépend d'un autre pour un de ses buts et l'autre dépend de lui pour un autre but. C'est-à-dire, les agents peuvent s'aider réciproquement à satisfaire leurs deux buts. Nous notons cette situation de dépendance par le prédicat *rec_depends(agent1,agent2,but_agent1,but_agent2)* et la formule utilisée pour son calcul est présentée ci-dessous. Nous voulons souligner que nous avons explicitement fait apparaître dans cette formule le fait que les agents ont les deux buts respectifs actifs (le fait qu'ils appartiennent à l'ensemble des buts de chaque agent), c.-à-d., ils se retrouvent dans une situation de dépendance réciproque seulement si chacun d'eux poursuit son but. Cette condition que le but appartienne à l'ensemble des buts actifs d'un agent existait déjà dans le calcul d'une relation de dépendance, mais nous avons préféré le mettre en évidence ici :

$$rec\_depends(x,y,g_1,g_2) \equiv believes(x,depends(x,y,g_1)) \wedge believes(y,depends(y,x,g_2)) \quad [7.7]$$
$$\text{où } x,y \in Ag,\ g_1 \in G_x,\ g_2 \in Gy,\ x \neq y,\ g_1 \neq g_2$$

La situation de *dépendance mutuelle* entre deux agents est similaire à la dépendance réciproque, sauf que dans ce cas il s'agit du même but, commun aux deux agents qui tous les deux dépendent l'un de l'autre pour sa satisfaction. C'est-à-dire, les agents peuvent s'aider mutuellement à satisfaire leur but commun. Nous notons cette situation de dépendance par le prédicat *mut_depends(agent1,agent2,but)* et la formule utilisée pour son calcul est présentée ci-dessous :

$$mut\_depends(x,y,g) \equiv believes(x,depends(x,y,g)) \wedge believes(y,depends(y,x,g)) \quad [7.8]$$
$$\text{où } x,y \in Ag,\ g \in G_x \cap Gy,\ x \neq y$$

A la première vue, il peut sembler difficile qu'une telle dépendance existe : un agent dépend d'un autre pour un but parce qu'il ne peut pas le satisfaire tout seul – si l'autre agent n'a pas non plus ce pouvoir, comment peut-il aider le premier agent ? L'exemple ci-dessous répond à cette question : aucun agent n'est capable de satisfaire le but tout seul, mais ils sont capables de le satisfaire ensemble.

Exemple 7.2 (cont.)

Dans le cadre de l'exemple précédent, dans le seul plan qui satisfait le but *neg*, il existe un conflit généré par l'exécution en parallèle de deux actions neg_1 et neg_2. Deux agents *x* et *y* ont les pouvoirs individuels pour ces deux actions, mais aucun n'a le pouvoir pour le but *neg* à cause de ce conflit. Comme nous l'avons argumenté dans l'exemple précédent, l'agent *x* dépend de l'agent *y* pour l'exécution de l'action neg_2 et ainsi pour la satisfaction du but *neg*. De la même manière, l'agent *y* dépend de l'agent *x* pour l'exécution de l'action neg_1 et ainsi pour la satisfaction du but *neg*. Si

nous considérons maintenant que les deux agents poursuivent le but *neg*, alors conforme à la Formule 7.8 ils se retrouvent dans une situation de dépendance mutuelle :

$$depends(x,y,neg) \wedge depends(y,x,neg) \xrightarrow{cf.\ 7.8} mut\_depends(x,y,neg)$$

(2) Dépendance d'une organisation

Bien que, comme nous le verrons dans la partie suivante, il est important pour un agent d'identifier les rôles dont il dépend pour pouvoir choisir lequel jouer, la relation de dépendance d'un agent envers un rôle a également une autre utilité. Si l'agent appartient à une organisation dans laquelle il joue un ou plusieurs rôles, il peut calculer s'il se trouve dans une *situation de dépendance vis-à-vis de l'organisation*. Si l'agent ne peut satisfaire son ou ses buts actif(s) qu'en jouant des rôles dans cette organisation, alors il dépend de l'organisation. Il essaiera ainsi de continuer à appartenir à cette organisation, pour continuer de jouer les rôles et d'être ainsi en mesure de satisfaire ses buts. Nous notons cette situation de dépendance par le prédicat *org_depends(agent,organisational_agent,but)* et la formule utilisée pour son calcul est présentée ci-dessous :

$$org\_depends(x,oa,g) \equiv \exists R \in \text{Roles}_{oa}\ x{:}R \wedge role\_depends(x,R,g), \quad [7.9]$$
$$\text{où } x \in \text{Ag},\ g \in \text{G}_x,\ oa \in \text{OA}$$

Bien évidemment, ce modèle nous permet de représenter des dépendances dans le sens inverse, comme par exemple la dépendance d'une organisation d'un agent qui joue un rôle. Cependant, comme l'utilisation que nous envisageons pour ce modèle est d'enrichir le raisonnement d'un agent, et non celui d'une organisation, nous ne représentons pas dans cette thèse ce type de dépendance. Dans les sections suivantes nous montrons l'importance des relations de dépendance d'un agent dans le calcul des notions plus génériques, comme le pouvoir indirect ou le pouvoir social.

7.3. Pouvoir indirect et pouvoir social

Malgré le fait que les types de pouvoir introduits dans les sections précédentes sont des pouvoirs individuels (qui appartiennent à un seul agent), la notion de pouvoir qui nous intéresse est une notion *sociale.* Autrement dit, nous nous intéressons aux relations de pouvoir qui existent entre les agents et aux pouvoirs qu'un agent a par rapport à d'autres. Les relations de dépendances décrites dans la section précédente constituent un premier pas vers ces pouvoirs sociaux : elles décrivent la relation entre deux agents générée par le pouvoir d'un et le manque de pouvoir de l'autre. Comme nous l'avons argumenté dans le Chapitre 3, un agent se retrouvant dans une situation de dépendance est généralement mené à interagir afin d'acquérir le pouvoir qui lui manque. Nous décrivons par la suite des relations de pouvoir qui étendent les relations de dépendance et le pouvoir indirect qu'un agent peut acquérir par l'intermèdiaire d'autres agents.

7.3.1. Relations de pouvoir social interpersonnel

Après avoir décrit les pouvoirs qu'un agent a individuellement, nous nous intéressons maintenant aux relations de pouvoirs qui peuvent être envisagées entre les agents. Parmi ces relations décrites dans le Chapitre 3, la relation de pouvoir social qui nous intéresse dans ce manuscrit est *le pouvoir qu'un*

agent a sur un autre (*power over*, en anglais). Comme nous le verrons par la suite, cette relation de pouvoir nous permet de réunir des aspects à la fois interpersonnels et institutionnels et elle est directement liée à la notion d'autonomie.

Dans un cadre purement interpersonnel, c.-à-d., sans considérer de mécanismes qui génèrent des contraintes institutionnelles, un agent acquiert du pouvoir sur un autre suite à une relation de dépendance entre cet agent et lui-même. Cette relation de pouvoir, notée par le prédicat *dep_power_over(agent,agent,tâche)*, (le préfixe dep fait référence à la source de ce pouvoir qui est liée à une relation de dépendance) décrit les influences qu'un agent subit dans sa prise de décision concernant une tâche, influences issues du fait qu'il dépend d'un autre agent pour la réalisation de cette tâche. Si un agent n'a pas le pouvoir individuel pour un de ses buts, alors il ne peut pas satisfaire son but tout seul et il doit chercher l'aide d'un autre agent qui a le pouvoir qui lui manque. Il dépend ainsi de cet agent pour la satisfaction de ce but, ce qui constitue une contrainte qui limite sa prise de décision : il ne peut plus décider librement comment et quand satisfaire son but, il est influencé en ceci par l'autre agent. Nous disons ainsi que l'autre agent acquiert un pouvoir sur lui : il peut influencer sa prise de décision concernant la satisfaction du but :

$$dep\_power\_over(y,x,\lambda) \equiv believes(x, depends(x,y,\lambda))$$
$$\text{où } x,y \in \mathrm{Ag}, \lambda \in \mathrm{T} \qquad [7.10]$$

Comme argumenté par Castelfranchi (Castelfranchi 2002), un élément très important apparaît dans la définition de ce pouvoir social : un agent est conscient de sa dépendance envers un autre (prédicat believes dans la définition 7.10). Si un agent ne croit pas qu'il dépend d'un autre agent, alors l'autre ne peut pas vraiment influencer sa prise de décision. Nous voulons souligner que les croyances d'un agent peuvent être fausses. Ainsi un agent peut croire qu'il dépend d'un autre et lui accorder ainsi du pouvoir sur lui, même si cette dépendance n'est pas vraiment réelle. L'exemple de Castelfranchi concernant le voyou qui menace avec un pistolet en caoutchouc est intéressant à ce titre : l'agent menacé croit qu'il dépend du voyou pour rester en vie, ce qui donne au voyou du pouvoir sur lui (il peut influencer ses décisions), même si, le pistolet n'étant pas réel, cette dépendance est basé sur une information fausse.

Comme nous l'avons précisé dans les sections précédentes, de nombreuses relations et situations de dépendance peuvent exister au sein d'un système. Nous pouvons ainsi parler de dépendances réciproques ou mutuelles. Toutes ces dépendances peuvent générer des relations de pouvoir social entre les agents : avant de discuter les pouvoirs réciproques ou mutuels, nous allons nous intéresser à d'autres sources pour les pouvoirs d'un agent sur un autre.

7.3.2. Relations de pouvoir social institutionnel

Les dépendances existantes entre les agents constituent une des sources de la relation de pouvoir d'un agent sur un autre agent, source qui est par définition (nature de la relation de dépendance) interpersonnelle. Cependant, la relation de pouvoir social peut avoir aussi des sources institutionnelles, c.-à-d., elle peut apparaître suite aux relations qui existent entre agents dans un contexte institutionnel. Ces sources sont représentées par la situation de dépendance d'un agent vers son organisation (voir la

Section 7.2.3), les obligations de l'agent ou les relations d'autorité au sein d'une organisation. Par la suite nous présentons les pouvoirs sociaux issus de ces trois sources.

(1) Pouvoir social issu de la dépendance d'une organisation

Comme nous l'avons précisé ci-dessus, un agent peut dépendre d'un rôle. Ceci signifie qu'il ne peut effectuer la tâche concernée par cette dépendance qu'en jouant ce rôle (qui lui donne des permissions, ressources ou même des connaissances). Ainsi, si un agent joue un rôle dans une organisation et s'il dépend de ce rôle, alors il se retrouve dans une situation de dépendance envers l'organisation. Il ne peut pas satisfaire un de ses buts s'il quitte l'organisation, ce qui constitue une contrainte sur sa prise de décision concernant la satisfaction de ce but. Il est influencé dans cette prise de décision par l'organisation (ou l'agent qui la représente), parce que cette prise de décision n'est plus indépendante de ce que l'organisation veut. Nous disons ainsi que l'organisation a du pouvoir sur l'agent pour la satisfaction de son but, noté par le prédicat *org_power_over(organisational_agent,agent,but)*, si un agent dépend de l'organisation pour satisfaire le but :

$$org\_power\_over(oa,x,g) \equiv org\_depends(x,oa,g)$$
$$\text{où } x \in \text{Ag}, oa \in \text{OA}, g \in \text{G} \qquad [7.11]$$

Nous voulons souligner que dans cette formule, comme d'ailleurs dans la Formule 7.9 définissant la dépendance d'une organisation, nous ne faisons pas apparaître la croyance d'un agent d'une relation de dépendance. Cette différence par rapport à la Formule 7.10 est liée à la nature différente des relations de dépendance considérées. Dans le cadre des relations interpersonnelles, la croyance d'un agent est importante : ces relations peuvent être générées par toute caractéristique d'un agent et généralement les agents ne sont pas conscients de toutes leurs relations de dépendances. Dans le cadre institutionnel, tous les facteurs utilisés dans le calcul des relations de dépendance envers une organisation sont connus par l'agent : ses caractéristiques, les caractéristiques du rôle qu'il joue et le fait qu'il le joue. Ceci fait que dans notre modèle nous considérons qu'un agent sait toujours s'il dépend ou non de son organisation pour un but, donc si l'organisation a ou non du pouvoir sur lui.

(2) Pouvoir basé sur des obligations

Comme nous l'avons présenté dans le Chapitre 6, quand un agent joue un rôle dans une organisation, il est le sujet de plusieurs contraintes que nous représentons à l'aide des engagements sociaux et des politiques sociales. Si les engagements sociaux décrivent les tâches qu'un agent doit effectuer, les politiques sociales décrivent les conditions dans lesquelles un agent doit s'engager d'effectuer telle ou telle tâche. Autrement dit, les politiques sociales associées à un rôle représentent des contraintes institutionnelles décisionnelles imposées aux agents jouant le rôle, telles que les obligations (souvent contextuelles) des agents. Dans un certain contexte, un agent se retrouve contraint d'adopter une tâche (de s'engager envers l'organisation pour la réalisation de cette tâche).

La décision d'un agent concernant un but peut être donc influencée par une organisation à travers les obligations que l'agent a parce qu'il joue des rôles dans l'organisation. Il existe ainsi une relation de pouvoir entre l'organisation et l'agent : le pouvoir que l'organisation a sur un agent pour la satisfaction d'un but, pouvoir issu d'une obligation de l'agent concernant le but et que nous notons par le

prédicat *norm_power_over(organisational_agent,agent,but)*. Quand l'organisation le décide (ou dans un certain contexte), l'agent doit poursuivre le but : il n'est plus libre de décider comme il veut sur la satisfaction de ce but, mais il doit prendre en compte l'organisation :

$$norm\_power\_over(oa,x,g) \equiv obliged\_to(x,oa,g) \quad [7.12]$$
$$\text{où } x \in \text{Ag},\ oa \in \text{OA},\ g \in \text{G}$$

Nous rappelons ici que le prédicat *obliged_to* est utilisé pour noter les obligations exprimées dans notre modèle par des politiques sociales définies dans le Formules 6.3 et 6.4 (voir Chapitre 6). Nous voulons souligner que dans ce cas les croyances de l'agent ne sont pas utilisées non plus, pour la même raison que dans la section précédente. Les spécifications du rôle étant connues du début, dans notre modèle nous considérons qu'un agent sait quand il a une obligation active envers l'organisation, ce qui donne à cette dernière du pouvoir sur l'agent.

(3) Pouvoir basé sur des relations d'autorité

Les obligations ne sont pas les seuls concepts institutionnels représentés à l'aide des politiques sociales. Dans notre modèle, ces politiques représentent également des liens d'autorité entre des agents, liens issus des relations d'autorité qui existent entre les rôles d'une organisation. Le fait de jouer un rôle dans une organisation impose des contraintes sur la prise de décision d'un agent, contraintes issues de ces liens d'autorité. Un agent peut ainsi avoir du pouvoir sur un autre, pouvoir qui est donné par l'autorité qu'il a sur lui et que nous notons par le prédicat *auth_power_over(agent1,agent2,but)*. Quand un agent a de l'autorité sur un autre pour un but, le deuxième agent ne peut plus décider librement concernant le but, ses décisions sont influencées par l'autre agent :

$$auth\_power\_over(x,y,g) \equiv \exists oa \in \text{Ag}\ autority(x,y,g,oa) \quad [7.13]$$
$$\text{où } x,y \in \text{Ag},\ g \in \text{G}$$

Nous rappelons ici que le prédicat *authority* est utilisé pour noter le lien d'autorité dans une organisation, lien exprimé dans notre modèle par des politiques sociales définies dans les Formules 6.5 et 6.6 (voir Chapitre 6). Pour les mêmes raisons qu'auparavant, les croyances des agents ne sont pas utilisées : nous considérons que tout agent connaît les spécifications des rôles qu'il joue, ce qui permet de transformer une relation d'autorité entre deux rôles en une relation de pouvoir d'un agent sur un autre.

7.3.3. Pouvoir social – *power over*

Dans les deux sections précédentes nous avons décrit plusieurs types de pouvoir qu'un agent peut avoir sur un autre. Ce pouvoir représente l'influence qu'un agent peut exercer sur la prise de décision d'un autre agent et peut être lié à une situation de dépendance, interpersonnelle ou envers une organisation, à une obligation ou un lien d'autorité au sein d'une organisation. Même si souvent il est utile pour un agent de connaître la source des contraintes que des entités externes imposent sur sa prise de décision, parfois dans le raisonnement d'un agent il suffit de savoir qu'une telle relation de pouvoir

existe, en ignorant d'où elle provient. Nous notons ainsi par le prédicat *power_over(agent1,agent2,but)* le pouvoir qu'un agent a sur un autre et qui peut être généré par un des quatre types de contrainte décrit dans les Formules 7.10-7.13 :

$$\begin{aligned} power\_over(x,y,g) \equiv\ & dep\_power\_over(x,y,g) \vee auth\_power\_over(x,y,g) \\ & \vee\ x \in \mathrm{OA} \wedge (org\_power\_over(x,y,g) \vee norm\_power\_over(x,y,g)) \end{aligned} \qquad [7.14]$$

où $x,y \in \mathrm{Ag}$, $g \in \mathrm{G}$

Vue cette diversité des sources possibles pour la relation de pouvoir social, un agent peut souvent se retrouver impliqué dans plusieurs relations de pouvoir avec d'autres agents, pour un ou plusieurs buts. Comme nous le verrons dans le chapitre suivant, dans son raisonnement basé sur ces relations, il peut identifier deux situations intéressantes, *le pouvoir réciproque et le pouvoir mutuel que deux agents ont l'un sur l'autre*. La définition de ces deux types de pouvoir (resp. notés par *rec_power_over(...)* et *mut_power_over(...)*) est similaire à la définition des dépendances réciproques et mutuelles (voir Formules 7.7 et 7.8) :

$$\begin{aligned} & rec\_power\_over(x,y,g_1,g_2) \equiv \exists g_1 \in \mathrm{G}\ power\_over(x,y,g_1) \wedge \exists g_2 \in \mathrm{G}\ power\_over(y,x,g_2) \\ & mut\_power\_over(x,y,g) \equiv \exists g \in \mathrm{G}\ power\_over(x,y,g) \wedge power\_over(y,x,g) \\ & \text{où } x,y \in \mathrm{Ag} \end{aligned} \qquad [7.15]$$

7.3.4. Pouvoir indirect – *indirect power of*

Un agent a des pouvoirs individuels pour satisfaire des buts, exécuter des plans ou des actions ou utiliser des ressources, mais il est souvent possible qu'il n'a pas tous les pouvoirs dont il a besoin. Il doit donc chercher l'aide d'un ou de plusieurs agents qui ont les pouvoirs qui lui manquent et qui pourront ainsi effectuer des tâches à sa place – nous verrons dans la partie suivante comment un agent peut chercher les agents qui peuvent l'aider. Si de tels agents acceptent de l'aider, l'agent peut considérer qu'il a acquis de nouveaux pouvoirs : même s'il n'est toujours pas capable de satisfaire son but tout seul, il est néanmoins capable d'avoir ce but satisfait par l'intermédiaire des autres.

Nous appelons *pouvoir indirect* le fait qu'un agent a un pouvoir pour une tâche par le biais d'un autre agent. Nous considérons deux formes de ce pouvoir, le *pouvoir individuel fort* et le *pouvoir individuel faible*. Nous notons le premier par le prédicat *strong_ind_power_of(agent,agent,tâche)*, prédicat défini par le manque de pouvoir du premier agent, le pouvoir individuel du deuxième et le fait que le deuxième a accepté d'aider le premier. Les deux premiers aspects sont représentés par une relation de dépendance qui lie les deux agents, tandis que dans notre modèle, nous représentons le troisième aspect par un engagement social (voir Chapitre 5). Autrement dit, dans notre modèle, en acceptant d'aider un autre pour une tâche, un agent s'engage envers l'autre pour effectuer la tâche :

$$\begin{aligned} strong\_ind\_power\_of(x,y,\lambda) \equiv\ & depends(x,y,\lambda) \wedge \\ & \exists ia \in IA \exists sc = < y,x,ia,\lambda,\beta > \in SC_{ia} \end{aligned} \qquad [7.16]$$

où $x,y \in \mathrm{Ag}, \lambda \in \mathrm{T}, \beta \in \mathrm{EF}$

Autrement dit, si un agent dépend d'un autre pour une tâche et l'autre s'est engagé envers lui pour effectuer cette tâche, alors l'agent a un pouvoir indirect, par l'intermédiaire de l'autre agent, pour avoir la tâche effectuée. Nous voulons souligner que l'état de l'engagement social dans cette définition doit être soit *active* (engagement créé), soit *fulfilled* (engagement rempli). Si l'engagement a été violé ou annulé, alors le premier agent n'a plus le pouvoir indirect fort. Nous appelons ce pouvoir indirect *fort* pour souligner qu'une contrainte forte, comportementale, oriente le comportement du deuxième agent vers la satisfaction d'une tâche pour le premier agent.

Contrairement à ce pouvoir fort, le pouvoir indirect faible, noté par le prédicat *weak_ind_power_of(agent,agent,tâche)*, représente le potentiel d'un agent à avoir une tâche effectuée à sa place par un autre : l'autre ne s'est pas (encore) engagé envers lui pour la tâche, mais il peut être convaincu de le faire. Autrement dit, le premier agent dépend du deuxième pour une tâche et peut influencer sa décision concernant la tâche, donc il a du pouvoir sur lui :

$$weak\_ind\_power\_of(x,y,\lambda) \equiv depends(x,y,\lambda) \wedge \quad power\_over(x,y,\lambda)$$
$$\text{où } x,y \in \mathrm{Ag}, \lambda \in \mathrm{T} \qquad [7.17]$$

Ce pouvoir qu'un agent à sur un autre, pouvoir qui est transformé par cette formule en un pouvoir indirect faible, peut avoir deux sources, comme nous l'avons vu dans les sections précédentes. Un agent peut acquérir du pouvoir sur un autre suite à une dépendance de l'autre envers lui ou suite à un lien d'autorité qui les li. Le deuxième cas est intéressant pour souligner le lien existant entre les deux formes de pouvoir indirect. Quand un agent joue un rôle qui a de l'autorité sur un autre rôle, à partir de la Formule 7.13, il a du pouvoir sur le deuxième agent. S'il dépend de cet agent, conforme à la Formule 7.17 ci-dessus, il a un pouvoir indirecte faibl par l'intermède de l'autre agent. Dans notre modèle, le lien d'autorité est représenté sous la forme d'une politique sociale (voir Formule 6.4) qui dit que le deuxième agent doit s'engager envers le premier quand le premier lui le demande. Si le deuxième ne viole pas cette politique sociale, alors il s'engage suite à une demande, ce qui, selon la Formule 7.16, donne au premier agent un pouvoir indirect fort par l'intermédiaire du second.

Dans cette section nous nous sommes intéressés aux pouvoirs qu'un agent a par rapport à un autre agent : le pouvoir indirect qu'un agent a pour effectuer une tâche par l'intermédiaire d'un autre agent et le pouvoir qu'un agent a sur un autre. Nous voulons souligner que nous avons défini certaines des formules précédentes pour représenter le pouvoir concernant des buts : cependant, elles utilisent des concepts définis pour tout type de tâche (action, plan ou but) et peuvent donc être facilement généralisées pour décrire le pouvoir d'un agent sur un autre pour la réalisation d'une tâche. Ces pouvoirs complètent notre modèle de représentation des contraintes interpersonnelles et institutionnelles qui limitent la prise de décision ou le comportement des agents. Par la suite nous décrivons comment ces contraintes peuvent être utilisées pour définir formellement l'autonomie sociale.

7.4. Autonomie d'un agent

Comme nous l'avons précisé dans le Chapitre 1, l'*autonomie* est souvent considérée comme la propriété d'un agent de prendre des décisions indépendamment des autres (non-influencées par eux). Cependant, un point de vue légèrement différent peut être pris quand un agent est appelé *autonome* parce qu'il refuse une demande d'adoption ou désobéit une norme. Dans cette section nous mettons en évidence les différences entre ces deux perspectives sur l'autonomie en les définissant formellement à l'aide des notions introduites auparavant.

7.4.1. Capacité d'autonomie d'un agent – *has autonomy*

Pour définir le concept de l'autonomie, nous partons d'une propriété communément admise : l'autonomie est une notion relationnelle. Un agent a de l'autonomie par rapport à une autre entité et pour quelque chose (l'objet de l'autonomie). D'une manière informelle, avoir de l'autonomie sociale de décision, par rapport à un autre agent, pour un but, signifie pour un agent que l'autre agent ne peut pas influencer ses décisions concernant la poursuite ou la satisfaction du but. Un agent avec de l'autonomie décide lui-même, sans pouvoir être contraint par une entité externe.

Avoir de l'autonomie signifie donc une liberté de décision, une absence de contraintes imposées sur le processus de décision. Afin de définir ce concept, nous avons besoin d'identifier ces contraintes *décisionnelles* possibles. Dans les sections précédentes nous avons construit graduellement un modèle dans lequel des contraintes décisionnelles, interpersonnelles ou institutionnelles, sont représentées à l'aide d'un unique concept : le pouvoir qu'un agent a sur un autre. La signification de ce pouvoir sur un autre agent est de pouvoir influencer sa prise de décision. Ceci constitue exactement le contraire de ce qu'avoir de l'autonomie signifie : nous notons cette capacité d'autonomie d'un agent par le prédicat *has_autonomy(agent1,agent2,tâche)* et nous le définissons ainsi :

$$has\_autonomy(x, y, \lambda) \equiv \neg power\_over(y, x, \lambda) \qquad [7.18]$$
$$\text{où } x, y \in \mathrm{Ag},\ \lambda \in \mathrm{T}$$

La relation du pouvoir sur un autre nous permet de regrouper dans un seul concept l'effet de différents types de contraintes imposées sur le comportement d'un agent, contraintes issues de ses dépendances ou du rôle qu'il joue. Ces contraintes font que l'agent ne soit plus libre dans ses décisions, mais influençable par un autre agent – l'agent n'a pas de l'autonomie pour prendre des décisions concernant la tâche. Quand un agent dépend d'un autre pour un de ses buts, il n'a pas d'autonomie par rapport à l'autre pour prendre des décisions de poursuivre ou satisfaire ce but. Quand un agent joue un rôle dans une organisation, sa décision peut être influencée par l'autorité qu'un autre a sur lui ou par le fait qu'il doit obéir l'organisation : quand un autre agent ou l'organisation le décide, il doit poursuivre le but.

Si nous considérons les différentes approches concernant l'autonomie sociale des agents, approches présentées dans le Chapitre 1, il est évident que notre définition de l'autonomie est compatible avec celles-ci. Par exemple, l'autonomie est souvent considérée comme une indépendance – dans notre modèle une dépendance implique une relation de pouvoir social, donc un manque d'autonomie. Un autre exemple définit l'autonomie d'un agent par sa capacité de refuser une délégation d'un but. Dans notre modèle, la délégation/adoption des buts sont représentées comme des opérations de demande et

de création d'un engagement social et la relation d'autorité telle que définie dans le Chapitre précédent dénote qu'une demande (délégation) doit être suivie par une création (adoption). Conformes à nos définitions, cette relation d'autorité implique une relation de pouvoir et donc un manque d'autonomie : si un agent doit accepter une délégation et ne peut pas refuser, alors il n'a pas d'autonomie.

7.4.2. Comportement autonome d'un agent – *is autonomous*

Un autre point de vue existant sur l'autonomie sociale considère ce concept du point de vue des contraintes qui limitent son comportement. Quand un agent a un comportement qui n'obéit pas une telle contrainte, il est considéré comme étant *autonome*. Les contraintes comportementales peuvent être issues des interactions entre les agents (interpersonnelles), mais aussi de l'appartenance de l'agent à une organisation (institutionnelles). Généralement, dans la littérature la première catégorie des contraintes est représentée par des engagements sociaux – dans notre modèle nous avons basé la représentation de la deuxième catégorie sur le même paradigme. Nous parlons ainsi des engagements d'un agent quand il joue un rôle : les buts qu'il s'est engagé de satisfaire et les méta-engagements (politiques sociales) qui limitent son autonomie (obligations et relations d'autorité – voir la section précédente). Nous représentons également le fait qu'un agent joue un rôle par un engagement social.

Si un agent est considéré autonome quand il désobéit une contrainte comportementale et si dans notre modèle une telle contrainte est toujours représenté par un engagement social (normal, politique sociale ou de rôle), la définition d'un comportement autonome d'un agent devient triviale. Nous notons le fait qu'un agent soit autonome par rapport à un autre agent pour la réalisation d'une tâche par le prédicat *is_autonomous(agent1,agent2,tâche)* et nous le définissons par le fait qu'il viole son engagement pris concernant la réalisation de la tâche :

$$
\begin{aligned}
&is\_autonomous(x,y,\lambda) \equiv \exists ia \in \mathrm{IA}\, \exists sc =< x,y,ia,done(\lambda),\beta > \in (SC_{ia} \cap VioSC) \\
&is\_autonomous(x,oa,R) \equiv x:R \wedge \exists ia \in \mathrm{IA}\, \exists \alpha \in EF,\ \exists sc =< x,oa,ia,\alpha,x:R >, \\
&\qquad\qquad sc \in (SC_{ia} \cap VioSC \cap RGComm_r \cap ROPol_r \cap RAPol_r) \\
&\text{où } x,y \in \mathrm{Ag}, oa \in \mathrm{OA},\ R \in \mathrm{Roles}_{oa},\ \lambda \in \mathrm{T},\ \beta \in \mathrm{T}
\end{aligned}
\qquad [7.19]
$$

En fonction du type d'engagements social violé, il est possible d'identifier deux types de comportement autonome d'un agent. Quand un agent viole un engagement qu'il avait envers un autre agent pour la réalisation d'une tâche, il est considéré autonome par rapport à cet agent pour cette tâche – autonomie interpersonnelle. Quand un agent viole un engagement social ou une politique sociale qu'il avait envers une organisation, alors il est considéré autonome par rapport à l'organisation – autonomie institutionnelle. Ceci est appelé dans la littérature autonomie par rapport aux normes – désobéir une norme (ce qui dans notre modèle représente la violation d'une politique sociale qui la représente). Cependant, il n'y a pas seulement les normes (obligations) qui peuvent être désobéies, mais aussi les relations d'autorité ou les buts du rôle qu'un agent joue.

Discussion

Les deux perspectives sur l'autonomie que nous utilisons et que nous avons appelé *avoir de l'autonomie* et *être autonome* font référence à des concepts différents. La première décrit les (le manque de) contraintes qui limitent la prise de décision d'un agent, contraintes issues des autres agents ou d'une organisation. La deuxième décrit la réponse donnée par un agent à une contrainte établie,

existante, qui limite son comportement. Pour mieux analyser la différence entre les deux, prenons comme exemple le cas d'une norme – une obligation dépendante de contexte.

Comme nous l'avons discuté dans le Chapitre 2, une norme représente à la fois une contrainte comportementale et décisionnelle. Une norme représente une contrainte claire, établie : quand un agent a accepté de jouer un rôle, il a accepté d'obéir à cette norme, c.-à-d., d'avoir un comportement qui suit ses spécifications. Cependant, la norme n'est pas toujours active, elle ne doit être obéie que dans un certain contexte. Quand l'agent se trouve dans ce contexte, alors il doit décider de suivre ou non la norme. Dans notre modèle, une telle norme est représentée comme une politique sociale de l'agent envers l'organisation, politique qui spécifie que si une condition (contexte) est vraie, alors l'agent s'engage pour une tâche (s'engage d'avoir le comportement spécifié par la norme). L'existence de cette politique sociale fait en sort que l'agent n'a pas d'autonomie par rapport à l'organisation : sa décision dans le contexte de la norme est influencée par l'organisation. Ceci ne dit rien sur le comportement de l'agent dans ce contexte : il est possible qu'il suit la contrainte et s'engage, mais il est également possible qu'il viole la politique sociale (la norme) – dans ce cas il est considéré autonome.

Pour résumer, le fait qu'un agent n'a pas d'autonomie représente une contrainte sur sa prise de décision, il est influencé par une entité externe. Cependant, cette influence n'est pas égale à un contrôle absolu : il peut toujours la désobéir - s'il la désobéit, il est considéré autonome. L'autonomie telle que nous l'avons définie peut ainsi représenter d'un coté les influences exercées sur la prise de décision d'un agent par des entités externes, mais aussi la réponse de l'agent à ces influences.

7.5. Synthèse

Dans ce chapitre nous avons défini formellement plusieurs concepts utilisés par les théories de la dépendance et du pouvoir social, concepts que nous utilisons pour représenter les contraintes qui limitent la prise de décision et le comportement des agents. Premièrement nous avons défini les pouvoirs individuels d'un agent, ce qu'il peut faire et ce qu'il a le droit de faire. Un agent peut calculer ces pouvoirs pour identifier les buts qu'il peut satisfaire tout seul et les buts pour lesquels il a besoin d'aide. S'il ne peut pas satisfaire seul un but, il peut calculer les pouvoirs des autres agents pour identifier des relations de dépendance envers eux. Dans la partie suivante nous montrerons comment l'identification des situations de dépendance dans lesquelles il se trouve peut aider un agent dans son raisonnement pour choisir avec qui interagir. Si une interaction est conclue avec succès, un agent peut acquérir un pouvoir indirect pour une tâche par l'intermède d'un autre agent qui s'est engagé envers lui pour effectuer la tâche.

L'existence des relations de dépendance envers un autre agent ou un rôle et le fait qu'il soit conscient de ces relations crée des relations de pouvoir entre les agents. Nous parlons ainsi du pouvoir qu'un agent a sur un autre concernant un but : le fait qu'il peut influencer la décision de l'autre sur la poursuite ou satisfaction du but. Cette relation de pouvoir, issue des dépendances interpersonnelles ou des concepts institutionnels tels que lien d'autorité ou norme, est directement liée à la notion d'autonomie. Nous avons ainsi défini le fait qu'un agent a de l'autonomie par le manque de pouvoir d'un autre agent sur lui ; nous avons également défini le fait qu'un agent est dans un comportement autonome par le fait qu'il viole ses engagements qui décrivent le comportement qu'il doit manifester.

Il nous semble intéressant de souligner que les concepts que nous avons définis et utilisés forment une sorte de pyramide. A la base restent les concepts de base, tels qu'action, ressource, permission, etc. A partir de ces concepts nous pouvons calculer des pouvoirs individuels et la présence et absence de ces pouvoirs génèrent des relations de dépendance. Nous généralisons encore plus et nous définissons les relations de pouvoir sur d'autres, relations qui uniformisent les sources de pouvoir, qu'elles soient interpersonnelles (dépendances) ou institutionnelles (relations d'autorité, obligations, etc.). Finalement, au sommet de la pyramide, nous nous retrouvons avec un seul concept, celui d'autonomie. Cette représentation uniforme à l'aide d'un seul concept de tout type de contrainte à un inconvénient : nous perdons de vue la source qui a provoqué le manque d'autonomie. Dans notre modèle nous exprimons par exemple le fait qu'un agent n'a pas d'autonomie par rapport à un autre pour la décision de poursuivre un but, mais nous ignorons (dans la formule finale) d'où provient ce manque d'autonomie. Cependant, cet inconvénient ne nous gène pas dans cette thèse, notre objectif est de permettre aux agents d'évaluer des différentes situations en terme d'autonomie (et pouvoirs) gagnée ou perdue et la source de cet autonomie n'est pas rélévante.

Avoir de l'autonomie décrit les contraintes qui limitent la prise de décision d'un agent, tandis qu'être autonome décrit la réponse de l'agent à de telles contraintes. Si dans les deux premiers chapitres de cette partie nous avons présenté comment notre modèle représente les agents, les interactions entre eux et les organisations, dans ce chapitre nous avons décrit la représentation que nous proposons pour les différents types de contraintes possibles. Cette représentation passe par le paradigme d'engagements sociaux, utilise des concepts de dépendance et pouvoir social, pour tout résumer à l'aide du concept d'autonomie. *Nous arrivons ainsi à un modèle dans lequel il est possible de calculer qui et comment influence la prise de décision des agents, de calculer leurs pouvoirs individuels ou par rapports à d'autres agents et de qualifier leur comportement à travers la notion d'autonomie.*

Synthèse

Dans cette partie nous avons décrit un modèle que nous proposons pour la représentation des diverses contraintes qui limitent la prise de décision et le comportement d'un agent, contraintes qui ont été utilisées pour définir le concept d'autonomie sociale. Comme nous l'avons précisé, nous basons notre modélisation de ces contraintes sur plusieurs théories, modèles ou paradigmes sociaux, notamment les engagements et politiques sociaux, les relations de dépendance et les pouvoirs individuels et sociaux des agents. Nous avons ainsi commencé par une description des éléments de base du raisonnement et comportement d'un agent dans notre modèle, tels que les actions, ressources, plans ou buts. Nous nous sommes ensuite concentrés sur la modélisation des interactions dans notre modèle : le concept d'engagement social à été utilisé. Les aspects liés à l'interaction qui nous intéressent ici sont la délégation d'une tâche d'un agent envers un autre – modélisée sous la forme d'une opération de demande de création d'un engagement social – et l'éventuelle adoption de cette tâche par l'autre agent – un engagement social est créé.

Dans notre modèle, les engagements sociaux représentent des contraintes qui limitent le comportement des agents (un agent engagé envers un autre pour satisfaire un but est censé d'avoir un comportement orienté vers la satisfaction de ce but). Ces contraintes (le modèle d'engagements sociaux) sont par leur nature des contraintes interpersonnelles – issues des interactions entre agents. Cependant, nous avons montré que le même modèle peut être utilisé pour la représentation des contraintes institutionnelles, qui limitent le comportement ou la prise de décision d'un agent qui joue un rôle. Nous représentons ainsi des normes sous la forme des politiques sociales (méta-engagements) et des rôles comme des collections d'engagements et politiques sociaux. Nous représentons également le fait qu'un agent joue un rôle à travers le même modèle : un agent qui joue un rôle devient engagé envers l'organisation pour joue le rôle et ceci fait qu'il est aussi engagé envers l'organisation pour obéir des normes ou des liens d'autorité, etc. Finalement, nous avons aussi décrit comment un agent représente le fonctionnement d'une organisation qui utilise ce modèle, notamment comment il considère qu'elle assure que des agents potentiellement autonomes obéissent les spécifications de leurs rôles.

Après avoir décrit la représentation des interactions et des concepts organisationnels dans notre modèle – en utilisant tout le temps le même paradigme des engagements sociaux, nous nous sommes intéressés à la représentation des contraintes qui limitent la prise de décision des agents. A travers les théories du pouvoir social et de la dépendance, nous avons défini les pouvoirs individuels qu'un agent peut avoir pour ses buts (plans, actions, etc.) : le pouvoir d'exécution et le pouvoir déontique. Un agent peut ainsi calculer quels sont les buts qu'il peut satisfaire sans l'aide des autres agents. Cependant, en général ces buts sont limités et un agent dépend d'autres pour la satisfaction de ses buts : nous avons montré comment des différentes relations de dépendances peuvent apparaître. Un agent peut identifier de telles relations existantes et peut ainsi calculer des relations de pouvoir social entre lui et d'autres agents : nous nous intéressons spécialement au pouvoir d'un agent sur un autre. Ce pouvoir peut être issu d'une situation de dépendance, mais aussi d'une relation d'autorité ou d'une obligation d'un agent et il signifie qu'un agent peut influencer la prise de décision d'un autre.

Nous avons ainsi obtenu un modèle qui représente des contraintes comportementales interpersonnelles et institutionnelles sous la forme des engagements sociaux et des contraintes décisionnelles sous la forme des relations de dépendance ou de pouvoir social. Comme nous l'avons précisé dans le Chapitre 1, ces deux types de contraintes sont directement liés aux deux perspectives sur l'autonomie présentes dans la littérature. Nous avons donc pu définir formellement le fait qu'un agent a de l'autonomie par rapport à une autre entité pour la réalisation d'une tâche par le fait que cette entité n'a pas du pouvoir sur lui pour la tâche – sa décision concernant la tâche ne peut pas être influencée par cette entité. Nous avons aussi défini le comportement autonome d'un agent comme le comportement qui viole un de ses engagements ou politiques sociaux – une de ses contraintes comportementales.

Nous pouvons ainsi imaginer que le modèle proposé ressemble à une pyramide. A la base se trouvent des concepts de bas niveau, comme action, plan, permission ou tâche. A ces concepts nous avons rajouté celui d'engagement social qui porte sur l'exécution des tâches. Ce concept peut être utilisé à un niveau supérieur (méta-engagements) pour représenter des concepts institutionnels. Des pouvoirs individuels d'un agent sont calculés à partir des notions de base et des réseaux de dépendance à partir de ces pouvoirs individuels. Le concept de relation de pouvoir sur un autre agent vient uniformiser à la fois les concepts institutionnels et les réseaux de dépendance, pour trouver au sommet de la pyramide l'autonomie sociale en décision d'un agent. Cette approche "pyramidale" qui cherche à uniformiser les différents types de contraintes qui limitent la prise de décision et le comportement des agents est intentionnelle et dans la partie suivante ses avantages deviendront plus claires.

La représentation de toutes ces contraintes dans notre modèle nous permet de définir formellement le concept d'autonomie, ce qui représente une contribution importante de ce travail dans la communauté scientifique multi-agents. Cependant, obtenir une définition formelle de l'autonomie ou représenter d'une manière uniforme les contraintes qui limitent la prise de décision et le comportement des agents ne constitue qu'un sous-objectif de ce travail. L'objectif principal est d'enrichir le raisonnement des agents dans certaines situations et cet enrichissement sera fait en utilisant le modèle proposé dans cette partie. Nous voulons souligner que ce modèle a été proposé avec ce but, celui d'être utilisé dans le raisonnement d'un agent – il s'agit donc d'un modèle interne à l'agent. Autrement dit, ce modèle décrit comment un agent se représente des différents concepts (limitations de prise de décision ou de comportement) – la partie suivante décrit pourquoi il se les représente.

III. Illustration et mise en œuvre du modèle

Introduction

Dans la partie précédente nous avons défini un modèle formel qui, à travers des concepts tels que les engagements sociaux, les politiques sociales, les pouvoirs individuels des agents et les relations de pouvoir social entre agents, nous a permis de proposer une définition formelle du concept d'autonomie sociale. Comme ce concept n'a pas de définition communément admise dans la communauté scientifique multi-agents, cette formalisation du concept d'autonomie représentait un des objectifs de cette thèse. Cependant, ceci reste un objectif secondaire du notre travail et représente le moyen utilisé pour attendre l'objectif principal, celui d'enrichir le raisonnement d'agents s'exécutant au sein d'un système ouvert et dynamique. Pour des raisons qui deviendront plus claires dans cette partie, nous avons choisi de baser ce raisonnement sur les contraintes qui limitent la prise de décision ou le comportement des agents, contraintes qui restent à la base de notre définition de l'autonomie.

Dans le Chapitre 4 nous avons analysé plusieurs modèles et architectures d'agent proposés dans la littérature afin d'identifier quatre types de raisonnement : *a priori*, pendant, *a posteriori* et planification. Le raisonnement *a priori* est le raisonnement effectué par un agent avant de choisir une direction d'action (p.ex.: de négocier avec un autre agent), suivi par le raisonnement effectué *pendant* la mise en œuvre de la direction d'action choisie (la négociation). Le raisonnement *a posteriori* représente une analyse du résultat de l'exécution de la direction d'action choisie (p.ex.: un contrat négocié). Il peut être suivi par une *planification* de mener jusqu'au bout ceci (remplir le contrat). Notre objectif est de proposer une approche qui permettra aux agents d'effectuer les deux raisonnements les moins analysés dans les travaux similaires, notamment le raisonnement *a priori* et *a posteriori*

Le raisonnement *a posteriori* représente ainsi le raisonnement effectué par un agent dans le cas où son comportement est contraint, c.-à-d., dans le cas où il existe une ou plusieurs contraintes sur le comportement de l'agent, contraintes visant à le pousser à suivre ce comportement. Dans l'exemple précédent, une telle contrainte est représentée par un contrat négocié par un agent, suite à quoi l'agent est censé avoir un comportement qui remplit sa partie du contrat. Dans le modèle proposé dans la partie précédente, nous avons représenté ce type de contraintes sur le comportement d'un agent à l'aide des engagements sociaux et nous avons lié le concept d'être autonome à la (non-)satisfaction de ces contraintes.

Le raisonnement *a priori* représente le raisonnement effectué par un agent sur les conséquences de ses futurs choix. Dans l'exemple précédent, ce raisonnement est utilisé par un agent pour détecter son besoin de négocier avec d'autres agents et il guide éventuellement le choix d'un partenaire de négociation. Ce raisonnement évalue ainsi plusieurs cours d'action possibles et permet à un agent de choisir celui qui correspond le plus à ses motivations (désirs, buts, etc.). L'agent choisira ainsi le cours d'action qui lui permettra de satisfaire le plus ses motivations. Cependant, cette évaluation de la meilleure direction possible est difficile : l'agent raisonne sur des possibilités et non sur des éléments concrets Par exemple, choisir un agent comme partenaire de négociation ne garanti pas un bon résultat de la négociation, le choix de l'agent sera orienté envers l'agent qui a le plus de potentiel pour offrir un bon résultat. Autrement dit, le raisonnement *a priori* évalue le potentiel de situations possibles dans lesquelles un agent peut se retrouver afin de choisir le cours d'action qui le mettra dans celle qui

correspond le plus à ses motivations. Nous avons argumenté dans le Chapitre 4 que cette évaluation de potentielle est faite sous la forme de liberté de prise de décision : plus un agent est libre à prendre les décisions comme il veut, plus il aura des chances à satisfaire ses motivations. Le modèle proposé dans la partie précédente nous a permis de lier le concept d'avoir de l'autonomie aux contraintes qui limitent la prise de décision d'un agent.

Ce bref passage en revue de deux types de raisonnement qui nous intéressent dans ce travail nous permet de souligner que, dans notre approche, ces raisonnements peuvent se baser sur la représentation de l'autonomie que nous avons proposée. Nous détaillerons ainsi dans cette partie ce point de vue. Bien évidemment, d'autres approches sont envisageables, comme le prouve la diversité des modèles de raisonnement rencontrés dans la littérature (cf. Chapitre 4). Cependant, ces raisonnements constituent une bonne illustration de la mise en œuvre de notre modèle et, comme nous le verrons par la suite, l'utilisation de ce modèle permet d'uniformiser des aspects de ces raisonnements.

Dans le chapitre suivant nous analyserons plus attentivement les deux types de raisonnement mentionnés, en prenant en compte des aspects interpersonnels (p.ex. interactions entre agents) et institutionnels (p.ex. organisations d'agents). Nous mettrons en évidence comment un modèle d'autonomie sociale tel que le notre rend ces raisonnements possibles. Dans le Chapitre 9 nous nous intéresserons spécialement au raisonnement *a priori* institutionnel, comme par exemple le choix d'un rôle à jouer dans une organisation. La raison pour ce cas particulier est donnée par la richesse des diverses dimensions de ce raisonnement et le manque des travaux similaires dans la littérature (cf. Chapitre 4). Le dernier chapitre de cette partie présentera l'implémentation effectuée pour valider nos propos. Fidèles au cadre général de cette thèse, nous nous sommes intéressés à une application dans le domaine de l'intelligence ambiante. L'implémentation de cette application à mis en évidence la possibilité pour les agents de choisir entre deux rôles à jouer dans une organisation et d'illustrer ainsi la nécessité d'un raisonnement *a priori* institutionnel.

8. Raisonnement à base d'autonomie

L'objectif de ce chapitre est de montrer comment la modélisation de la notion d'autonomie que nous avons proposée et décrite dans les chapitres précédents peut enrichir le raisonnement des agents. Cette modélisation utilise des concepts variés tels que les pouvoirs individuels et sociaux des agents, les relations de dépendance ou les engagements et politiques sociales. Pour être plus précis, nous avons lié la notion d'autonomie aux contraintes externes qui limitent la prise de décision et le comportement des agents. Il devient ainsi évident que les raisonnements qui portent sur ces contraintes seront ceux qui bénéficieront le plus du modèle d'autonomie proposé.

Les raisonnements auxquels nous nous intéressons plus particulièrement dans ce manuscrit sont les raisonnements *a priori* et *a posteriori*. Comme nous l'avons argumenté ci-dessus, le premier concerne une évaluation des choix à la disposition de l'agent pour choisir le cours d'action qui résultera dans le moins des contraintes sur la prise de décision d'un agent, donc le plus d'autonomie. Le deuxième type de raisonnement concerne la décision à prendre vis-à-vis du respect de contraintes qui limitent le comportement d'un agent : leur obéir ou non. Selon notre définition, lorsqu'une de ces contraintes n'est pas respectée, l'agent est considéré comme étant autonome. Les raisonnements des agents sont différents si les agents agissent dans un contexte institutionnel (où des structures organisationnelles, rôles, normes, etc. sont présentes) ou dans un contexte purement interpersonnel. Dans ce chapitre nous analysons dans les deux contextes les situations qui impliquent une prise de décision des agents, en montrant où interviennent le raisonnement *a priori* ou le raisonnement *a posteriori*.

Cette analyse sera faite en employant les termes et formules introduits dans notre modèle pour souligner comment ce modèle peut être utilisé dans ces raisonnements. De ce point de vue, ce chapitre constitue une illustration ou une mise en œuvre du modèle proposé dans des différents cas. Cependant, l'analyse des différents types de raisonnement reste indépendante du modèle-même et pourrait être faite différemment, ainsi que d'autres utilisations du modèle peuvent être envisagées aussi. Le raisonnement *a priori* dans un contexte institutionnel représente un cas particulier qui est moins traité dans la littérature. Nous considérons qu'il bénéficiera le plus de l'utilisation du modèle proposé dans cette thèse. Pour cette raison, il ne sera pas présenté dans ce chapitre, mais il fera le sujet du chapitre suivant.

8.1. Raisonnement dans un contexte interpersonnel

Pour simplifier notre présentation, considérons d'abord un agent appartenant à un système multi-agent dans lequel il perçoit et agit sur l'environnement, interagit avec d'autres agents, mais où aucun concept organisationnel (rôle, norme, etc.) n'est présent. Cet agent a des buts qu'il veut satisfaire, en formant des plans pour leur satisfaction et exécutant chaque étape de ce plan. Cependant, il existe souvent des buts que l'agent ne peut pas satisfaire tout seul, d'où le besoin d'interagir avec d'autres agents afin de s'entraider pour la satisfaction de leurs buts.

Un agent doit ainsi identifier les buts qu'il peut satisfaire tout seul et les buts pour lesquels il a besoin d'aide. Pour un but qu'il ne peut pas satisfaire tout seul, il doit choisir un agent qui peut l'aider à le

satisfaire. Il interagit ensuite avec cet agent pour lui déléguer la satisfaction du but ou d'un élément d'un plan satisfaisant le but (il peut lui déléguer un sous-but, une action, etc.). S'il a du succès dans la délégation, il a obtenu l'aide de l'autre agent pour satisfaire son but, si non il doit essayer une autre fois, avec peut être un autre agent. Du point de vue de l'autre agent, qui aide le premier à satisfaire son but, il a une décision à prendre quand le premier agent lui délègue la satisfaction d'un de ses buts. Notamment il doit décider s'il aide ou non cet agent : si oui, il adopte le but qui lui a été délégué. A un moment donné il devra ainsi décider de poursuivre ou non ce but : s'il le poursuit, il le fait comme si c'était son propre but et il utilise le même raisonnement que ci-dessus, s'il ne le poursuit pas, il doit subir les conséquences d'avoir adopté un but sans l'avoir poursuivi.

Le processus décrit ci-dessus inclut à plusieurs reprises des décisions prises par un agent, donc des formes de raisonnement. Nous allons analyser par la suite ce raisonnement à travers notre modèle de pouvoirs sociaux et autonomie.

Buts qu'il peut satisfaire tout seul et buts pour lesquels il dépend d'autres

Un agent peut calculer les *pouvoirs individuels* (cf. Formules 7.1, 7.2 et 7.3) qu'il a pour ses buts – seuls les pouvoirs d'exécution sont considérés dans notre cas, vu que les permissions dans notre modèle sont purement institutionnelles. L'objectif principal d'un calcul de pouvoirs est d'identifier les buts que l'agent peut satisfaire tout seul (p.ex. le but g_1 ci-dessous) et d'aider l'agent à décider quels buts poursuivre. Nous voulons souligner que vue la dynamique probable de l'environnement, les pouvoirs d'un agent évoluent en fonction des ressources de l'environnement en sa possession, des contraintes temporelles sur des tâches, etc.

$$power\_of(x, g_1), \quad \neg power\_of(x, g_2)$$

En général, les pouvoirs individuels ne permettent pas à un agent de satisfaire tous ses buts, comme le cas de l'agent x et du but g_2 ci-dessus. Un agent peut calculer les pouvoirs des autres agents en se basant sur leur description externe ou sur des informations provenant par exemple des rôles qu'ils jouent. Et il peut ainsi identifier d'autres agents qui ont les pouvoirs qu'il manque, donc qui peuvent l'aider à satisfaire ces buts :

$$power\_of(y, g_2), \quad power\_of(z, g_2)$$

Un agent peut calculer les relations de dépendance (en utilisant les Formules 7.4) qui le lient à d'autres agents : en faisant ce calcul et en devenant conscient de ces relations, il se retrouve dans des situations de dépendance (cf. au chapitre précédent). Ce calcul de dépendances est effectué en faisant des hypothèses sur les pouvoirs individuels des autres agents : ces hypothèses peuvent être fausses ou incomplètes et donc l'agent peut croire qu'il se trouve dans une situation de dépendance qui n'existe pas en réalité. Ce raisonnement à base de dépendance permet à un agent d'identifier ce qu'il manque pour satisfaire un de ses buts et qui a le pouvoir vis-à-vis de ce qui lui manque.

$$depends(x, y, g_2), \quad depends(x, z, g_2)$$

Choix de partenaire d'interaction – raisonnement *a priori* interpersonnel

Un agent peut ainsi choisir avec qui interagir pour chercher de l'aide dans la satisfaction d'un but en se basant sur ses situations de dépendance identifiées – ce choix de partenaire d'interaction représente ce que nous appelons dans ce manuscrit un *raisonnement a priori interpersonnel*. Cependant, ce raisonnement peut être amélioré en identifiant d'autres situations de dépendance et relations entre agents : l'agent peut essayer de calculer les dépendances que les autres ont envers lui et il peut ainsi identifier des situations de dépendance réciproque ou mutuelle (cf. Formules 7.7 et 7.8). Il dépend d'un autre agent pour satisfaire un de ses buts, mais l'autre dépend également de lui pour satisfaire un autre ou le même but. Il aura ainsi plus de chances de convaincre l'autre agent de l'aider, en lui proposant son aide en retour. En continuant l'exemple précédent dans lequel l'agent *x* dépend de *y* ou de *z* pour la satisfaction de son but g_2, supposons que *x* ne connaît pas une dépendance de *y* envers lui, mais il connaît la dépendance de *z* pour la satisfaction d'un autre but g_3, ce qui constitue une situation de dépendance réciproque entre les deux agents :

$$depends(x,z,g_2) \wedge depends(z,x,g_3) \xrightarrow{cf.\ 7.7} rec\_depends(x,z,g_2,g_3)$$

Nous voulons souligner que, en utilisant la Formule 7.10, une situation de dépendance d'un agent représente en fait un pouvoir qu'un autre agent a sur lui ce qui ensuite signifie que, selon la Formule 7.18, il n'a pas d'autonomie par rapport à l'autre agent.

$$power\_over(x,z,g_3) \xrightarrow{cf.\ 7.18} \neg has\_autonomy(z,x,g_3)$$
$$power\_over(z,x,g_2) \xrightarrow{cf.\ 7.18} \neg has\_autonomy(x,z,g_2)$$

Du point de vue d'un contexte interpersonnel, ce raisonnement peut être fait directement sur les situations de dépendance existante et non sur les relations de pouvoir ou d'autonomie. Cependant, nous verrons par la suite que l'utilisation des concepts plus génériques nous permet d'envisager le même raisonnement dans d'autres situations, par exemple dans un contexte institutionnel.

Pour résumer, le raisonnement *a priori* interpersonnel d'un agent, c'est-à-dire le choix d'un partenaire d'interaction se résume à utiliser le calcul de dépendances et de relation de pouvoir afin de trouver un autre agent qui pourrait être persuadé d'aider cet agent de satisfaire ce but. Comme nous l'avons défini dans la Formule 7.17, si un agent dépend d'un autre pour la satisfaction d'un de ses buts et s'il a un pouvoir sur cet autre agent, alors il a un pouvoir indirect faible pour la satisfaction de ce but. Autrement dit, il peut éventuellement convaincre l'autre de satisfaire ce but (ou généralement d'effectuer une tâche) à sa place.

Dans l'exemple ci-dessus, le raisonnement *a priori* interpersonnel, donc le choix d'un partenaire d'interaction se base sur l'autonomie que d'autres agents ont ou non par rapport à l'agent *x.* Cet agent dépend de *y* et de *z* pour la satisfaction de ce but : il n'a pas de pouvoir sur *y* , mais il a sur *z*, ce dernier n'a pas d'autonomie par rapport à lui. Il a ainsi plus des chances à avoir une interaction réussie avec *z*.

Délégation / adoption

Convaincre un agent d'effectuer une tâche pour un autre agent signifie une interaction entre les deux agents, interaction qui consiste en une délégation de l'exécution de la tâche par un agent et une

éventuelle adoption de cette tâche par l'autre. La décision du deuxième agent d'adopter ou non l'exécution de la tâche dépend de plusieurs facteurs, comme par exemple si cette tâche lui permet de satisfaire un de ses propres buts. Cependant, si le premier agent l'a choisi en tant que partenaire d'interaction, il l'a fait non seulement parce qu'il croit qu'il peut effectuer la tâche, mais aussi parce qu'il croit qu'il a du pouvoir sur lui. Si le deuxième agent se retrouve dans cette situation, ceci signifie qu'il n'a pas d'autonomie par rapport au premier agent, donc, par définition, qu'il existe une contrainte sur sa prise de décision.

De point de vue de notre modèle, nous ne nous intéressons pas à la prise de décision effective dans cette situation – ce que nous avons appelé un raisonnement pendant l'interaction. Cependant, ce modèle nous permet d'identifier les contraintes externes sur cette prise de décision, contraintes qui seront plus diversifiées dans un contexte institutionnel.

Satisfaction de but adopté

Si l'interaction décrite ci-dessus est conclue avec du succès, le deuxième agent adopte la satisfaction d'un but (ou l'exécution d'une tâche) appartenant à un autre agent. Dans notre modèle nous avons représenté ceci sous la forme d'un engagement social du deuxième agent envers le premier pour effectuer la tâche. Ceci confère au premier agent un pouvoir indirect fort (cf. Formule 7.16) : une partie de la satisfaction d'un de ses buts sera effectuée par un autre agent.

$$depends(x,z,g_2) \wedge power\_over(x,z,g_3) \xrightarrow{cf.\ 7.17} weak\_ind\_power\_of(x,z,g_2)$$

$$depends(x,z,g_2) \wedge \exists sc = \langle z,x,ia,achieved(g_2),\beta \rangle \in SC_{ia} \xrightarrow{cf.\ 7.16} strong\_ind\_power\_of(x,z,g_2)$$

Cet engagement social créé par un agent représente une contrainte qui limite son comportement : il est censé avoir un comportement orienté vers la satisfaction de l'objet de l'engagement, donc l'exécution d'une tâche. Cependant, même si un agent adopte une tâche d'un autre agent et s'engage de l'exécuter, ceci ne signifie pas forcément que cette tâche sera effectuée. Un agent peut changer d'avis et ne pas essayer de l'effectuer ou il peut essayer mais ne pas réussir. En tout cas, un tel engagement social peut être violé, ce qui dans notre modèle est équivalent au fait que l'agent qui s'était engagé est considéré autonome. Nous rappelons que la décision qu'un agent prend de poursuivre ou non une contrainte comportementale telle qu'un but adopté d'un autre agent représente un raisonnement *a posteriori* de l'agent. Ce raisonnement porte en fait sur la décision de l'agent de se retrouver ou non dans un comportement autonome, de violer ou non un engagement social.

Nous analyserons à la fin de ce chapitre les raisons qui poussent les agents à être dans un comportement autonome et comment différents types d'agents effectuent ce raisonnement *a posteriori*. Mais avant, nous allons présenter des raisonnements similaires à ceux qui ont été présentés ci-dessus, en prenant en compte des concepts institutionnels.

8.2. Raisonnement dans un contexte institutionnel

Si dans la section précédente nous avons analysé le raisonnement des agents lié aux interactions sans prendre en compte l'existence des différents mécanismes de coordination institutionnel, par la suite nous allons nous intéresser aux implications de la présence de ces mécanismes. Nous pouvons ainsi

identifier deux cas possibles dans notre analyse : l'interaction entre agents dans un contexte institutionnel et l'interaction entre un agent et une organisation.

8.2.1. Interactions entre agents dans un contexte institutionnel

Par rapport au raisonnement des agents lié aux interactions que nous avons analysé dans la section précédente, la présence des mécanismes de coordination institutionnels crée des situations intéressantes. Notre modèle unifie les contraintes interpersonnelles et institutionnelles en les représentant à l'aide de mêmes concepts. Par exemple, les obligations imposées à un agent, ainsi que les relations d'autorité entre agents sont toutes deux représentées sous la forme d'engagements sociaux. Les formules introduites dans le Chapitre 7 construisent en fait une abstraction des contraintes interpersonnelles ou institutionnelles en utilisant des concepts d'un niveau de plus en plus haut pour aboutir à la notion d'autonomie.

Dans l'exemple précédent, un agent x choisit comme partenaire d'interaction l'agent z parce qu'il existe une relation de dépendance réciproque entre eux. Cette relation de dépendance signifie que x n'a pas d'autonomie par rapport à z pour la satisfaction d'un but g_2 et que z n'a pas d'autonomie par rapport à x pour la satisfaction d'un autre but g_3. Autrement dit, x base son raisonnement *a priori* sur l'hypothèse que z a un but qu'il ne peut satisfaire, hypothèse qui peut s'avérer fausse. Il sera mieux pour x s'il calcule le manque d'autonomie d'un autre agent, manque issu d'une autre source, comme par exemple une relation d'autorité entre les agents x et y suite aux rôles qu'ils jouent. Cette relation est connue par les agents appartenant à l'organisation et, plus qu'une simple dépendance, elle est renforcée par l'organisation : l'agent x a plus des garanties que y adoptera la satisfaction du but g_2 pour lui.

$$authority(x,y,g_2) \xrightarrow{cf.\ 7.13} auth\_power\_over(x,y,g_2) \xrightarrow{cf.\ 7.14,7.18} \neg has\_autonomy(y,x,g_2)$$

Les formes de raisonnement présentées dans la section précédente peuvent être effectuées dans un contexte institutionnel aussi, sans changements nécessaires. Par exemple, le choix d'un partenaire d'interaction (raisonnement *a priori*) sera toujours fait en privilégiant les agents qui n'ont pas d'autonomie par rapport à l'agent qui délègue une tâche, c.-à-d., les agents sur lesquels l'agent déléguant a du pouvoir. Ce pouvoir sur d'autres peut être issu d'une dépendance, mais aussi d'une relation d'autorité (cf. Formule 7.13). Parfois, comme dans l'exemple ci-dessus, la source de ce pouvoir qu'un agent a sur un autre peut être importante, parfois elle peut être ignorée, mais le raisonnement de l'agent reste le même.

Une modification introduite par le contexte institutionnel apparaît dans le raisonnement *a posteriori* d'un agent, le raisonnement sur la violation ou non d'une contrainte comportementale telle qu'un engagement social. La présence des institutions électroniques sert à renforcer l'obéissance à ces contraintes, l'agent institutionnel étant capable d'imposer des sanctions aux agents qui décident de se retrouver dans un comportement autonome. Dans son raisonnement *a posteriori* l'agent doit maintenant prendre en compte les éventuelles sanctions associées à la violation de ses engagements. Cependant, la discussion sur les différents types d'agents et leurs décisions d'être ou ne pas être dans un comportement autonome reste valide : un agent égoïste préfère toujours violer un engagement social si les gains associés à cette violation (p.ex. la satisfaction d'un but) sont plus grands que les sanctions associées.

8.2.2. Interactions entre un agent et une organisation

Si nous considérons maintenant les interactions entre un agent et l'organisation à laquelle il appartient et dans laquelle il joue des rôles, le raisonnement des agents n'est pas très différent. Comme nous l'avons discuté dans la partie précédente, le fait qu'un agent joue un rôle dans une organisation crée des contraintes sur son comportement. Notamment, la présence des obligations qu'il a envers l'organisation – représentées dans notre modèle sous la forme de meta-engagements sociaux, l'engagement qu'il a pris pour jouer le rôle et les engagements pour satisfaire les buts du rôle, tous ces engagements représentent des contraintes comportementales.

Par exemple, un agent peut être engagé envers l'organisation de satisfaire un but associé au rôle qu'il joue dans cette organisation. De point de vue de la représentation que nous utilisons, il n'y a pas des différences entre cet engagement envers l'organisation et l'engagement envers un autre agent (une adoption d'une tâche suite à une interaction, comme dans l'exemple précédent) :

$$rolecomm=\langle x,oa,ia,achieved(g),\beta\rangle\in SC_{ia}$$

Un agent doit être ainsi capable d'un raisonnement *a posteriori* institutionnel dans lequel il s'intéresse à la possibilité de violer une contrainte comportementale de provenance institutionnelle. Ceci se traduit dans notre modèle par la violation ou non d'un engagement social (envers l'organisation) et donc la décision d'être ou de ne pas être autonome. Ce type de raisonnement reste donc inchangé dans sa nature, indifféremment de la nature de la contrainte à violer. Cependant, la nature de la contrainte qu'un agent envisage de violer peut jouer un rôle intéressant : en règle générale, la violation des engagements qu'un agent a envers une organisation est punie plus sévèrement que la violation des engagements entre agents. Un agent doit ainsi bien calculer les conséquences de son comportement autonome.

Pour mieux comprendre le raisonnement *a posteriori* dans un contexte institutionnel et les implications d'un comportement autonome d'un agent, prenons l'exemple suivant. Une relation d'autorité existe entre deux agents suite à deux rôles qu'ils jouent dans l'organisation. Cette relation est représentée dans notre modèle sous la forme d'un engagement de l'agent « subordonné » envers l'organisation, engagement qui l'oblige d'accepter toute délégation de but de la part de l'autre agent (Formule 1 ci-dessous). Les agents interagissent et l'agent « supérieur » délègue un but à son subordonné (Formule 2 ci-dessous): ci ce dernier refuse l'adoption du but, il viole son engagement envers l'organisation (Formule 3 ci-dessous). Il se retrouve ainsi dans un comportement autonome envers l'organisation (Formule 4 ci-dessous), ce qui peut être puni sévèrement, par exemple par l'exclusion de l'organisation. Cependant, s'il adopte le but délégué, il s'engage envers l'autre agent pour satisfaire ce but (Formule 5 ci-dessous). S'il décide de violer cet engagement, il se retrouve dans un comportement autonome, mais par rapport à un autre agent, ce qui peut ne pas être puni de la même manière (Formule 6 ci-dessous). Le raisonnement *a posteriori* porte donc sur le fait d'être ou ne pas être dans un comportement autonome – il faut savoir choisir quand il faut être autonome et quand il ne le faut pas.

1. $aut=\langle y,oa,ia,\left(\forall x{:}R_2\ request(x,y,achieved(g,\beta))\Rightarrow\exists sc=<y,x,achieved(g),\beta>\in SC_{ia}\right)\ y{:}R_1\rangle\in SC_{ia}$
2. $request(x,y,g_2)$

Soit il n'adopte pas :

3. $<y,x,achieved(g_2),\beta> \notin SC_{ia} \xrightarrow{cf.\ 5.11} aut \in VioSC_{ia}$

4. $\xrightarrow{cf.\ 7.19} is\_autonomous(y,oa,R_1)$

Soit il adopte :

5. $\exists sc = <y,x,achieved(g_2),\beta> \in SC_{ia}$

6. $sc \in VioSC_{ia} \xrightarrow{cf.\ 7.19} is\_autonomous(y,x,g_2)$

Si le raisonnement *a posteriori* institutionnel est effectué par un agent qui joue un rôle, le raisonnement *a priori* institutionnel se réfère aux situations menant à celle-ci : il est utilisé par un agent avant de choisir un rôle à jouer ou par un agent qui envisage d'abandonner un rôle. Nous décrivons dans plus de détails ce raisonnement dans le chapitre suivant, après avoir analysé comment des types d'agents différents effectuent le raisonnement *a posteriori*.

8.3. Raisonnement *a posteriori* – autonomie – types d'agents

L'acquisition de pouvoirs indirects forts et la capacité de satisfaire des buts par l'intermédiaire des autres agents fonctionnent seulement si les autres agents remplissent leurs engagements. Il est possible que, bien qu'il soit engagé pour effectuer une tâche, un agent ne parvienne pas à l'effectuer dans les conditions prévues (avant la date limite, par exemple). Cet échec de réalisation de l'objet de son engagement peut être accidentel ou volontaire – dans notre modèle, indifféremment de la raison de l'échec, l'engagement social est considéré *violé* et des mesures peuvent être prises. Les caractéristiques de ces mesures dépendent du concepteur du système/institution électronique, comme nous l'avons discuté dans les Chapitres 2 et 6. Dans notre approche, un engagement social violé est équivalent à un comportement autonome, comportement qui peut être le résultat d'un raisonnement *a posteriori* d'un agent ou non, comme nous l'avons vu dans les sections précédentes.

Etiqueter un agent comme étant ou n'étant pas dans un comportement autonome représente une perspective externe à l'agent, le point de vue d'un observateur externe (celui des agents institutionnels). Cet observateur n'a pas accès au raisonnement de l'agent, il peut juger le caractère autonome de l'agent seulement à partir de son comportement. L'utilisation d'une représentation explicite des contraintes comportementales sous la forme des engagements sociaux permet ainsi à un observateur d'étiqueter le comportement d'un agent comme autonome ou non-autonome. Cependant, il est parfois impossible pour un agent de remplir un de ses engagements sociaux (p.ex., s'il est engagé d'effectuer une action qui a une durée trop longue par rapport à sa date limite) ou, à cause d'un environnement qu'il ne contrôle pas, l'agent peut essayer de remplir un engagement qu'il a et échouer. Un tel agent est donc étiqueté en tant qu'autonome et il peut se retrouver redevable pour une sanction associée à un engagement violé, même s'il n'était pas dans son intention de le violer. Ce genre de cas doit être pris en compte par le concepteur d'un système multi-agents pour éviter des sanctions non-justifiées.

Ce qui nous intéresse plus dans notre travail n'est pas le comportement accidentellement autonome d'un agent, mais celui qui résulte d'une délibération explicite. Nous sommes ainsi intéressé par le cas où un agent *se retrouve délibérément dans comportement autonome*, c.-à-d., la violation d'une

contrainte externe sur son comportement, donc d'un engagement social est intentionnelle. La question qui se pose ainsi est d'identifier les raisons qui poussent un agent à désobéir à ses engagements ? La réponse à cette question est donnée en considérant que le comportement d'un agent est orienté vers la satisfaction de ses désirs, motivations ou buts, comme par exemple de gagner de l'argent ou, pour généraliser, de maximiser une utilité. Un agent utilise un raisonnement *a posteriori* pour décider d'être ou non dans un comportement autonome par rapport à une contrainte et cette décision est prise en fonction de ses motivations.

Prenons par exemple l'engagement social d'un agent d'exécuter une action pour un autre agent. L'agent a le choix entre deux comportements possibles, un orienté vers l'exécution de cette action, l'autre qui ne l'inclut pas dans ses plans. Son raisonnement *a posteriori* doit ainsi prendre en compte lequel des deux comportements possibles satisfait le plus ses motivations. Cependant, nous considérons nécessaire de distinguer entre deux types des motivations. D'un coté nous avons les motivations propres à l'agent, que l'on peut qualifier d'égoïste, qui cherchent à maximiser le gain de l'agent (p.ex. : gagner le plus d'argent), de l'autre nous avons les motivations orientées vers le gain de la société ou d'autres agents, que l'on peut qualifier d'altruiste (p.ex. : permettre à l'organisation de gagner le plus d'argent). De ce point de vue, nous pouvons diviser les agents en deux catégories :

- *agents égoïstes* (self-interested en anglais) qui sont motivés seulement par des désirs ou motivations égoïstes comme par exemple de maximiser leur utilité individuelle. Ces agents obéissent aux contraintes comportementales seulement parce que c'est dans leur meilleur intérêt (le but d'une obligation leur convient ou la sanction associée ne leur convient pas). Ce type d'agent peut facilement se retrouver dans un comportement autonome et fait appel souvent au raisonnement *a posteriori*.

- *agents altruistes* qui ont un comportement orienté vers le gain de l'organisation (société, système) ou des autres agents. Souvent, ces agents ne prennent même pas en considération l'éventualité de désobéir aux contraintes comportementales et ils se retrouvent donc dans un comportement autonome seulement par accident. Cependant, dans certains systèmes, le raisonnement *a posteriori* reste nécessaire même pour les agents altruistes. Il est possible par exemple que les normes définies dans une organisation soient contradictoires : un agent altruiste se retrouve ainsi contraint de poursuivre deux comportements contradictoires. Le raisonnement *a posteriori* d'un tel agent lui permettra de choisir le cours d'action qui sera le plus avantageux pour la société, même si ceci implique une violation d'un engagement social ou une norme. Comme nous l'avons discuté dans le Chapitre 2, il est possible que parfois une norme (ou un engagement social) soit violée pour le *bien de l'organisation* – le raisonnement *a posteriori* reste ainsi utile même pour les agents altruistes.

Bien évidement, il est possible que des agents soient à la fois égoïstes et altruistes s'ils ont les deux types de motivations, mais ceci n'est pas important pour notre travail. Notre objectif n'est pas de concevoir des agents. Nous ne sommes donc pas intéressé par la prise de décision effective pour violer ou non un engagement social. Cette prise de décision dépend de l'agent, du système, du contexte, etc. Ce que nous avons voulu montrer est que le raisonnement *a posteriori*, donc le raisonnement d'être ou non autonome, peut être utile dans certaines situations et que le modèle d'autonomie que nous avons proposé rend ce raisonnement possible.

8.4. Synthèse

Dans ce chapitre nous avons analysé des différents raisonnements effectués par un agent, soit dans un contexte interpersonnel où tout concept institutionnel peut être ignoré, soit dans un contexte institutionnel. Cette analyse a été effectuée en utilisant le modèle d'autonomie que nous avons introduit dans la partie précédente et qui se base sur des concepts tels que les engagements sociaux ou le pouvoir social. Nous avons ainsi analysé le raisonnement effectué par un agent dans un contexte interpersonnel pour satisfaire ses buts, notamment comment l'agent identifie les buts qu'il ne peut pas satisfaire tout seul (calcul des pouvoirs), comment il choisit un agent pour lui déléguer la satisfaction d'un but – raisonnement *a priori* (situations de dépendance), comment représenter une adoption d'un but de la part de l'autre agent (engagement social) et comment raisonner s'il faut remplir ou non un tel engagement social – raisonnement *a posteriori*.

La décision de remplir ou non un engagement social peut être traduite dans une décision d'être ou non dans un comportement autonome. Ce raisonnement *a posteriori*, qu'il soit interpersonnel ou institutionnel, est ainsi directement lié à la notion d'autonomie. Dans ce chapitre nous avons discuté les raisons qui poussent les agents à être dans un comportement autonome et les différents types d'agents qui peuvent exister en fonction de comment ils prennent cette décision.

Le modèle que nous avons proposé dans la partie précédente agrége les contraintes interpersonnelles et institutionnelles à l'aide de la théorie du pouvoir social et de l'autonomie. Ceci fait que les raisonnements d'un agent dans un contexte interpersonnel sont identiques à ceux effectués si des concepts institutionnels sont pris en compte aussi. Autrement dit, le raisonnement sur les contraintes sur la prise de décision ou le comportement d'un agent peut être fait en représentant ces contraintes d'une manière uniforme à travers le concept d'autonomie. De ce point de vue, la description des raisonnements effectuée dans ce chapitre ne représente qu'une illustration de la mise en œuvre du modèle de pouvoirs et autonomie que nous avons proposé. Nous sommes conscients que d'autres approches peuvent être envisagées pour ces raisonnements (comme nous l'avons vu dans les travaux similaires présentés dans le Chapitre 4). Cependant, une forme de raisonnement a été délibérément ignorée – le raisonnement *a priori* institutionnel – le manque de travaux similaires dans la littérature et son importance dans le cadre des systèmes ouverts et dynamiques font de ce raisonnement un cas à part qui illustre mieux l'utilité du modèle proposé et qui sera traité dans le chapitre suivant.

9. Raisonnement *a priori* institutionnel

Dans le chapitre précédent nous nous sommes intéressés aux différents types de raisonnement et à la manière dont ces raisonnements peuvent bénéficier du modèle d'autonomie proposé dans cette thèse, nous analysons dans ce qui suit le cas particulier du raisonnement *a priori* institutionnel. Celui-ci concerne l'évaluation du point de vue d'un agent des implications qu'un changement de son statut institutionnel peut avoir sur ses capacités à satisfaire ses buts (motivations, désirs, etc.). Ici, nous regroupons dans le terme 'changement de statut institutionnel' tout changement lié à la relation entre un agent et une organisation/institution. Ces changements concernent par exemple, l'entrée ou la sortie de l'agent d'une organisation, son changement de rôle au sein d'une structure organisationnelle, le fait qu'il commence à jouer un rôle ou arrête d'en jouer un autre, etc.

Ce changement de statut institutionnel est particulièrement important dans le cadre général de notre travail, notamment dans les systèmes ouverts et dynamiques. Dans de tels systèmes, qui selon notre vision sont associés à des mécanismes institutionnels, les agents peuvent entrer ou sortir, la structure organisationnelle ou l'attribution des rôles aux agents peuvent changer à tout moment. Nous considérons ainsi qu'il est important pour un agent de comprendre les implications de ces changements sur ses performances futures, donc d'être capable d'un raisonnement *a priori* institutionnel.

Dans ce chapitre, nous allons résumer ce changement de statut institutionnel à un seul cas, celui d'un agent qui commence à jouer un rôle. Cette hypothèse est justifiée dans notre travail parce que, comme nous l'avons vu au Chapitre 6, nous avons utilisé la notion de rôle comme une notion centrale pour les contraintes institutionnelles imposées aux agents. Ainsi, même si dans ce chapitre nous n'analysons que le cas d'un agent qui analyse un rôle qu'il pourrait jouer, cette analyse peut être facilement étendue à des situations où un agent veut entrer dans un groupe (où il jouera un rôle) ou un agent qui veut abandonner un rôle. En fait, comme nous le verrons par la suite, cette analyse réside dans l'identification des différences qui existent entre l'agent quand il ne joue pas le rôle et quand il joue le rôle. Ces différences sont liées à ses capacités de satisfaire ses buts. Dans ce chapitre nous allons analyser la signification pour un agent de jouer un rôle et comment il peut raisonner en terme des pouvoirs et autonomie gagnés et perdus, tandis que dans le chapitre suivant nous allons donner un exemple concret de ce raisonnement.

9.1. Signification pour un agent de jouer un rôle

Dans notre modèle, en nous basant sur des travaux similaires dans la littérature, nous avons défini les caractéristiques d'un rôle vis-à-vis d'un agent qui le joue, comme étant composées de ressources et de permissions reçues, de connaissances acquises, d'obligations à suivre, de buts à satisfaire et de structures hiérarchiques dans lesquelles se situer. Cependant, la notion de rôle est beaucoup plus riche et multidisciplinaire. Elle est utilisée dans des domaines tels que les bases de données, les ontologies ou les langages de programmation et non seulement dans la modélisation des organisations humaines

ou d'agents[9]. Même si, comme dans notre travail, nous considérons cette notion seulement de ce dernier point de vue, la modélisation des organisations multi-agents représente un concept très riche et les implications de jouer un rôle pour un agent peuvent être très variées et subtiles. Dans un travail précédent, (Boissier *et al.* 2005), nous nous somme intéressés à la signification de jouer un rôle dans une organisation humaine pour un individu. Par la suite nous détaillons cette signification, en soulignant les parties où notre modèle, conçu pour des agents artificiels, peut être utilisé.

9.1.1. Influence du rôle sur l'individu

Dans (Boissier *et al.* 2005), nous argumentons que non seulement le comportement d'un individu est sérieusement modifié par le rôle qu'il joue, mais nous considérons que sa manière de raisonner et ses relations avec d'autres individus sont également touchées.

Buts adoptés

Une des sources de ces modifications concerne l'adoption des buts ou désirs de son rôle : un individu qui joue un rôle est censé poursuivre les buts qui caractérisent ce rôle. Cette adoption des buts est problématique pour un individu, surtout si les buts du rôle sont contradictoires avec les buts propres de l'individu. Dans le meilleur des cas, il n'existe pas de contradictions entre les deux ensembles de buts, mais si une telle contradiction existe, l'individu doit choisir quel but privilégier : le sien ou celui de son rôle.

Notre modèle considère cet aspect : un agent qui joue un rôle devient engagé pour la satisfaction des buts du rôle (cf. Formule 6.1) – son comportement doit être orienté vers leur satisfaction, mais l'agent peut toujours violer ses engagements, c.-à-d., ne pas poursuivre les buts.

Croyances adoptées

A part des buts ou désirs, un autre type d'élément qu'un individu doit adopter en jouant un rôle concerne les croyances du rôle (voir (Boella & van der Torre 2003), comme nous l'avons vu au sein du Chapitre 2) : le comportement d'un individu qui joue un rôle doit être conforme aux croyances associées au rôle. Même si l'individu considère ces croyances fausses (différentes des siennes), il ne doit pas le montrer, sous peine de sanction.

Nous ne modélisons pas cet aspect dans notre étude. La raison réside principalement dans le fait que les seules croyances que nous associons aux agents font référence à leurs relations de dépendance ou de pouvoir. Cependant, un modèle similaire au notre peut être envisagé. Ce modèle prendrait en compte les croyances des agents et des rôles – une description de comment ces croyances influencent le comportement d'un agent est alors nécessaire.

Nouvelles connaissances

Quand un agent entre dans une organisation (où qu'il joue un rôle), un individu passe souvent par une phase d'apprentissage : il acquiert de nouvelles connaissances et gagne l'accès à de nouvelles informations. Le but de ce processus est d'assurer qu'il est compétent pour pouvoir mieux satisfaire

[9] Voir par exemple les actes de « Roles : an interdisciplinary perspective » - AAAI Fall Symposium 2005.

les buts que l'organisation lui délègue. Cependant, il arrive souvent qu'un individu utilise ses nouvelles connaissances pour la satisfaction de ses buts propres et pas uniquement ceux de son rôle.

Dans notre modèle, nous prenons en considération ce gain de connaissance en explicitant les éléments qu'un agent apprend quand il joue un rôle, notamment de nouvelles actions qu'il sait exécuter et de nouveaux plans qu'il peut utiliser (Formules 6.7 et 6.10). Avec ces nouveaux éléments appris, un agent augmente ses pouvoirs individuels. Il devient capable de satisfaire un plus grand nombre de buts, ce qui peut être dans l'intérêt de l'organisation. Notre modèle permet ainsi de représenter les nouveaux pouvoirs individuels qu'un agent acquiert en jouant un rôle – ces pouvoirs peuvent concerner les buts de son rôle, mais aussi ses buts propres qui peuvent être satisfaits grâce aux nouveaux pouvoirs acquis.

Ressources et permissions

Un individu augmente ses pouvoirs individuels en recevant de la part de l'organisation des ressources et des permissions. Par exemple, dans les organisations humaines, un individu qui entre dans une organisation peut recevoir une voiture ou un ordinateur portable (des ressources), mais aussi le droit d'utiliser une connexion internet ou une imprimante (des permissions).

Nous modélisons tout ceci au travers des formules 6.7 et 6.10 : un agent reçoit de son rôle des ressources et des permissions, éléments qui augmentent respectivement ses pouvoirs d'exécution et déontiques, donc ses pouvoirs individuels.

Prohibitions et obligations

Mis à part les permissions, d'autres éléments normatifs peuvent être associés à un rôle. Ce sont notamment les interdictions et les obligations. Un individu qui joue un rôle peut être interdit d'effectuer certaines tâches et peut être obligé d'en effectuer d'autres.

Dans notre modèle, tout manque de permission (explicite) est considéré comme une interdiction. Les obligations sont représentées sous la forme de politiques sociales (Formule 6.3). Alors que les interdictions (le manque de permissions) peuvent impliquer un manque de pouvoir déontique d'un agent (Formule 7.2), les obligations génèrent un pouvoir de l'organisation sur l'agent (Formule 7.12), relation de pouvoir qui à son tour implique un manque d'autonomie de l'agent (Formule 7.18).

Ce pouvoir d'une organisation sur un de ses membres représente dans notre modèle une relation très intéressante : la mise à disposition de pouvoirs de l'organisation humaine vers l'un individu qui y appartient. Quand il entre dans une organisation, un individu doit mettre à la disposition de celle-ci une partie de (parfois tous) ses pouvoirs, c.-à-d., quand l'organisation le désire, l'individu doit effectuer la tâche demandée (doit utiliser son pouvoir concernant la tâche). Dans les organisations humaines, les pouvoirs mis à disposition de l'organisation ne sont pas toujours formellement définis. Ceci peut générer des abus de la part de l'organisation, qui peut demander à un de ses membres d'utiliser des pouvoirs qu'il n'avait pas explicitement mis à sa disposition. Dans notre modèle, cette est impossible du fait des obligations d'un rôle ou des dépendances d'un agent vis-à-vis d'un rôle (Formule 7.6). L'organisation a des pouvoirs sur un agent, mais ces pouvoirs sont tous définis par rapport à des tâches précises. Un agent jouant un rôle peut ne pas avoir d'autonomie par rapport à

l'organisation. L'autonomie étant une notion relative, il peut ne pas l'avoir pour satisfaire un but, mais il peut l'avoir pour la satisfaction d'un autre. Autrement dit, dans notre modèle, un agent met à la disposition d'une organisation une partie de ses pouvoirs, mais il garde son autonomie décisionnelle pour d'autres.

9.1.2. Influence du rôle d'un individu sur ses relations sociales

Relations d'autorité

Les relations de pouvoir social mentionnées ci-dessus ne concernent pas uniquement une organisation et un de ses membres mais elles peuvent également apparaître entre deux individus. En dehors des dépendances entre individus (cf. Formule 7.4), une des principales sources de pouvoir d'un individu sur un autre est représentée par les liens d'autorité existants dans les organisations. Des hiérarchies de rôles sont souvent utilisées au sein des organisations. Selon notre modèle ceci donne aux membres du pouvoir sur d'autres (Formule 7.13). Un individu jouant un rôle dans une organisation devient ainsi impliqué dans de nombreuses relations de pouvoir avec d'autres individus.

Dans notre modèle, jouer un rôle dans une hiérarchie de rôles signifie une perte d'autonomie pour un agent et un gain de pouvoir pour un autre. Ce gain de pouvoir est lié à la représentation du lien d'autorité sous la forme d'une politique sociale. Ainsi, une délégation de but sera suivie par la création d'un engagement social pour satisfaire ce but. Un agent qui a du pouvoir sur un autre, peut ainsi acquérir du pouvoir indirect par l'intermédiaire de l'autre agent, qui mettra en place les éléments nécessaires à la satisfaction des buts à sa place (Formules 7.16 et 7.17).

Count-as

Quand il joue un rôle, le comportement d'un individu peut être modifié par les aspects considérés ci-dessus. Cependant, la manière dont les autres individus interprètent son comportement est également modifiée. Souvent, le comportement n'est plus considéré comme le comportement de l'individu même, mais du rôle joué. Cet effet, appelé l'effet *count-as* dans la littérature REF, signifie que tout ce qu'un individu fait ou dit est interprété comme ce que son rôle fait ou dit. Un individu doit ainsi prendre en compte cet effet et adapter son comportement en conséquence : souvent dans la société humaine des individus sont sanctionnés parce que leur comportement n'a pas été conforme au comportement que leur rôle devrait avoir.

Cet aspect manque à notre modèle. Nous ne nous intéressons pas à ce type de modélisation du comportement des autres individus. Cependant, nous sommes conscients que l'effet count-as représente une extension très intéressante. Un agent devrait être capable de représenter les pouvoirs gagnés sur d'autres agents en jouant un rôle même s'il n'agit pas dans un contexte organisationnel. Un agent a des pouvoirs provenant du rôle qu'il joue même s'il n'agit pas en tant que représentant du rôle.

Relations de dépendances

Comme nous l'avons précisé, en dehors du contexte institutionnel, les individus dépendent les uns des autres pour satisfaire leurs buts – ils doivent s'aider réciproquement. Bien évidemment, suite à la modification des pouvoirs individuels par les rôles qu'ils jouent (voir ci-dessus), les « réseaux de

dépendance » changent pour les individus quand ils commencent à jouer des rôles. Cependant, une modification encore plus subtile peut exister : parfois les rôles dans les organisations humaines ont leur propre réseau de dépendance défini et tout agent qui joue un tel rôle doit adopter et utiliser ce réseau. Par exemple, le rôle directeur dépend du rôle secrétaire pour fixer un rendez-vous : tout individu qui joue le premier rôle doit utiliser cette relation de dépendance et non une autre (ou son pouvoir individuel de fixer un rendez-vous).

Dans notre modèle nous ne considérons pas les dépendances pré-établies des rôles que les agents doivent adopter, mais nous considérons seulement les dépendances inter-agents. Cependant, vu les concepts que nous utilisons, notre modèle peut être facilement étendu dans cette direction.

Confiance

Finalement, il existe encore une source de modification des relations entre individus suite aux rôles qu'ils jouent. Le fait de jouer un rôle ou d'appartenir à une organisation confère à un individu un statut social, statut souvent assimilé à la notion de confiance. Des individus peuvent faire confiance à un autre tout simplement parce qu'il appartient à une organisation, organisation qui joue le garant d'un bon comportement de l'individu. Un exemple de cette relation sociale est la confiance existante dans la société humaine dans les individus qui jouent le rôle de médecin – les autres leur font confiance pas nécessairement à cause de leurs compétences, mais à cause des rôles qu'ils jouent.

Le domaine de la confiance dans les systèmes multi-agents REF est encore jeune et la proposition et l'utilisation d'un modèle de confiance est en dehors des objectifs de cette thèse – pour cette raison nous ne considérons pas dans notre modèle les changements de confiance dans un agent qui joue un rôle.

9.2. Evaluer un rôle en terme de pouvoirs sociaux et d'autonomie

Cette revue des différentes manières dont jouer un rôle peut transformer le comportement, le raisonnement ou les relations sociales d'un agent illustre la richesse de la notion de rôle et la complexité de ce problème. Même si tous les aspects ne sont pas pris en compte, nous avons argumenté comment notre modèle, utilisant les concepts d'engagement social, de dépendance, de pouvoir social et d'autonomie, s'intéresse à une bonne partie d'entre eux. Il est donc possible d'utiliser ce modèle pour décrire les implications de jouer un rôle pour un agent et de prendre ainsi des décisions en fonction de ces implications. Par la suite nous décrivons de quelle manière l'utilisation de ce modèle peut améliorer ce type de raisonnement.

9.2.1. Pouvoirs et autonomie gagnés et perdus en jouant un rôle

Une des principales raisons pour laquelle un agent accepte de jouer un rôle est liée au grand nombre de pouvoirs individuels qu'il gagne en le jouant. Ces nouveaux pouvoirs sont issus des ressources, des permissions reçues et des connaissances apprises quand il commence à jouer le rôle (voir les caractéristiques d'un rôle telles que nous les avons définies dans le Chapitre 6). Les pouvoirs individuels d'un agent augmentent ainsi sensiblement quand il joue un rôle et diminuent quand il ne le joue plus. Si un agent a plus de pouvoirs individuels, il dépend moins des autres agents (ce qui fait

qu'il a plus d'autonomie sociale). Dans le même temps il est plus probable que d'autres dépendent de lui. Les relations de pouvoir social et d'autonomie changent ainsi grâce aux nouveaux pouvoirs individuels des agents qui jouent des rôles. Ceci constitue une motivation pour un agent à jouer un rôle : gagner des pouvoirs individuels et acquérir de l'autonomie sociale.

Ces nouveaux pouvoirs individuels ne représentent pas la seule source de modification de relations de pouvoirs : dans notre modèle nous considérons qu'un agent jouant un rôle est sujet de plusieurs relations d'autorité qui relient les rôles d'une organisation. Le fait de jouer un rôle qui a de l'autorité sur un autre rôle confère à un agent un pouvoir sur tout agent qui joue le deuxième rôle. En jouant un rôle, un agent gagne ainsi du pouvoir sur d'autres agents, mais d'autres gagnent également du pouvoir sur lui – les relations d'autorité peuvent être dans les deux sens. En jouant un rôle, un agent gagne et perd également de l'autonomie sociale par rapport à d'autres agents.

Faire partie d'une organisation ou institution multi-agents confère plus de sécurité à un agent. Comme nous l'avons discuté dans le chapitre précédent, un agent cherche à interagir avec des agents sur lesquels il a du pouvoir (grâce par exemple à une relation d'autorité ou à une dépendance) – cette relation de pouvoir lui donne un pouvoir indirect faible d'effectuer une tâche. Cependant, dans une organisation où un comportement autonome est sanctionné, il est plus probable que ce pouvoir indirect faible se transforme en un pouvoir indirect fort et qu'une tâche adoptée par un autre agent soit effectuée. Autrement dit, appartenir à une organisation signifie une probabilité plus forte que tout comportement autonome soit puni : ce qui rassure un agent qui dépend d'autres pour la satisfaction de ses buts.

Le gain des pouvoirs sur d'autres agents et la garantie que dans une institution les autres sont pénalisés s'ils ne respectent pas leurs engagements constituent d'autres raisons qui poussent les agents à accepter de jouer des rôles dans des organisations. Cependant, même s'ils gagnent des pouvoirs sur d'autres, d'autres gagnent aussi des pouvoirs sur eux : les agents perdent ainsi de l'autonomie. Le rapport entre l'autonomie perdue et le pouvoir sur d'autres acquis par un agent varie selon le rôle, l'organisation et l'application. Ceci explique pourquoi les agents cherchent généralement à jouer des rôles « avec plus de pouvoir » : non seulement pour plus de ressources, mais aussi parce que ces rôles leur offrent plus de pouvoirs sur d'autres tout en minimisant leur perte d'autonomie.

Quand il entre dans une organisation pour jouer un rôle, la principale perte d'autonomie d'un agent n'est pas envers d'autres agents, mais envers l'organisation. Si l'agent dépend du rôle qu'il joue, selon la Formule 7.11, l'organisation acquiert un pouvoir sur lui – il est contraint par le fait qu'il doit continuer de jouer ce rôle s'il veut satisfaire une partie de ses buts. Ce pouvoir que l'organisation a sur un de ses membres représente un manque d'autonomie du membre par rapport à l'organisation. Cependant, une grande partie du manque d'autonomie d'un agent par rapport à l'organisation est provoquée par les obligations associées au rôle joué par l'agent. Ces obligations, représentées dans notre modèle par des politiques sociales, limitent la prise de décision des agents en spécifiant ce qu'ils doivent faire quand l'organisation le souhaite (ou quand un contexte défini par l'organisation est actif). Ceci constitue un pouvoir de l'organisation sur l'agent et donc une perte d'autonomie de la part de l'agent.

Synthèse

Pour résumer, un agent qui accepte de jouer un rôle dans une organisation gagne des pouvoirs individuels pour satisfaire ses buts et donc de l'autonomie, gagne du pouvoir sur d'autres agents, mais d'autres agents et l'organisation gagnent aussi de pouvoir sur lui entraînant ainsi une perte de son autonomie. Un agent peut ainsi évaluer ce que le fait de jouer un rôle lui apportera, en terme de pouvoirs et d'autonomie ou lui supprimera, afin de prendre une décision informée de jouer ou non le rôle.

9.2.2. Raisonnement à base d'autonomie

Le raisonnement *a priori* institutionnel concerne, dans l'analyse que nous effectuons dans ce chapitre, l'évaluation par un agent de ce qu'il gagne ou perd en jouant un rôle, en terme de pouvoir et d'autonomie. Il ne faut pas oublier que les pouvoirs ou l'autonomie qu'un agent gagne ou perd sont des concepts relationnels. Autrement dit, un agent peut gagner le pouvoir individuel de satisfaire un but, mais il peut perdre son autonomie de décider (librement) concernant la satisfaction d'un autre but. En utilisant notre modèle et notre approche, un agent peut ainsi se représenter si le fait qu'il joue un rôle lui facilite ou non la satisfaction des buts. Ensuite, la décision de l'agent de jouer ou non le rôle dépend des buts mêmes : un agent privilégie les buts qui vont à la rencontre de ses motivations, désirs.

S'il est assez claire pourquoi un agent voudrait avoir le pouvoir individuel pour satisfaire les buts satisfaisant ses motivations, pourquoi est-ce que l'autonomie devrait être recherchée par un agent ? Nous rappelons que dans ce travail nous considérons l'autonomie sociale en décision d'un agent, c'est-à-dire la capacité d'un agent à prendre des décisions sans être influencé par une entité sociale externe (autre agent, organisation). Avoir de l'autonomie par rapport à une entité pour une tâche signifie dans notre modèle que cette entité n'a pas de pouvoir sur l'agent concernant l'exécution de la tâche. Autrement dit, cette entité externe ne peut ni forcer ni empêcher l'agent d'effectuer la tâche. Si, par exemple, il s'agit d'un but propre à l'agent, qu'il veut satisfaire, il est important pour l'agent d'avoir de l'autonomie, donc de décider librement de le poursuivre. Ou, si l'agent n'a pas d'autonomie et il est forcé de poursuivre un but qui ne satisfait pas ses motivations, ceci signifie que l'agent consomme des ressources et surtout du temps en poursuivant des buts qui ne lui sont pas propres et ne peut donc pas se focaliser sur la satisfaction de ses propres buts.

Il est donc important pour un agent d'avoir de l'autonomie en général et surtout pour certains buts, afin de pouvoir satisfaire ses motivations ou désirs. Cependant, comme argumenté par Castelfranchi (Castelfranchi 1995), avoir de l'autonomie représente un désir intrinsèque pour tout agent. Cependant, dans les systèmes multi-agents, il est possible d'envisager des agents qui ne cherchent pas à avoir de l'autonomie. Par exemple, un agent *altruiste* (voir le chapitre précédent) peut chercher les situations ou les rôles qui lui confèrent le moins d'autonomie possible, ce qui signifie qu'il sera le plus possible à la disposition d'autres agents ou de l'organisation.

Pour résumer, le raisonnement *a priori* institutionnel d'un agent lui permet d'évaluer les pouvoirs et l'autonomie qu'il gagne ou perd en jouant un rôle, évaluation rendue possible par l'utilisation de notre modèle. Cette évaluation lui permet ainsi de comprendre comment il peut mieux satisfaire ses motivations, en jouant ou non un rôle, en en choisissant un autre, etc. Bien évidemment, la décision

effective concernant le rôle à jouer dépend de l'application, de l'agent et de ses motivations. Nous avons tout simplement illustré comment cette décision peut être plus informée et quels sont les critères possibles qui peuvent être pris en compte.

9.3. Point de vue d'une organisation

Jusqu'à maintenant nous nous sommes intéressés au point de vue d'un agent : comment son raisonnement utilise le modèle à base des pouvoirs et d'autonomie que nous avons proposé. Cependant, ce modèle permet d'expliquer également le fonctionnement (et pourquoi pas le raisonnement) d'une organisation. Généralement une organisation est conçue pour satisfaire des buts de son concepteur. La manière pour arriver à ceci est de déléguer des sous-buts aux rôles qui la composent et de mettre en place des mécanismes qui assurent une satisfaction coordonnée de ces sous-buts. Notre modèle représente ces buts par des engagements associés aux rôles : un agent qui joue un rôle devient engagé envers l'organisation pour satisfaire les buts de son rôle (Formule 6.2). Même si nous ne nous sommes pas intéressés à ceci, il est possible d'exprimer dans notre modèle le fait qu'une organisation dépend de ses membres pour satisfaire ses buts : dans la mesure où il n'existe pas d'actions qu'une organisation peut exécuter en tant que telles, l'organisation n'a pas de pouvoirs individuels pour ses buts et elle dépend ainsi de ses membres. Ceci fait que, selon la Formule 7.16, *une organisation a un pouvoir indirect fort de satisfaire ses buts par l'intermède de ses membres*.

Pour convaincre les agents d'y entrer, mais aussi pour les aider à satisfaire ses buts, l'organisation met à leur disposition des ressources, des permissions et des connaissances, comme nous l'avons précisé auparavant. Cependant, déléguer des buts aux agents jouant des rôles ne suffit pas généralement pour assurer que les buts globaux de l'organisation soient satisfaits. Des mécanismes de coordination doivent être utilisés, mécanismes qui prennent la forme des obligations contextuelles : comme nous l'avons discuté, ces obligations génèrent un pouvoir de l'organisation sur ses membres. A partir de la Formule 7.17, *une organisation a un pouvoir indirect faible de satisfaire ses buts par l'intermédiaire de ses membres.* Ce pouvoir indirect faible complète le pouvoir indirect fort mentionné ci-dessus et permet à l'organisation de s'adapter aux changements extérieurs (déléguer de nouveaux buts, etc.).

Il semble ainsi tout à fait possible d'envisager une organisation capable du même type de raisonnement que celui *a priori* institutionnel décrit dans la section précédente. Si un agent peut représenter les conséquences que jouer un rôle a sur ses capacités de satisfaction de buts, une organisation peut obtenir la même représentation mais en s'intéressant à ses buts, et non à ceux de l'agent. Si un agent compare ainsi des rôles possibles à jouer afin de choisir celui qui lui convient le plus, une organisation peut également évaluer les candidats (agents) possibles pour un rôle afin de choisir l'agent qui, jouant le rôle, convient le plus pour la satisfaction de ses buts organisationnels. Cette approche intéressante représente une nouvelle utilisation de notre modèle à base d'autonomie et pouvoir social, utilisation que nous considérons comme une direction pour un travail futur.

9.4. Synthèse

Dans ce chapitre nous avons analysé la signification de jouer un rôle pour un agent. Cette problématique est spécialement pertinente dans le cadre des systèmes ou organisations ouverts, dans lesquels un agent qui rentre doit choisir parmi plusieurs rôles à sa disposition. Dans ce chapitre nous

avons analysé les différents aspects de la relation agent-rôle, notamment comment le comportement, le raisonnement et les relations sociales d'un agent changent quand il joue un rôle. Comme nous l'avons argumenté, le modèle que nous avons proposé dans les chapitres précédents nous permet de considérer une bonne partie de ces aspects d'une manière formelle.

Un agent utilisant notre modèle est ainsi capable d'identifier les pouvoirs individuels ou indirects qu'il gagne et perd en jouant un rôle et comment évoluent les relations de pouvoirs qui le lient à d'autres agents ou à l'organisation. Autrement dit, un agent peut évaluer les changements que jouer un rôle aura sur son autonomie, sur ses capacités de mieux satisfaire ses buts. Le raisonnement *a priori* institutionnel consiste dans cette évaluation et le choix de jouer ou non un rôle qui devient maintenant informé. Bien évidement, le même type de raisonnement peut s'appliquer facilement à d'autres situations similaires, telles que la décision d'entrer ou non dans une organisation, d'arrêter ou non de jouer un rôle, etc.

Nous avons également discuté la possibilité d'un raisonnement similaire mais dans le sens inverse : une organisation peut aussi évaluer les changements produits par le fait qu'un agent commence à jouer un rôle, changements qui vont bénéficier ou nuire à la satisfaction des buts de l'organisation. Même si notre travail est orienté dans le point de vue d'un agent, cette discussion nous permet d'envisager des perspectives intéressantes qui peuvent être prises en utilisant le modèle d'autonomie proposé.

10. Réalisation : ADOMO

Afin de mieux illustrer l'utilité d'un raisonnement à base de pouvoirs sociaux et d'autonomie tel que nous l'avons décrit dans le modèle formel proposé, nous avons développé un prototype d'application dans le domaine de l'intelligence ambiante. Ce dernier est bien adapté au cadre général de ce travail qui est celui des systèmes ouverts et dynamiques. Dans ce chapitre nous décrivons le scénario d'intelligence ambiante que nous avons envisagé. Nous l'avons intitulé *ADOMO* pour *ADvertising On MObile devices* (publicité sur des objets portables) et ce scénario est une version simplifiée d'une application réelle réalisée suite à une collaboration de l'auteur avec Siemens AG. Nous décrivons ensuite l'implémentation de ce scénario et son exécution sur des objets portables tels que les téléphones portables ou les PDA.

Des difficultés d'implémentation et surtout des contraintes issues de l'usage de l'application réalisée nous ont mené à chercher plusieurs solutions d'implémentation de ce scénario. Ces solutions ont ainsi résulté en une structure organisationnelle mettant à la disposition d'un agent entrant dans le système (organisation) deux rôles. L'agent doit choisir entre ces deux rôles en tenant compte d'une part de ses motivations et d'autre part des caractéristiques de ces deux rôles. L'agent met donc en œuvre un raisonnement *a priori* institutionnel. Nous décrivons ainsi comment ce raisonnement est effectué à l'aide de notre modèle dans ce cas précis. A la fin de ce chapitre nous décrivons aussi l'implémentation de ce modèle et de ce raisonnement sous la forme d'une application écrite en Prolog.

10.1. Description du scénario

Avec les améliorations récentes de la technologie mobile, un sous-domaine des applications multi-agents de commerce électronique est apparu : le *commerce mobile* (m-commerce) (Langer 2002). Dans un scénario de *m-commerce*, l'utilisateur, équipé d'un objet portable tel qu'un téléphone portable ou un PDA, se déplace et effectue des transactions (commerciales) avec d'autres utilisateurs, indépendamment de sa position. Des agents intelligents s'exécutent sur les objets portables afin d'aider les utilisateurs à effectuer des transactions. La capacité limitée de traitement des objets portables sur lesquels les agents s'exécutent, la mobilité de l'utilisateur (le changement continu de l'environnement de l'agent) et les connexions non filaires limitées sont les causes des nombreux défis techniques associés à ces applications. Par exemple, un des défis est de permettre la négociation entre agents (un processus qui peut nécessiter une longue durée), tout en sachant que la mobilité des utilisateurs et le temps de création d'une place de marché *ad-hoc* (Perkins 2000) dans laquelle les agents peuvent négocier limiteront le temps disponible aux agents pour négocier.

Le scénario que nous avons choisi d'utiliser dans notre travail s'appèle *ADOMO – ADvertising On MObile phones*. Ce scénario est une version adaptée d'un problème réel décrit dans (Carabelea and Berger, 2005) et implémenté dans une collaboration de l'auteur avec Siemens AG. Dans ce scénario un utilisateur est équipé d'un objet portable sur lequel un agent *AgentUtilisateur* (UA) s'exécute. Une partie de l'écran de l'appareil est utilisée pour afficher de la publicité pour différentes compagnies situées dans le voisinage de l'utilisateur, telles que des bars, magasins, restaurants, etc. Chacune de ces compagnies est représentée par un *AgentCommercial* (CA). Chacun de ces CA fait une offre pour

pouvoir afficher son logo dans une partie de ou sur tout l'écran pendant un laps de temps. Un CA peut aussi spécifier si l'utilisateur doit être averti quand un nouveau logo est affiché.

Les contrats publicitaires pour lesquels les agents négocient sont de deux types : court-terme ou long-terme. Les premiers sont créés d'une manière ad-hoc via une connexion temporaire et non fixe, quand l'utilisateur se trouve dans une zone commerciale. L'UA négocie de tels contrats avec les CA qui se trouvent dans son voisinage. Par rapport à ce type de contrat, l'UA peut aussi négocier des contrats long-terme avec des CA situés sur un serveur, d'une manière non ad-hoc, via une connexion fixe.

Pour chaque logo affiché, l'utilisateur gagne une petite somme d'argent. Dans notre cas nous considérons que les CA transfèrent une somme d'argent dans le compte bancaire de l'utilisateur pour chaque logo affiché. A la fin de la journée, ou quand il le demande, l'utilisateur peut contacter sa banque pour vérifier son compte. D'autres modèles de paiement plus élaborés peuvent être envisagés, mais ils ne sont pas d'intérêt pour cette thèse.

Le choix parmi plusieurs contrats proposés par des CA est influencé par les préférences de l'utilisateur. Par exemple, un utilisateur peut préférer des logos d'une certaine catégorie de compagnie (ex.: restaurants), ou il ne veut pas être averti quand un nouveau contrat est créé. Nous voyons donc que l'UA doit être capable de négocier des contrats publicitaires tout en prenant en compte les préférences de l'utilisateur.

Par la suite nous décrivons l'implémentation réalisée pour ce scénario et les défis techniques qui ont été résolus. Les solutions trouvées à ces défis constitueront le cadre d'un raisonnement *a priori* institutionnel.

10.2. ADOMO sur des objets portables

Le scénario décrit ci-dessus a été implementé comme un premier prototype. L'UA a été déployé sur un objet portable et tous les CA ont été exécutés sur un PC. Bluetooth[10] est un standard pour la communication sans-fil de plus en plus utilisé ces dernières années. Comme il semble bien adapté aux réseaux ad-hoc, nous avons décidé que les agents dans notre scénario communiqueront en utilisant Bluetooth. Une grande variété d'objets portables existent, comme les téléphones portables de diverses générations (nous parlons aujourd'hui de *smart phones*) ou les PDA comme ceux utilisant PALM-OS ou les PocketPC. Un bon nombre de ces objets portables sont équipés de Bluetooth et les PC actuels peuvent eux aussi être équipés facilement en utilisant un adaptateur USB-Bluetooth.

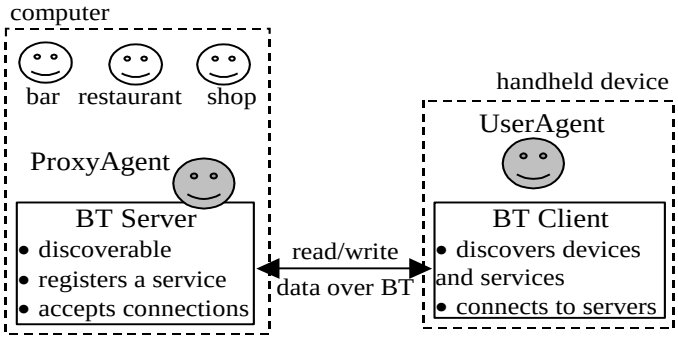
Pour faciliter le déploiement des applications multi-agents, plusieurs plateformes existent. Un état de l'art sur les plateformes multi-agents pour les objets portables est présenté dans (Carabelea et Boissier 2003). Leur utilisation pose des problèmes, particulièrement sur des objets avec des capacités de traitement limitées. Au moment de l'implémentation de ce prototype aucune plateforme multi-agents ne permettait la communication via Bluetooth. En prenant tous ces aspects en considération, nous avons décidé d'implémenter notre application sans utiliser de plateforme multi-agents. Pour rendre notre application portable sur différents objets, nous avons décidé d'utiliser Java. En ce qui concerne les objets portables sur lesquels nous avons déployé nos agents, nous avons décidé d'utiliser un

[10] The Official Bluetooth Site: http://www.bluetooth.org

PocketLoox Siemens-Fujitsu, mais nous avons aussi testé la même application sur un téléphone portable Siemens SX1 équipé avec Java et Bluetooth.

Un objet Bluetooth agissant comme *client* peut rechercher dans sa proximité (jusqu'à 10m) des objets (*serveurs*) qui offrent un service spécifique et peut se connecter à cet objet. Les services sont identifiés en utilisant un identifiant Bluetooth unique (UUID). Pour permettre à nos agents de communiquer via Bluetooth, nous avons créé une bibliothèque de communication – *Bluetooth Communication Package (BCP)*. Cette bibliothèque permet aux agents de découvrir d'autres agents dans leur voisinage et d'envoyer ou de recevoir des messages à/de ces agents. Un agent qui s'exécute sur un objet portable Bluetooth et qui veut découvrir les agents dans son voisinage lancera une découverte d'objets et pour chaque objet trouvé, il vérifiera s'il contient des services avec le UUID spécifique aux agents. Si c'est le cas, sur cet objet se trouve un agent avec qui il peut communiquer. Une connexion série peut ainsi se former entre les deux objets. L'agent client ouvre une connexion avec l'agent serveur au travers de laquelle les deux agents peuvent écrire et lire des octets.

Tous les agents commerciaux (CA) de notre scénario sont exécutés sur un PC. Quand plusieurs agents sont exécutés sur le même objet Bluetooth, l'utilisation de notre approche de communication n'est plus possible : un seul agent peut s'enregistrer comme service Bluetooth. Nous avons donc introduit un agent spécial, appelé *Agent Proxy* (PA), qui joue le rôle de serveur Bluetooth sur le PC. Il contient



une liste de tous les CA exécutés sur le PC et chaque fois qu'il reçoit un message adressé à un CA, il le transmet au destinataire et envoie la réponse si nécessaire.

Figure 10.1. L'architecture du système

L'agent utilisateur (UA) est déployé sur un objet portable (PocketLoox) jouant le rôle de client Bluetooth. Il est donc capable de découvrir des objets sur lesquels des agents sont exécutés, c.-à-d., des objets sur lesquels un PA s'est enregistré comme serveur. Après avoir ouvert une connexion à un PA, l'UA peut lui demander la liste d'agents CA qu'il connaît. Le PA joue donc dans notre système le rôle d'un agent médiateur (middle-agent).

10.2.1. Négociation entre agents

Dans le scénario ADOMO, un agent utilisateur négocie avec plusieurs agents commerciaux afin de créer un contrat publicitaire. Pour tous les types de contrat, l'UA demande à l'agent proxy la liste de CA disponibles et commence un processus de négociation avec ceux-ci. Ces négociations sont effectuées dans des réseaux ad-hoc, où un nombre limité de messages peut être échangé avant que les agents soient hors de portée de la communication. Le protocole de négociation doit être très simple

afin d'assurer la création des contrats le plus vite possible. Nous avons donc décidé d'utiliser un protocole d'enchère premier-prix offre-cachée (first-price sealed-bid – Figure 10.2), parce que ce protocole utilise moins de messages que d'autres protocoles tels que les enchères anglaises, la négociation par compromis, etc.

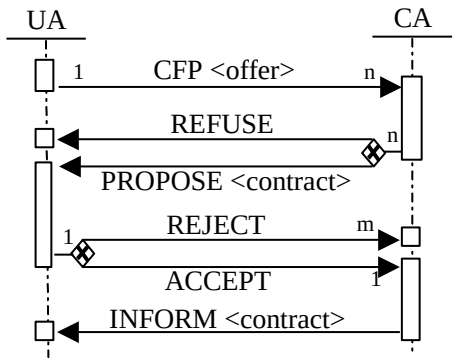


Figure 10.2. Le protocole de négociation

Quand un UA envoie un message de type appel d'offres (call for proposals – CFP) à tous les CA disponibles, le contenu de ce message représente les caractéristiques désirées du contrat : la partie de l'écran occupée par le logo (25%, 50%, 75%, 100% ou *any*), si l'utilisateur doit être averti quand le logo est affiché et si le contrat est pour un court-terme ou un long-terme. Chaque CA recevant ce message décidera s'il veut faire une proposition ou non. Un CA peut modifier les caractéristiques d'une offre, par exemple si le CFP reçu était pour un logo qui occupe 50% de l'écran, la proposition d'un CA peut être pour seulement 25%. L'UA attend une période de temps (ex. : 20 sec) pour recevoir des propositions. A la fin de cette période ou dès que tous les CA ont répondu, l'UA analyse les propositions reçues.

L'UA envoie un message de refus (*reject*) pour toutes les propositions qui ne respectent pas l'offre envoyée avec le message CFP, par exemple un logo occupant 100% de l'écran alors que l'offre initiale était pour 50%, etc. L'UA choisi parmi les propositions restantes celle de plus grande valeur en terme de gain d'argent par seconde par partie d'écran. Il envoie ensuite un message d'acceptation (*accept*) à l'agent CA qui a fait cette proposition et des messages de refus aux autres. Le CA avec la proposition acceptée répondra au message d'acceptation avec un message d'information (*inform*) qui contient les termes du contrat et un identifiant de la transaction qui pourra être utilisé par l'utilisateur (ou son agent) pour recevoir le paiement associé au contrat.

Si aucune proposition n'est acceptée, l'UA n'a pas réussi à créer un contrat. Il peut réessayer tout de suite ou ultérieurement, en fonction des préférences de l'utilisateur. Si un contrat a été créé, l'UA affiche le logo associé sur l'écran comme spécifié par le contrat : sur une partie de l'écran, pour une période de temps, en informant l'utilisateur ou non. Si l'écran n'est pas entièrement occupé par le(s) logo(s), l'UA essaie de créer d'autres contrats pour le reste de l'écran. A la fin de la période spécifiée, le logo est enlevé de l'écran, c.-à-d., le contrat est rempli.

En utilisant ce prototype nous avons été capables de mesurer le temps nécessaire à un UA pour créer un contrat. Ce temps inclue le temps nécessaire pour découvrir des agents/objets dans son voisinage,

celui d'obtention la liste de CA disponibles et le celui de négociation. La Table 1 montre la durée moyenne de chacune de ces phases.

Tableau 10.1. Durée de création des contrats

Phase de création de contrat	**Durée (sec.)**
Découverte d'objets/agents	12-15
Demande d'une liste d'agents	2-3
Négociation de contrat	15-20
Total	**30-38**

La découverte d'objets qui contient des agents inclue la découverte d'objets Bluetooth et la découverte de services sur ces objets. A la fin de cette phase qui dure entre 12 et 15 secondes, le PocketLoox a une liste d'agents proxy existants dans son voisinage. Le volume de données transférées entre les deux objets via Bluetooth étant très petit, leur transfert est rapide. La longue durée du processus de négociation est donc due au temps d'ouverture d'une connexion. Bien que le protocole de négociation soit simple, il spécifie que, pour chaque CA, l'UA envoie plusieurs messages et attend les réponses. Pour chaque message envoyé et reçu une nouvelle connexion est ouverte, la négociation peut ainsi prendre jusqu'à 20 secondes.

L'évaluation du premier prototype montre qu'un agent a besoin de rester entre 30 et 38 secondes dans la proximité d'autres agents pour créer un contrat. Même si cette durée ne semble pas très importante, elle peut décourager l'utilisation des agents qui négocient en utilisant Bluetooth dans les réseaux ad-hoc réels. Si l'utilisateur portant l'objet se déplace à une vitesse normale, il est probable que l'agent sur l'objet n'a pas le temps de découvrir tous les agents autour de lui *et* négocier avec eux avant d'être hors de portée. Des solutions doivent être donc trouvées pour diminuer la durée de la création des contrats.

10.2.2. Améliorations – utilisation d'un agent *broker*

Le temps nécessaire pour créer un contrat n'est pas satisfaisant et des améliorations sont nécessaires. La première que nous avons faite est d'ouvrir une seule connexion pour toute la négociation, et non une pour chaque paire de messages envoyés/reçus. La deuxième a été d'envoyer des messages similaires vers la même destination comme un seul message, et pas comme des messages séparés. Par exemple, l'UA n'envoie plus au début de chaque négociation trois messages *cfp* différents aux trois CA qui se trouvent sur le même objet. Il met ces trois messages dans un seul qu'il envoie au PA se trouvant sur cet objet. Le PA reçoit le message, le divise en trois et envoie chaque partie à son CA destinataire. La troisième amélioration faite a été de modifier légèrement le protocole de négociation. L'UA envoie seulement un message *accept* (s'il accepte une proposition) et n'envoie plus des messages *reject*. Un CA considère implicitement que sa proposition est refusée s'il ne reçoit pas un message *accept*.

Comme nous l'avons mentionné auparavant, le processus de découverte d'objets portables et d'agents prend jusqu'à 15 secondes. C'est à dire qu'un UA passe beaucoup de temps en cherchant autour de lui

des agents avant de négocier avec eux. Nous voulons diminuer ce temps passé dans la phase de découverte d'agents. Ce n'est malheureusement pas possible puisque cette durée est un des paramètres du port Bluetooth. En plus d'imposer un retard dans la création des contrats, la découverte d'objets/agents consomme la batterie de l'objet portable. Pour résoudre ces problèmes, nous avons décidé de changer l'architecture du système. Pour minimiser le temps de découverte et la consommation de la batterie de l'objet portable, l'UA ne cherche plus des objets dans son voisinage. C'est l'agent proxy qui effectue ce processus de découverte. Nous avons ainsi décidé d'échanger les rôles de serveur et client Bluetooth entre l'agent utilisateur et l'agent proxy. Dans cette configuration, l'agent utilisateur s'enregistre comme un service Bluetooth spécifique sur l'objet portable sur lequel il s'exécute. Quand il veut créer des contrats, il met l'objet dans l'état *non-caché*, c.-à-d., l'objet peut être découvert par d'autres objets Bluetooth. Il attend ensuite des connexions entrantes d'autres objets.

Cependant, l'amélioration la plus significative a été faite dans le processus de négociation qui était l'étape la plus longue de la formation d'un contrat. Ainsi, nous avons décidé de transformer l'agent PA en un *broker* (Sycara *et al.*, 1997). Un UA délègue le processus de négociation à l'*agent broker* (BA), qui négocie ensuite pour son compte avec les CA disponibles. L'UA ne demande plus au BA la liste d'agents CA disponibles, mais il lui demande de négocier pour son compte en lui envoyant une offre. Le BA négocie avec les CA exécutés sur le même objet portable ou PC pour créer un contrat. Si la négociation se termine avec succès, le contrat créé est envoyé par le BA à l'UA qui l'exécute.

Le plus grand avantage d'utiliser un agent broker dans notre système est que la négociation des contrats devient plus rapide. Le BA connaît la liste de CA disponibles qui sont en général exécutés sur la même machine. La négociation est alors presque instantanée. Du point de vue de l'agent utilisateur, un contrat est créé en envoyant un seul message et en recevant sa réponse. Un autre avantage est qu'un protocole de négociation plus complexe peut être mis en place. Puisque le BA n'est pas contraint par les ressources limitées d'un objet portable, des protocoles ou des stratégies de négociation plus complexes peuvent être envisagés. En revanche, la délégation du processus de négociation à un tiers présente un désavantage pour l'agent utilisateur. Dans ce cas, l'UA (et son utilisateur) n'a plus de contrôle sur la création des contrats. L'UA doit accepter le contrat créé par le BA à sa place, il n'a aucune possibilité de le refuser. Ceci peut constituer un désavantage parce que l'UA et l'utilisateur doivent faire confiance à l'agent broker. L'utilisateur et son UA cèdent ainsi leur pouvoir de choisir d'accepter des contrats qui sont avantageux, même s'ils ne respectent pas les conditions initiales.

Malgré ces désavantages, l'utilisation d'un broker dans notre scénario diminue la durée de la création des contrats. Le tableau ci-dessous montre que la durée totale de création d'un contrat après les améliorations effectuées est de 12-13 secondes :

Tableau 10.2. Durée de création des contrats après toutes les améliorations

Phase de création de contrat	**Durée (sec.)**
Découverte d'objets/agents	10
Négociation de contrat	2-3
Total	**12-13**

Un autre avantage prototype amélioré est qu'un UA peut être découvert alors qu'il négocie et/ou exécute un contrat. Si nous considérons que plusieurs BA, exécutés sur plusieurs PC, sont distribués aux alentours comme des bornes d'accès, ils peuvent chercher continuellement les objets portables. Quand l'utilisateur se déplace, son UA négocie et exécute des contrats, tout en étant découvert par plusieurs BA. C'est à dire que chaque fois qu'un UA veut créer un contrat, il peut le faire parce qu'il a déjà été découvert par au moins un BA. Idéalement, dans cette situation la phase de découverte peut être ignorée et un contrat peut être créé en 2-3 secondes !

10.2.3. Synthèse

Nous avons implémenté le scénario ADOMO présenté dans ce chapitre et nous l'avons déployé sur des objets portables tels que un PDA et un téléphone portable, comme montrer dans la Figure 10.3 ci-dessous. Pour nous assurer que les agents auront le temps nécessaire de négocier des contrats avant que les utilisateurs se déplacent hors de portée, nous avons du envisager plusieurs solutions. Notamment, les agents utilisateurs peuvent négocier des contrats directement avec des agents commerciaux ou ils peuvent déléguer la tâche de négociation à un agent broker qui négociera pour eux. Dans le premier cas un agent utilisateur garde le contrôle sur le processus de négociation, mais il court le risque de ne pas pouvoir finir la négociation si elle prend trop de temps et son utilisateur se déplace hors de la porté de l'objet portable. Dans le deuxième cas, l'agent utilisateur doit faire confiance à l'agent broker – la négociation est très rapide et donc il a plus de chances de former un contrat, mais il ne peut pas être sûr des termes de ce contrat.

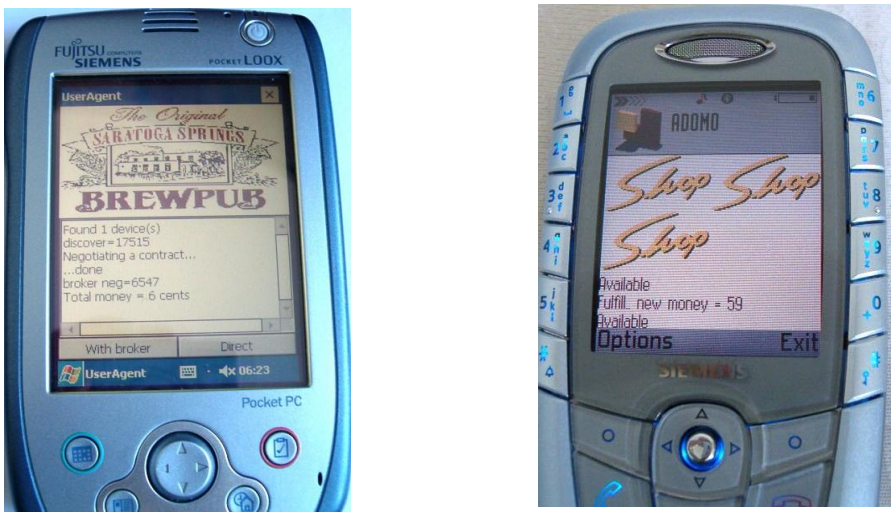


Figure 10.3. ADOMO sur un PDA et un téléphone portable

Dans la section suivante nous décrivons ce choix entre deux possibilités sous la forme d'une structure organisationnelle dans laquelle un agent utilisateur à le choix entre deux rôles et nous décrivons la représentation des contraintes existante à l'aide de notre modèle et le raisonnement *a priori* institutionnel utilisé pour faire le choix.

10.3. Raisonnement à base d'autonomie dans ADOMO

Pour mieux illustrer le raisonnement à base d'autonomie et pouvoir social dans le cadre du scénario ADOMO, nous allons nous intéresser seulement à un aspect de ce scénario et nous allons montrer comment le modèle formel proposé dans cette thèse peut y être appliqué. Par la suite, nous nous focalisons sur la négociation entre les agents pour former des contrats qu'ils exécutent ensuite. Les agents à qui nous nous intéressons seront notés par *ua, ca* et *ba* correspondant aux agents ADOMO décrits dans les sections précédentes (agent utilisateur, commercial et broker). Nous prenons le point de vue d'un agent *ua* qui a un but *g*, celui de gagner de l'argent pour son utilisateur. Le seul plan qui satisfait ce but et qui est connu par tous les agents du système est similaire à celui présenté dans le Chapitre 5. Il est décrit dans la formule suivante.

$$pl_g=\{g, \{negotiate\}, \{display\}, \{\ \}, enables(negotiate, display)$$
$$pl_{negotiate}=\{negotiate, \{\ \}, \{neg_1,neg_2\}, \{\ \}, conflict(neg_1,neg_2)$$

Pour satisfaire son but, un agent *ua* doit ainsi satisfaire le sous-but de négocier, ce qui implique l'exécution en parallèle de deux actions neg_1 et neg_2 – d'où le conflit entre les deux actions. Une négociation réussie définit les termes d'un contrat entre un agent *ua* et un agent *ca*, notamment la somme gagnée par *ua* et la durée de l'action *display* que celui-ci doit effectuer – correspondant à l'action physique d'afficher une publicité sur l'écran de l'objet portable. Dans notre modèle, ce contrat est représenté sous la forme d'un engagement social entre les deux agents pour l'exécution de l'action *display* : l'engagement a une condition de validité négociée (p.ex.: t<100 – date limite de l'exécution 100) et il est créé devant un agent institutionnel que nous notons par *ma* (agent marché) :

$$sc=\langle ua,ca,ma,done(display,ua,ca,10,1\$),t\leq 100\rangle$$

10.3.1. Description de l'organisation ADOMO

Cet agent institutionnel *ma* représente aussi l'organisation que nous utilisons pour modéliser le marché virtuel dans lequel les agents se situent. Cet agent regroupe ainsi les agents *ia* et *oa* présents dans notre modèle, étant à la fois les témoins des engagements sociaux et les représentants de l'organisation. Il existe quatre rôles dans cette organisation, organisation que nous représentons graphiquement comme dans la figure suivante :

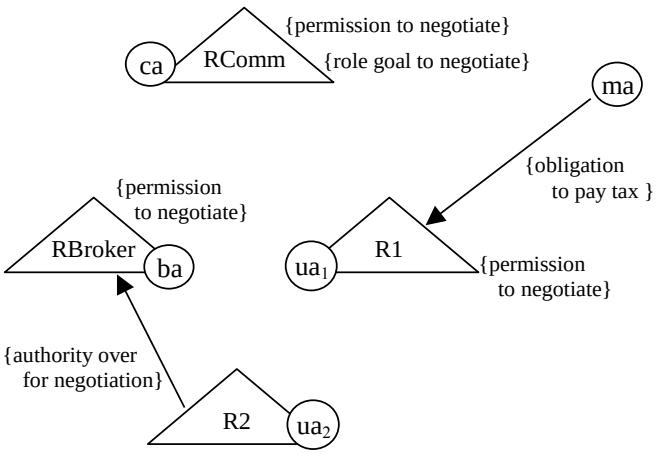


Figure 10.4. Structure de l'organisation ADOMO

Le rôle *RComm* correspond aux agents commerciaux d'ADOMO. Ce rôle est associé à un but de négocier (*negotiate*) et à la permission d'exécuter l'action neg_2. Le rôle *RBroker* correspond à un agent broker d'ADOMO et il a la permission d'exécuter l'action neg_1, tout comme le rôle *R1*. Ce dernier correspond aux agents utilisateurs d'ADOMO qui négocient directement avec les agents commerciaux. Quant aux agents utilisateurs qui négocient via un broker, le rôle correspondant est *R2*, rôle qui n'a pas la permission d'exécuter une action de négociation. Par contre, ce rôle a une relation d'autorité sur le rôle *RBroker*, autorité concernant la satisfaction d'un but de type *negotiate* – autrement dit, un agent jouant le rôle *RBroker* doit négocier pour un agent jouant le rôle *R2* si ce dernier le lui demande. A part cette relation d'autorité, une autre différence entre les rôles *R1* et *RBroker*, qui ont droit tous les deux d'exécuter neg_1, est l'obligation du rôle *R1* envers l'organisation (représentée par l'agent *ma*). Nous utilisons cette obligation de payer un pourcentage de tout contrat négocié pour modéliser le risque couru dans ADOMO de négocier directement un contrat, sans passer par un broker. Vu que statistiquement une partie des négociations vont échouer due à la mobilité de l'utilisateur, nous représentons ceci comme une "taxe" payée pour chaque contrat formé. Enfin, les rôles *R1* et *R2* sont les seuls à avoir la permission d'exécuter l'action *display*, donc les seuls à pouvoir être joués par un *ua.*

L'organisation décrite ci-dessus et les caractéristiques envisagées pour les agents y appartenant nous permettent de représenter un aspect du scénario ADOMO du point de vue du modèle d'autonomie que nous avons proposé. Analysons ainsi les pouvoirs des agents jouant chacun des rôles proposés.

Agent ca *jouant le rôle* RComm

Par le fait qu'il joue le rôle *RComm*, l'agent *ca* est engagé envers l'agent *ma* (l'organisation) de satisfaire le but associé au rôle, celui de *negotiate*. Il poursuit ainsi ce but, mais il ne peut pas exécuter tout seul le plan qu'il connaît pour sa satisfaction. Il n'a donc pas de pouvoir individuel pour satisfaire ce but : le seul pouvoir individuel qu'il a se réfère à l'exécution de l'action neg_2 - il a le pouvoir déontique (venant du rôle) et le pouvoir d'exécution pour cette action.

$$rolecomm=\langle ca,ma,ma,achieve(negotiate),ca:RComm\rangle$$
$$power\_of(ca,neg_2),\ \neg power\_of(ca,negotiate)$$

Agent ba *jouant le rôle* RBroker

Un agent *ba* n'a pas de but propre, mais il sait effectuer l'action neg_1. Ceci, plus la permission reçue du rôle *RBroker* qu'il joue lui donnent le pouvoir individuel pour cette action. La relation d'autorité associée au rôle joué est représentée dans notre modèle par une politique sociale. Comme c'est la seule relation d'autorité que nous envisageons dans cet exemple, nous utilisons la notation *autority_broker* pour la désigner, afin de simplifier nos formules:

$$power\_of(ba,neg_1),\quad \neg\, power\_of(ba,negotiate)$$
$$authority\_broker=\langle ba,ma,ma,\alpha,ba:RBroker\rangle\in SComm_{ma}, \text{où } \alpha=\forall x\in Ag, x:R2\wedge$$
$$request(x,ba,achieve(negotiate,\beta))\Rightarrow\exists\langle ba,x,ma,achieve(negotiate),\beta\rangle\in SComm_{ma}$$

Agent ua_1 *jouant le rôle* R1

Similaire à un agent *ba*, un agent ua_1 jouant le rôle *R1* a le pouvoir individuel pour l'action neg_1. Il a le but propre *g*. Etant donné le seul plan connu, il tente de satisfaire le sous-but *negotiate*. Cet agent se retrouve ainsi dans une situation similaire à l'agent *ca* : il a le pouvoir individuel pour seulement une des deux actions correspondant au but *negotiate*. L'agent ua_1 a aussi le pouvoir individuel pour l'exécution de l'action *display*.

$$power\_of(ua_1, neg_1),\ power\_of(ua_1, display),\ \neg power\_of(ua_1, negotiate)$$

De plus, cet agent est sujet d'une obligation de payer une taxe quand un contrat est créé, obligation représentée dans notre modèle par une politique sociale. Comme c'est la seule obligation que nous envisageons dans cet exemple, nous utilisons la notation *obligation_tax* pour la désigner, afin de simplifier nos formules :

$$obligation\_tax = \langle ua_1, ma, ma, \alpha, ua_1{:}R1\rangle \in SComm_{ma}$$

où $\alpha = \forall sc \in SComm_{ma}\ sc = \langle ua_1, x, ma, done(display, ua_1, x, dur, sum), \beta\rangle \Rightarrow$

$$\exists \langle ua_1, ma, ma, done(pay, ua_1, ma, sum/10), ua_1{:}R1\rangle \in SComm_{ma}$$

Agent ua_2 *jouant le rôle* R2

Un agent ua_2 a les même caractéristiques qu'un agent ua_1. Il a le même but *g* (et ainsi *negotiate*), mais il n'a pas le pouvoir individuel pour ce but. En plus, vu que le rôle *R2* qu'il joue n'a pas la permission d'exécuter l'action neg_1, il n'a même pas le pouvoir individuel pour cette action. Cependant, il peut profiter de l'existence de la politique sociale représentant une relation d'autorité, *authority_broker*, comme nous le verrons par la suite.

$$\neg power\_of(ua_2, neg_1),\ \ power\_of(ua_2, display),\ \ \neg power\_of(ua_2, negotiate)$$

10.3.2. Evaluation de rôle dans ADOMO

Considérons maintenant le cas d'un agent de type *ua* qui a le choix entre les rôles *R1* et *R2*. Notre approche consiste à permettre à l'agent d'évaluer les pouvoirs et l'autonomie qu'il gagne ou perd en jouant chacun de ces rôles. Pour simplifier notre description, nous analysons les pouvoirs sociaux et l'autonomie de deux agents, ua_1 et ua_2, qui jouent respectivement les deux rôles. Un agent ayant le choix pourrait ensuite comparer les différences entre les deux cas et prendre une décision informée.

Suite à son manque de pouvoir pour le but *negotiate* et donc pour le but *g*, l'agent ua_1 dépend de l'agent *ca* pour satisfaire ses buts. A son tour, *ca* dépend aussi de lui pour le même but *negotiate*, nous pouvons donc parler d'une dépendance mutuelle (Formule 7.8) :

$$mut\_depends(ua_1, ca, negotiate),\ dep\_power\_over(ua_1, ca, negotiate),$$

La dépendance de *ca* envers lui donne à l'agent ua_1 un pouvoir sur *ca* et, à partir de la Formule 7.17, un pouvoir indirect faible pour satisfaire le but *negotiate*. Cependant, l'agent *ca* a aussi du pouvoir sur lui (suite à la dépendance mutuelle), ce qui limite l'autonomie de décision de l'agent ua_1 (Formule 7.18) :

$$weak\_ind\_power(ua_1, ca, negotiate), \neg has\_autonomy(ua_1, ca, negotiate)$$

De plus, suite à l'obligation de payer une taxe, l'organisation représentée par l'agent *ma* a du pouvoir sur l'agent ua_1 ce qui limite encore plus son autonomie :

$$\neg has\_autonomy(ua_1, ma, pay),$$

Quant à l'agent ua_2, il manque aussi le pouvoir individuel pour son but *negotiate* et donc *g*. Cependant, la présence de la relation d'autorité sur l'agent *ba* lui confère un pouvoir sur cet agent, donc un pouvoir indirect faible de satisfaire le but *negotiate* :

$$weak\_ind\_power(ua_2, ba, negotiate)$$

Si nous devons comparer les deux cas, des agents *ua* jouant respectivement les rôles *R1* et *R2*, il semble évident que le rôle *R2* est mieux pour un agent qui a comme motivation de gagner le plus d'argent pour son utilisateur dans cette organisation. Dans les deux cas les agents ont seulement un pouvoir indirect faible pour satisfaire leur but, sauf que l'agent jouant le rôle *R1* n'a pas d'autonomie par rapport à un autre agent et par rapport à l'organisation. Si la première perte d'autonomie n'est pas grave (il n'a pas d'autonomie pour un but qui fait partie de ses plans), la deuxième l'est. L'organisation a du pouvoir sur lui pour le faire payer de l'argent, ce qui vient contre ses motivations – ce manque d'autonomie peut coûter cher à l'agent.

Cependant, avant de trancher entre les deux rôles, nous voulons souligner que le pouvoir indirect que l'agent ua_1 a grâce à l'agent *ca* a plus de chances de mener à la satisfaction du but *negotiate*, vu qu'il est du à une dépendance mutuelle : l'agent *ca* est intéressé à satisfaire le même but aussi. De l'autre coté, l'agent *ba* peut décider de se retrouver dans un comportement autonome par rapport à l'organisation et ne pas adopter la délégation du but *negotiate* de la part de l'agent ua_2 – ce qui empêchera ce dernier de satisfaire ses buts. Cette situation peut être remédiée par l'introduction d'une sanction implicite pour un agent *ba* qui n'obéit pas à la relation d'autorité, comme moyen de rassurer les agents jouant le rôle *R2* que leurs buts seront satisfaits.

10.3.3. Réalisation en Prolog

Comme notre modèle est fondé sur la logique des prédicats, son implémentation a pu être facilement faite dans Prolog[11]. Des prédicats sont utilisés pour décrire l'état du monde et l'application, comme par exemple quels sont les agents, les actions, buts, etc. qu'ils connaissent ou qu'ils ont. Les engagements sociaux existants sont aussi représentés sous la forme des prédicats qui sont rajoutés ou supprimés en fonction de l'évolution du système. La présence de ces prédicats-engagements nous permet de représenter les rôles et les contraintes associées facilement (voir Chapitre 6). Ensuite, tous les concepts que nous avons introduits formellement dans le Chapitre 7, les pouvoirs d'exécution, déontique, individuel, les relations de dépendance et de pouvoir social et l'autonomie des agents sont facilement inférés dans Prolog. Nous avons ainsi réalisé un module Prolog, décrit dans l'annexe, qui, à partir d'une description du monde, des agents et d'une organisation, nous permet d'identifier les pouvoirs et l'autonomie des agents jouant des rôles.

[11] L'implémentation de Prolog utilisé est SWI-Prolog www.swi-prolog.com

Dans la réalisation Prolog nous avons été fidèles à notre approche qui considère que ce modèle d'autonomie est une valeur rajoutée dans le raisonnement d'un agent, sans pour ceci remplacer des modèles de base existants. Nous avons ainsi réalisé des agents Java qui agissent et interagissent les uns avec les autres d'une manière classique (envoi des messages, etc.). Cependant, chaque agent est équipé d'un moteur d'inférence Prolog qui utilise le module décrit ci-dessus et qui implémente donc le modèle que nous avons proposé dans ce manuscrit. Quand un agent est dans une situation d'un raisonnement *a priori* ou *a posteriori*, donc dans le besoin de raisonner en terme d'autonomie, il utilise ce module pour pouvoir prendre une décision informée. Ceci correspond à notre approche qui envisage de rajouter un module de raisonnement à base d'autonomie dans l'architecture d'un agent existant, sans avoir besoin de concevoir de zéro un tel agent ou un système multi-agents. La Figure suivante décrie l'architecture d'un agent réalisé, tandis que l'implémentation Prolog est décrite dans l'annexe de ce manuscrit.

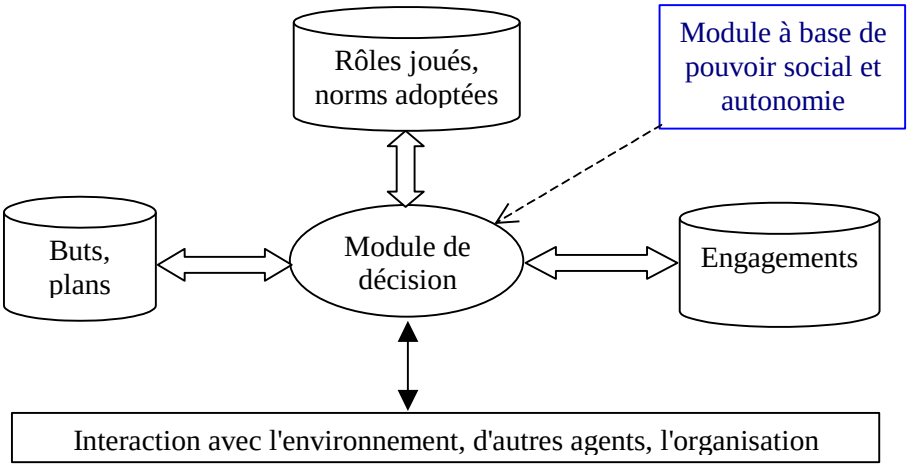


Figure 10.5. Schéma architectural d'un agent ADOMO

10.4. Synthèse

Dans ce chapitre nous avons décrit le scénario d'intelligence ambiante appelé ADOMO. Dans ce scénario, les agents des utilisateurs entrent dans des marchés virtuels afin de négocier des contrats de publicité pour leurs utilisateurs. Nous avons décrit comment de tels agents ont été déployés sur des téléphones portables ou des PDA ainsi que la structure des marchés implémentés. Un agent doit ainsi entrer dans une nouvelle organisation (un marché) où il joue un rôle qui lui permet de satisfaire ses motivations – les désirs de l'utilisateur. Des soucis techniques et de performance nous ont mené à offrir deux rôles possibles aux agents, chacun avec ses avantages et inconvénients.

Un agent dans le scénario ADOMO doit choisir entre deux rôles dans une organisation (marché virtuel), ce qui constitue un exemple d'un raisonnement *a priori* institutionnel. Nous avons décrit comment ce raisonnement est effectué, c.-à-d., comment l'agent évalue les deux rôles en terme des pouvoirs et autonomie gagnés ou perdus et comment le choix de rôle est effectué. Ce raisonnement est possible grâce au modèle d'autonomie proposé dans cette thèse. Le modèle à été implémenté sous la forme des prédicats Prolog, ce qui nous permet d'équiper facilement les agents d'un raisonnement à base d'autonomie. Le scénario présenté dans ce chapitre, ainsi que l'implémentation Prolog proposée ne sont que des exemples de l'utilité du raisonnement à base d'autonomie et de l'utilisation de notre modèle dans un tel raisonnement, d'autres exemples peuvent être envisagés.

Synthèse

L'objectif principal de cette thèse était de proposer un raisonnement à base d'autonomie dans les systèmes ouverts et dynamiques. Plus précisément, ce travail se proposait de montrer comment l'utilisation de la notion d'autonomie peut enrichir des différentes formes de raisonnement. Pour ceci, nous avions besoin d'un modèle formel qui permet de définir l'autonomie sociale en décision des agents, dans ses différents aspects, ce qui constituait un objectif secondaire de cette thèse. Ce dernier objectif à été atteint dans la partie précédente, tandis que le premier objectif a été traité dans cette partie.

Dans notre approche, être autonome signifie pour un agent d'être dans un comportement qui n'obéit pas à une contrainte comportementale. Comme nous l'avons vu dans les chapitres précédents, le raisonnement *a posteriori* d'un agent concerne la décision d'un agent d'être dans le comportement qu'il est contraint d'être. Autrement dit, ce raisonnement d'un agent porte sur le fait d'être ou ne pas être autonome, d'obéir ou non aux contraintes comportementales. Même si dans la littérature plusieurs approches ont été proposées pour ce type de raisonnement (cf. Chapitre 4), comme par exemple le calcul d'utilité, notre approche a l'avantage de baser ce raisonnement sur le même concept – autonomie – que le raisonnement *a priori* (voir ci-dessous). En plus, l'utilisation de ce concept permet de cacher la nature de la contrainte qui limite le comportement, qu'elle soit interpersonnelle ou institutionnelle.

Avoir de l'autonomie signifie pour un agent que sa prise de décision n'est pas influencée par une source externe, autrement dit l'agent est libre de prendre les décisions qui correspondent au mieux avec ses motivations. Le raisonnement *a priori* d'un agent concerne une évaluation de l'agent des différentes situations dans lesquelles il peut se retrouver afin de choisir celle avec le plus grand potentiel pour satisfaire ses motivations. Autrement dit, le raisonnement *a priori* permet à un agent d'évaluer dans quelle situation il aura le plus d'autonomie, il pourra décider le plus librement. Dans les précédents chapitres nous avons analysé plus attentivement ce raisonnement et son utilisation d'un modèle à base d'autonomie.

Nous considérons que le raisonnement *a priori* institutionnel, comme par exemple le choix d'un rôle à jouer dans une organisation, représente un raisonnement très utile dans le cadre général de notre travail, celui des systèmes ouverts et dynamiques. Nous avons ainsi analysé ce que signifie pour un agent de jouer un rôle, quelles sont les implications pour l'autonomie de l'agent de jouer ou ne pas jouer un rôle. A l'aide d'une application d'intelligence ambiante qui se situe dans notre cadre général, nous avons illustré comment le modèle de pouvoir social et autonomie proposé dans la partie précédente peut être utilisé par un agent pour choisir entre deux rôles qu'il peut jouer.

Conclusions et perspectives

Conclusions

Cette thèse se situe dans le cadre général des systèmes ouverts et dynamiques, tels que, par exemple, les applications d'intelligence ambiante. Comme l'illustre par exemple le scénario d'intelligence ambiante que nous avons présenté, ADOMO, dans une telle application les agents peuvent entrer et sortir de différents systèmes multi-agents. Ils peuvent également, à l'intérieur d'un système, changer de rôle, changer de groupe. Dans ce cadre général, il nous apparaît important qu'un agent soit capable de comprendre lequel des systèmes disponibles lui permettra de mieux atteindre ses objectifs. Autrement dit, nous voulons permettre aux agents d'évaluer les différentes possibilités mises à leur disposition et de pouvoir choisir celle qui convient le plus à leurs objectifs.

Pourquoi considérer l'autonomie ?

Le problème que nous essayons de résoudre que l'on peut exprimer en terme d'évaluation des différentes actions possibles, n'est pas trivial. Deux raisons essentielles peuvent être évoquées. En premier lieu, il n'est pas facile de prendre en considération tous les aspects d'un système ou au moins ceux qui sont pertinents pour un agent. De plus, il existe une grande diversité de concepts et de mécanismes qui règlent le fonctionnement d'un système multi-agent. Il n'est pas facile de les traduire tous dans un concept qui permettra une telle évaluation. En second lieu, il est souvent impossible pour un agent d'envisager tout ce qui peut ce passer dans un système multi-agents avant même d'y entrer – il est impossible de connaître le comportement futur des autres agents ou comment les changements de contexte se réflèteront dans le système.

Pour répondre à ces deux problèmes nous avons dû nous intéresser à ce qu'un agent doit évaluer ou prendre en considération avant de décider d'entrer ou non dans un système. La réponse à cette question est donnée par les mécanismes de coordination qui existent dans un système multi-agents et qui ont comme rôle d'assurer que le système a le comportement souhaité par son concepteur, indifféremment des agents qui le composent. Nous avons argumenté que l'existence de la coordination au niveau d'un système implique une existence d'un contrôle sur le comportement des agents, souvent sous la forme d'une influence externe aux agents. Autrement dit, le besoin de coordination dans un système multi-agents implique la présence de contraintes qui limitent la prise de décision et le comportement des agents. De plus, notre passage en revue de la littérature nous a montré que le concept d'autonomie est parfois lié à ces contraintes. *L'autonomie sociale de décision d'un agent – donc l'autonomie qu'un agent a pour prendre des décisions dans un contexte social – peut être considérée comme complémentaire aux contraintes imposées à l'agent par le besoin de coordination dans le système*.

L'idée que nous avons promue au cours de cette thèse et qui constitue la principale valeur ajoutée de ce travail est de *permettre aux agents de conduire un raisonnement à base d'autonomie*, c'est-à-dire d'évaluer les différentes séquences d'actions possibles, comme par exemple dans quel système entrer, en fonction de l'autonomie qu'ils auront dans chacun d'entre eux. Cependant, nous rencontrons un autre problème : le concept d'autonomie, malgré son utilisation et sa présence dans presque toutes les définitions d'agents, n'a pas une définition communément admise ou formalisée. Pour qu'il soit utilisé

par les agents dans leur raisonnement, nous devons d'abord le définir et proposer un modèle formel de ce concept.

Modèle d'autonomie et pouvoir social

Comme l'autonomie sociale de décision d'un agent caractérise la prise de décision d'un agent qui est limitée ou non par des contraintes provenant des sources sociales externes (d'autres agents, une organisation multi-agents, etc.), l'approche que nous avons envisagée a été de baser notre proposition de définition de l'autonomie sur ces contraintes. A leur tour, ces contraintes sont en provenance de mécanismes de coordination présents dans un système. Un passage en revue des différentes approches envisagées dans la littérature pour assurer la coordination dans un système multi-agents nous a permis d'illustrer leur grande diversité. Nous avons pu mettre en avant deux grandes catégories de contraintes sociales imposées aux agents : celles provenant des autres agents (interpersonnelles) et celles provenant des organisations/institutions multi-agents (institutionnelles). Parmi les notions rencontrées dans la littérature, nous avons pu identifier des concepts de référence, tels que rôle ou norme – des concepts institutionnels – ou engagements sociaux, dépendances ou bien pouvoirs sociaux – des concepts interpersonnels.

Pour définir l'autonomie sociale de décision d'un agent, nous avons du définir les contraintes qui limitent cette autonomie, tâche compliquée par la grande diversité des concepts utilisés pour ces contraintes. Cependant, nous avons pu proposer un modèle formel qui utilise des engagements sociaux pour modéliser des contraintes interpersonnelles en s'appuyant sur des notions de base telles que action, but, plan ou permission. Nous avons ainsi proposé une représentation des engagements sociaux qui nous permet de modéliser les contraintes sur le comportement des agents qui sont issues des interactions avec les autres agents. Nous avons ensuite utilisé la même représentation d'engagements sociaux (et des méta-engagements ou politiques sociales) pour modéliser les normes, les relations d'autorité et autres contraintes qui limitent le comportement ou la prise de décision des agents situés dans un contexte institutionnel. Finalement, nous avons modélisé le fait qu'un agent joue un rôle dans une organisation à travers les engagements sociaux aussi.

Malgré la puissance de représentation de ce concept d'engagement social, il ne nous suffit pas : il permet de représenter des contraintes sur le comportement des agents, mais il y a d'autres types de contraintes qui nous intéressent, notamment celles qui poussent un agent à interagir avec d'autre ou à appartenir à une organisation. Pour modéliser ces contraintes nous nous sommes appuyés sur la théorie du pouvoir social qui inclut la théorie de la dépendance. Nous avons ainsi modélisé les pouvoirs individuels d'un agent : pouvoir d'exécution et déontique et le manque de pouvoir qui crée des relations de dépendance entre les agents. L'existence de ces relations, ainsi que les contraintes institutionnelles (normes, liens d'autorité) représentées dans notre modèle par des engagements sociaux, constituent des sources de création d'un lien de pouvoir d'un agent sur un autre et des pouvoirs indirects des agents.

Nous avons ainsi pu aboutir à un modèle de pouvoir social et des engagements sociaux. Ce modèle nous a permis de définir formellement le fait qu'un agent a de l'autonomie pour prendre une décision (par le fait qu'un autre agent n'a pas de pouvoir sur lui) ou qu'un agent se trouve dans un comportement autonome (par le fait qu'il a violé un engagement social).

Ce modèle formel proposé décrit dans ce manuscrit représente une des contributions majeures de cette thèse, surtout en considérant le manque de définitions formelles de l'autonomie dans la littérature. Ce modèle peut être vu comme une pyramide : à la base se trouve la diversité des concepts décrivant les relations entre un agent et son environnement (action, plan), avec d'autres agents (interactions) ou avec une organisation (rôles, normes). Ces concepts sont graduellement unifiés en utilisant des concepts de plus en plus généraux et de haut niveau, pour arriver au sommet avec l'utilisation d'un seul concept : l'autonomie. Nous sommes conscients qu'en utilisant cette approche nous ignorons des détails, par exemple nous considérons qu'un agent n'a pas d'autonomie par rapport à un autre agent, mais sans forcement représenter si ce manque d'autonomie provient d'une action qu'il ne sait pas exécuter ou d'une autorité que l'autre a sur lui. Cependant ceci ne constitue pas un problème dans notre cas si nous considérons l'utilisation envisagée pour ce modèle que nous décrivons par la suite.

Raisonnement à base d'autonomie

Le modèle à base de pouvoir social et d'autonomie que nous avons proposé dans cette thèse a été conçu dans le but d'enrichir le raisonnement des agents. Nous ne nous sommes pas fixés comme objectif de s'intéresser à la conception des systèmes multi-agents (ouverts ou non), ni à la conception des agents capables de fonctionner dans ces systèmes. Nous considérons que de tels agents et systèmes existent et fonctionnent comme prévu. Nous voulons tout simplement améliorer un aspect des performances des agents. Ce que nous proposons dans cette thèse est d'équiper les agents d'un module supplémentaire, un module à base de notre modèle de pouvoir social et d'autonomie. L'agent continue de fonctionner comme son concepteur l'a envisagé, tandis que ce module analyse les pouvoirs et l'autonomie de l'agent dans diverses situations. Cette analyse, ou bien cette évaluation des situations présentes ou potentielles, n'est pas toujours utile, mais il existe des cas où nous argumentons qu'elle a une grande utilité.

Notre modèle permet de regrouper sur le même concept*, être dans un comportement autonome*, le fait qu'un agent désobéit à des contraintes existantes qui limitent son comportement. Indifféremment du type de la contrainte, qu'elle soit en provenance d'une interaction avec un autre agent (p.ex. l'agent a adopté un but de l'autre) ou en provenance d'une organisation (p.ex. l'agent a une obligation parce qu'il joue un rôle), le fait de désobéir à la contrainte est représenté dans notre modèle par le fait que l'agent se trouve dans un comportement autonome. Ceci permet un raisonnement unifié sur les conséquences d'un tel comportement, raisonnement que nous appelons *a posteriori* parce qu'il est nécessaire *après* que la contrainte soit établie.

Notre modèle permet aussi de représenter le fait qu'un agent ne peut pas prendre une décision sans être influencé par un autre agent, c.-à-d., le fait qu'il n'a pas d'autonomie de décision. Ceci est très utile par exemple dans le choix d'un partenaire d'interaction : un agent devrait choisir comme partenaire un agent qui n'a pas d'autonomie par rapport à lui, pour augmenter ainsi les chances que l'autre prenne la décision désirée. Un autre exemple est le choix qu'un agent doit faire entre deux rôles : en règle générale, un agent choisira le rôle qui lui confère plus d'autonomie, ce qui signifie qu'il pourra prendre plus de décisions comme il veut. Il aura donc plus de chances d'atteindre ses objectifs. Nous appelons ce type de raisonnement utilisé pour choisir un partenaire *avant* l'interaction ou un rôle *avant* de le jouer, un raisonnement *a priori.*

Le choix entre deux rôles présenté ci-dessus est un raisonnement *a priori* sur un concept *institutionnel* et représente en faite l'objectif initial que nous nous étions fixé dans cette thèse. Nous voulions permettre à un agent d'évaluer les conséquences d'entrer ou de sortir d'un système, de changer de rôle dans un système ou de faire le choix entre deux rôles, groupes, systèmes, etc. Tout ceci représente un raisonnement *a priori* institutionnel et cette évaluation peut être faite en terme de pouvoir et d'autonomie. Un agent peut ainsi se représenter quels sont les pouvoirs et l'autonomie qu'il gagne et qu'il perd en fonction des possibilités qui lui sont offertes pour prendre ensuite une décision plus informée.

Pour illustrer ces propos nous avons implémenté un scénario d'intelligence ambiante (fidèles à notre cadre général d'application) dans lequel des agents intelligents sont déployés sur des PDA ou des téléphones portables. Les différentes contraintes techniques et de réalisation ont conduit à la création d'une organisation multi-agents dans lequel un agent a le choix entre deux rôles. Nous avons ensuite décrit comment ces deux rôles sont évalués par un agent pour voir celui qui lui offre le plus des pouvoirs et d'autonomie, c.-à-d., le raisonnement *a priori* institutionnel des agents dans ce scénario.

Perspectives

Si nous devons résumer les conclusions présentées ci-dessus, nous pouvons identifier trois dimensions différentes auxquelles cette thèse s'intéresse et des extensions sont envisageables sur chacune de ces dimensions :

- le problème de permettre aux agents de choisir dans quel système entrer ou quel rôle jouer et l'idée de baser ce choix sur le concept d'autonomie
- un modèle formel qui regroupe et lie les concepts d'engagement social, pouvoir social et autonomie sociale de décision débouchant ainsi sur une définition formelle de celle-ci
- utilisation de ce modèle dans le raisonnement des agents, pour leur permettre ainsi d'évaluer les pouvoirs et l'autonomie qu'ils perdent ou gagnent afin de prendre une décision plus informée

En prenant en considération les nouveaux domaines d'application des systèmes multi-agents, tels que ceux de l'intelligence ambiante, nous sommes convaincus que le problème général adressé par cette thèse, celui de permettre aux agents de faire un choix informé sur le système dans lequel entrer, reste très important, malgré les travaux similaires peu nombreux qui s'y intéressent. Nous avons fait le choix argumenté de baser cette décision des agents sur le gain ou la perte d'autonomie – d'autres solutions peuvent être envisagées. Par exemple, un agent peut utiliser un simple calcul d'utilité pour évaluer les possibilités qu'il a, même si nous restons sceptiques par rapport à cette solution, vu qu'il n'est pas facile de calculer l'utilité des situations potentielles. Cependant, pour revenir au problème attaqué par cette thèse, nous aimerons bien que dans le futur de plus en plus d'applications prennent le point de vue d'un agent, et non seulement celle d'un système. Autrement dit, nous considérons qu'il est important de s'intéresser à la conception des agents aussi : si un système multi-agents ouvert est conçu et spécifié, s'intéresser plus à comment améliorer les performances des agents au sein du système, pas seulement les performances du système.

Comme nous avons voulu définir le concept d'autonomie d'une manière utilisable par les agents artificiels, nous avons utilisé des concepts tels que engagement social, dépendance, pouvoirs individuels et sociaux que nous avons regroupés dans un modèle formel. Ce modèle peut être remplacé par un modèle similaire basé sur d'autres concepts, sans que cela change notre approche. De plus, ce modèle peut être également enrichi pour prendre en compte des aspects que nous avons ignorés. Nous avons essayé de marquer au cours du manuscrit les hypothèses faites et qui peuvent être relaxées dans les extensions de ce modèle. Par exemple, la possibilité d'obtenir des permissions de la part d'un autre agent, et pas seulement de l'organisation, ou de représenter le fait qu'en jouant un rôle un agent est vu différemment par d'autres agents (l'effet *count-as*). De plus, d'autres formalisations de la théorie du pouvoir social sont proposées dans la littérature et un modèle similaire au notre mais à base d'une autre formalisation que la logique des prédicats peut être envisagé. Autrement dit, la modélisation que nous avons proposée est adaptée à nos besoins, par exemple à celui de permettre à un agent d'évaluer ce que signifie pour lui de jouer un rôle, une modélisation différentes pourrant être nécessaire dans un autre objectif.

Nous avons illustré l'utilisation et la mise en œuvre du modèle proposé dans le raisonnement des agents, notamment nous avons discuté comment différentes formes de raisonnement peuvent bénéficier d'une représentation explicite des pouvoirs et de l'autonomie d'un agent. Ceci ne constitue qu'un exemple d'utilisation, exemple provenant du problème principal auquel nous nous sommes intéressés, celui des agents désirant entrer dans un système (organisation, groupe, etc.) ouvert. D'autres utilisations dans d'autres cas d'applications peuvent être envisagés, ce qui nous permettra de faire évoluer le modèle proposé pour l'adapter à ces nouvelles utilisations. Une extension possible que nous avons mentionnée est aussi l'utilisation du modèle *par une organisation* afin d'évaluer lequel des agents candidats pour un rôle est le mieux adapté pour le jouer. Une autre extension possible concerne le raisonnement a posteriori d'un agent, celui d'être ou non autonome, de violer ou non une contrainte. Un modèle tel que le notre permet de rendre explicite une telle violation et éventuellement les sanctions associées, mais nous ne sommes intéressés à ce problème que superficiellement. Même si ceci sort des objectifs de notre travail, nous considérons que ce type de raisonnement est très important dans la conception des agents intelligents et qu'il mérite d'être analysé plus attentivement. De plus, les essais d'adapter notre modèle à ce problème ne feront que nous permettre d'enrichir ce modèle en adressant ses limitations.

Pour résumer, nous sommes conscients que le modèle proposé est limité et qu'il peut être enrichi – il a était conçu avec une utilisation précise comme objectif. De plus, nous sommes conscients aussi que l'utilisation envisagée dans cette thèse pour un modèle formel d'autonomie n'est pas la seule d'intérêt dans les systèmes multi-agents. Cependant, nous restons convaincus que le problème adressé dans cette thèse reste important et que l'approche proposée, comprenant un modèle formel d'autonomie et son utilisation, est une solution adaptée à ce problème.

Bibliographie

Ashri,R.; Luck, M. and d'Inverno, M. 2003. On Identifying and Managing Relationships in Multi-Agent Systems. In Proceedings of the 18th International Joint Conference on Artificial Intelligence – IJCAI'03, Acapulco, Mexico.

Barbuceanu, M. and Fox, M.S. 1995. COOL: A Language for Describing Coordination in Multi-Agent Systems. In *Lesser, V. (ed.): Proceedings of the First International Conference on Multi-Agent Systems*, AAAI Press/MIT Press, p.17-25.

Beavers, G. and Hexmoor, H. 2004. Types and Limits of Agent Autonomy. In *Nikles, M.* et al. *(eds.): Agents and Computational Autonomy: Potentials, Risks and Solutions*. LNAI 2969, Springer-Verlag Berlin, p.95-102.

Boella, G. and van der Torre, L. 2004a. Contracts as legal institutions in organizations of autonomous agents. In *Proceedings of the 3rd Autonomous Agents and Multiagent Systems Conference – AAMAS'04*, ACM Press.

Boella, G. and van der Torre, L. 2004b. Attributing mental attitudes to roles: The agent metaphor applied to organizational design. In Proceedings of ICEC'04.

Boella, G.; Sauro, L. and van der Torre, L. 2004. Power and Dependence Relations in Groups of Agents. In Proceedings of the 2004 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Beijing, p.1-7.

Boissier, O. 2001. Modèles et architectures d'agents. In *Briot, J.B. and Demazeau, Y. (eds.): Systèmes Multi-Agents*, Editure Hermes.

Boissier, O. 2003. *Contrôle et Coordination orientées multi-agent*. Memoire HdR, ENS Mines de Saint-Etienne, France.

Bonnet, G. and Tessier, C. 2007. In *Proceedings of the 6th Autonomous Agents and Multiagent Systems Conference – AAMAS'07*.

Bourne, R.; Excelente-Toledo, C. and Jennings, N. 2000. Run-time selection of coordination mechanisms in MAS. In *Proceedings of the 14th European Conference on Artificial Intelligence – ECAI'00*, Berlin, p.348–352.

Brainov, S. and Hexmoor, H. 2001. Quantifying Relative Autonomy in Multiagent Interaction. In Proceedings of the IJCAI-01 workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents, Seattle, WA, p.27-35.

Brazier, F.M.T.; Jonker,C. and Treur, J. 1999. Compositional Design and Reuse of a Generic Agent Model. In Gaines, B. and Munses, M. (eds.): Proceedings of the Knowledge Acquisition Workshop – KAW'99, Banff.

Broersen, J. *et al.* 2001. The BOID architecture. In Proceedings of the 5th International Conference on Autonomous Agents, Montreal.

Brooks, R. and Connel, J.H. 1986. Asynchronous Distributed Control System for a mobile robot, *SPIE 727 Mobile Robots*.

Castelfranchi, C. 1995a. Commitments: From Individual Intentions to Groups and Organizations. In *Lesser, V. (ed.): Proceedings of the International Conference on Multi-Agent Systems*, AAAI Press/MIT Press, p.41-48.

Castefranchi, C. 1995b. Guarantees for Autonomy in Cognitive Agent Architectures. In *Jennings, N. and Wooldridge, M. (eds.): Agent Theories, Architectures and Languages*, Spinger-Verlag.

Castefranchi, C. 1998. Modelling social action for AI agents. *Artificial Intelligence*, Vol 103, p.157-182.

Castefranchi, C. 2000. Founding Agent's 'Autonomy' On Dependence Theory. In *Proceedings of the 14th European Conference on Artificial Intelligence*, IOS Press, p.353-357.

Castelfranchi, C. 2002. A micro and macro definition of power. *ProtoSociology – An International Journal of Interdisciplinary Research*, Vol 18-19, p.208-268.

Castelfranchi C. 2005. Formalising the informal?. In *Nordic Journal of Philosophical Logic*.

Castelfranchi, C.; Dignum, F.; Jonker, C. and Treur, J. 2000. Deliberate Normative Agents: Principles and Architectures, In *Jennings, N. and Lesperance, Y. (eds.): Intelligent Agents VI*, LNAI 1757, Springer-Verlag, p.364-378.

Chopinaud, C. 2007. *Contrôle de l'émergence de comportements dans les systèmes d'agents cognitifs autonomes*. PhD thesis, LIP6, Paris, France.

Cohen, P. and Levesque, H. 1990. Intention is Choice with Commitment, *Artificial Intelligence*, Vol. 42, p.213-261.

Conte, R. and Castelfranchi, C. 1999. From conventions to prescriptions. Towards a unified view of norms. *Artificial Intelligence and Law*, Vol 7 (3), p.23-40.

Conte, R.; Castelfranchi, C. and Dignum, F. 1999. Autonomous Norm-Acceptance. In *Muller, J.* et al. *(eds.), Intelligent Agents V*, LNAI 1555, Springer-Verlag, p.319-334.

Dastani, M.; Dignum, V. and Dignum, F. 2003. Role-Assignment in Open Agent Societies. In *Proceedings of the 2nd Autonomous Agents and Multiagent Systems Conference – AAMAS'03*, ACM Press.

Demolombe, R. and Louis, V. 2006. Norms, Powers and Roles : Towards a Logical Framework. In *Esposito, F.* et al. *(eds.) : Foundations of Intelligent Systems*. LNCS 4203, Springer-Verlag Berlin, p.514-523.

Dignum, F. 1999. Autonomous Agents with Norms. In *Artificial Intelligence and Law*, Vol. 7(3), p.69-79.

Dignum, F.; Kinny, D. and Sonenberg, E. 2001. Motivational Attitudes of Agents: On Desires Obligations and Norms. In Proceedings of the 2nd International Workshop of Central and Eastern Europe on MAS, p.61-70.

Dignum, V. 2004. *A Model for Organizational Interaction: based on Agents, founded in Logic*. SIKS Dissertation Series 2004-1, SIKS. PhD thesis.

Dignum, V. ; Meyer, J-J. and Weigand, H. 2002. Towards an Organisational Model for Agent Societies Using Contracts. In *Proceedings of the 1st Autonomous Agents and Multiagent Systems Conferences – AAMAS'02*, ACM Press, p.694-695.

Drogoul, A. and Ferber, J. 1992. Multi-Agent Simulation as a Tool for Modeling Societies: Application to Social Differentiation in Ant Colonies. *Decentralized Artificial Intelligence*, Vol. 4.

Dunin-Keplicz, B. and Verbrugge, R. 2002. Collective Intentions. *Fundamenta Informaticae*, Vol. 51(3), p.271-295.

Esteva, M. 2003. *Electronic Institutions: from specification to development*. Ph.D. Thesis, Technical University of Catalonia.

Ferber, J. 1995. *Les Systèmes multi-agents : Vers une intelligence collective*. InterEditions, Paris.

Ferber, J. and Gutknecht, O. 1998. A Meta-Model for the Analysis and Design of Organizations in Multi-Agent Systems. In *Proceedings of the 3rd International Conference on Multi-Agent Systems*, IEEE Press Paris.

FIPA 2005. Agent Communication Standards. http://www.fipa.org/repository/aclspecs.php3.

Gâteau, B.; Khadraoui, D.; Boissier, O. and Dubois, E. 2004. Multi-agent organizational model for e-contracting. In *Proceedings of the 3rd Autonomous Agents and Multiagent Systems Conference – AAMAS'04*, ACM Press, p.1452-1453.

Georgeff, M. and Lansky, A. 1987. Reactive Reasoning and Planning. In Proceedings of the Conference of the American Association of Artificial Intelligence, Seattle, WA, p.677–682.

Glass, A. and Grosz, B. 2000. Socially Conscious Decision-Making. In Proceedings of the Agents2000 Conference, Barcelona, Spain, p. 217 - 224.

Hannoun, M.; Sichman, J.; Boissier, O. and Sayettat, C. 1998. Dependence Relations between Roles in a Multi-Agent System. In Proceedings of the Workshop on Multi-agent systems and Agent-Based Simulation – MABS'98, Paris, France.

Hexmoor, H. 2000. Case Studies of Autonomy. In *Proceedings of FLAIRS 2000*, AAAI Press, p.246-249.

Hexmoor, H.; Castelfranchi, C. and Falcone, R. 2003. *Agent Autonomy*, In *Multi-Agent Systems, Artificial Societies, and Simulated Organizations*, Vol 7, Kluwer Publishing.

Hübner, J. 2003. *Un Modèle de Reorganisation dans les Systémes Multi-Agents*. PhD thesis, Universidade de São Paulo, Escola Politécnica.

d'Inverno, M. and Luck, M. 2001. *Understanding Agent Systems*, Springer-Verlag.

Jennings, N. 1993. Commitments and conventions: the foundation of coordination in multi-agent systems. *Knowledge Engineering Review*, Vol 2(3), p.223-250.

Jennings, N.; Norman, T. and Faratin, P. 1998. ADEPT: An Agent-based Approach to Business Process Management. *ACM SIGMOD Record*, Vol. 27 (4), p.32-39.

Jones, A. and Sergot, M. 1996. A formal characterisation of institutionalised power. *Journal of the IGPL*, Vol 4(3), p.429-445.

Kollingbaum, M. 2005. *Norm-governed Practical Reasoning Agents*. PhD thesis, University of Aberdeen, Scotland.

Langer, A. 2002. *Applied E-Commerce: Analysis and Engineering for E-Commerce Systems*. John Wiley Inc..

Lesser, V. and Gasser, L. 1995. *ICMAS Proceedings of the First International Conference on Multiagent Systems*.

Lesser, V., *et al.* 2004. Evolution of the GPGP/TAEMS Domain-Independent Coordination Framework. *International Journal of Autonomous Agents and Multi-Agent Systems*, Vol 9(1), p.87-143.

López y López, F. 2003. *Social Powers and Norms: Impact on Agent Behaviour*. PhD thesis, University of Southampton, UK.

Lorini, E. *et al.* 2007. Grounding Power on Actions and Mental Attitudes. In *Proceedings of FAMAS'07.*

Luck, M. and d'Inverno, M. 2001. Autonomy: A Nice Idea in Theory. In *Castelfranchi, C. and Lesperance, V. (eds.) : Intelligent Agents VII: Proceedings of the7th International Workshop on Agent Theories, Architectures and Languages – ATAL'01*, LNAI 1986, Springer-Verlag.

Matson, E. and deLoach, S. 2005. Formal Transition in Agent Organizations. In *Proceedings of the International Conference on Knowledge Intensive Multiagent Systems*, Waltham, USA, IEEE Press.

Maudet, N. and Chaib-draa, B. 2002. Commitment-based and dialog-game based protocols - new trends in agent communication languages. *Knowledge Engineering Review*, Vol. 17(2), p.157-179.

Morge, M. 2005. *Système dialectique multi-agents pour l'aide à la concertation*. Ph.D. thesis, ENS Mines, Saint Etienne, France.

Meyer, J.-J.Ch. and Wieringa, R.J. 1991. *Deontic Logic in Computer Science:Normative Systems Specification*. John Wiley and sons.

Muller, J.P. 1996. *The Design of Intelligent Agents: A Layered Approach*. Springer-Verlag.

Muller, G. 2006. *Utilisation de normes et de réputations pour détecter et sanctionner les contradictions*. Ph.D. thesis, ENS Mines, Saint Etienne, France.

Munroe, S. and Luck, M. 2005. Motivation-Based Selection of Negotiation Opponents. In *Gleizes, M.-P.* et al. *(eds.) : Engineering Societies in the Agents World V*. LNCS 3451, Springer-Verlag Berlin, p.119-138.

Nikles, M.; Rovatsos, M. and Weiss, G. (eds.): *Agents and Computational Autonomy: Potentials, Risks and Solutions*. LNAI 2969, Springer-Verlag Berlin.

Omicini, A. 2000. Hybrid Coordination Models for Handling Information Exchange among Internet Agents. In Proceedings of the 7th AIIA Convention, Workshop Agenti intelligenti e Internet: teorie, strumenti e applicazioni, Milano, Italy.

Omicini, A. and Zambonelli, F. 1998. The TuCSoN Coordination Model for Mobile Information Agents. In Proceedings of the First Workshop on Innovative Internet Information Systems, Pisa, Italy.

Pasquier, Ph.; Flores, R. and Chaib-draa, B. 2005. Modeling flexible social commitments and their enforcement. In *Gleizes, M-P.* et al. *(eds.): Engineering Societies in Agents World V*, LNCS 3451, Springer-Verlag Berlin, p.111-116.

Perkikns, Ch. 2000. *Ad-hoc Networking.* Addison-Wesley, Boston.

Ricci, A.; Viroli, M. and Omicini, A. 2004. Role-Based Access Control in MAS using Agent Coordination Contexts. In Proceedings of the AOTP'04 Workshop, San Jose, USA.

Sabater, J. 2003. *Trust and Reputation for agent societies*. PhD thesis, IIIA, Barcelona, Spain.

Schumacher, M. 1999. *Designing and Implementing Objective Coordination in Multi-Agent Systems*. PhD thesis, University of Fribourg, Fribourg.

Sichman, J.S. 1995. *Du raisonnement social chez les agents: une approche fondée sur la théorie de la dépendance*. PhD thesis, INPG, France.

Singh, M. 1991. Social and psychological commitments in multiagent systems. In AAAI Fall Symposium on Knowledge and Action at Social and Organizational Levels, p.104-106.

Singh, M. 1999. An Ontology for Commitments in Multiagent Systems: Towards a Unification of Normative Concepts. *AI and Law*, Vol. 7, p.97-113.

Stratulat, T. 2002. *Systèmes d'agents normatifs: concepts et outils logiques*. PhD thesis, University of Caen, France.

Sycara, K.; Dedrer, K. and Williamson, M.1997. Middle-agents for the Internet. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*. Morgan-Kauffman, p.578-583.

Rao, A.S. and Georgeff, M. 1991. Modeling Rational Agents within a BDI-Architecture. In *Allen, J.F. et al. (eds.): Proceedings of the International Conference on Knowledge Representation and Reasoning*, Morgan Kaufmann, Cambridge, MA, p.473–484.

Tambe, M.; Scerri, P. and Pynadath, D. 2002. Why the elf acted autonomously: Towards a theory of adjustable autonomy. In *Proceedings of the 1st Autonomous Agents and Multiagent Systems Conferences – AAMAS'02*, ACM Press.

van der Torre, L. 2003. Contextual Deontic Logic: Normative Agents, Violations and Independence. In *Annals of Mathematics and Artificial Intelligence*, Vol 37(1-2), p.33-63.

Tuomela, R. 1995. *The Importance of Us: A Phylosophical Study of Basic Social Norms*. Stanford University Press.

Valckenaers, P. 2004. *Challenges of Next Generation Manufacturing Systems*. SoftSpez Final Report 2004, p.23-28.

Vázquez-Salceda, J. 2004. *The Role of Norms and Electronic Institutions in Multi-Agent Systems*. Whitestein Series in Software Agent Technology, Birkhäuser Verlag AG.

van der Vecht, B. *et al.* 2007. A Dynamic Coordination Mechanism Using Adjustable Autonomy. In *Proceedings of the COIN'007 Workshop*, Durham, UK, p.169-180

Verhagen, H. 2000. *Norm Autonomous Agents*. PhD thesis.

Weiser, M. 1991. The Computer for the 21st Century. *Scientific American,* Vol 165(3), p.94-104.

Wooldridge, M. 1999. Intelligent Agents, In *Weiss, G. (ed.): Multi-agent systems: A modern approach to Distributed Artificial Intelligence*, MIT Press.

von Wright, G.H. 1981. On the logic of norms and actions. In *New Studies in Deontic Logic*, p.3-35.

Publications de l'auteur

O.Boissier, C.Carabelea, C.Castelfranchi, J.Sabater and L.Tummolini. 2005. The Dialectics between an Individual and His Role. In *Roles: An Interdisciplinary Perspective - AAAI Fall Symposium*.

C.Carabelea, O.Boissier and C.Castelfranchi. 2005. Using social power to enable agents to reason about being part of a group. In *M-P.Gleizes et al. (eds.): Engineering Societies in the Agents World V: 5th International Workshop, ESAW 2004, Toulouse, France, October 20-22, 2004. Revised Selected and Invited Papers*. LNCS 3451, Springer-Verlag Berlin, p.166-177.

C.Carabelea, O.Boissier. 2005. Coordinating agents in organizations using social commitments. Accepted for publication in *the ENTCS Journal - special issue on postproceedings of the 1st International Workshop on Coordination and Organisation*, Namur, Belgium.

C.Carabelea, M.Berger. 2005. Agent Negotiation in Ad-Hoc Networks. In *Proceedings of the Ambient Intelligence Workshop at AAMAS'05 Conference, Utrecht, The Netherlands*, p.5-16.

C.Carabelea, M.Berger. 2005 Negotiating Agents in Ad-Hoc Networks. Poster at the Mobiquitous 2005 Conference, Las Vegas.

C.Carabelea, M.Berger. 2005. Negociation entre agents au sein des reseaux ad-hoc. In *Proceedings of the 2emes Journees Francophones Mobilite et Ubiquite 2005*, Grenoble, France, p.145-151.

C.Carabelea, C.Castelfranchi, O.Boissier. 2004. Autonomie et pouvoir social. In *O.Boissier et al. (eds.): Actes des Journées Francophones des Systèmes Multi-Agents (JFSMA04)*, p.195-208.

C.Carabelea, O.Boissier and A.Florea. 2004. Autonomy in MAS: A Classification Attempt. In *M.Nikles et al. (eds.): Agents and Computational Autonomy: Potentials, Risks and Solutions*. LNAI 2969, Springer-Verlag Berlin, p.103-113.

C.Carabelea, O.Boissier and F.Ramparany. 2003. Benefits and Requirements of Using MAS on Smart Devices. In *H.Kosch et al. (eds.): Euro-Par 2003 Parallel Processing: 9th International Euro-Par Conference, Klagenfurt, Austria, August 2003, Proceedings*. LNCS 2790, Springer-Verlag Berlin, p.1091-1098.

C.Carabelea, O.Boissier and A.Florea. 2003. Deploying Multi-Agent Systems on Small Devices. In *Proc. of the 14th International Conference on Control Systems and Computer Science CSCS14*, Bucharest.

C.Carabelea and Ph.Beaune. 2003. Engineering a Protocol Server Using Strategy Agents. In *V.Marik et al. (eds.): Multi-Agent Systems and Applications III: 3rd International Central and Eastern European Conference on Multi-Agent Systems*, *CEEMAS 2003, Prague, Czech Republic, June 2003, Proceedings*. LNAI 2691, Springer, p.413-422.

C. Carabelea and O.Boissier. 2003. Multi-agent Platforms for Smart Devices: Dream or Reality?. In *Proc. of the Smart Objects Conference*, Grenoble, p.126-129.

C.Carabelea, O.Boissier and A.Florea. 2003. Autonomie dans les systèmes multi-agents: essai de classification. In *J-P.Briot and K.Ghadira (eds.): Deploiement des SMA – vers un passage a l'echelle, JFSMA 2003 Proceedings*, Hammamet, Tunisia, p.191-204.

A. Florea, E. Kalisz, and C. Carabelea, 2002. Genetic Multi-agent Planning of Self-interested Agents. In *Proc. of the Genetic and Evolutionary Computation Conference GECCO 2002*, New York.

A. Florea, C. Carabelea. 2002. Genetic Models for the Rational Exploitation of Resources. In *Proc. of 7th Online World Conf. on Soft Computing in Industrial Applications WSC7*.

C.Carabelea, 2001. Adaptive Agents in Argumentation-Based Negotiation. In *V. Marik et al. (eds.): Multi-Agent Systems and Application II, Selected Revised Papers: 9th ECCAI-ACAI/EASSS 2001, AEMAS 2001, HoloMAS 2001*. LNAI 2322, Springer-Verlag Berlin, p.176-183.

A.Florea, E.Kalisz, C.Carabelea: Genetic Prediction of a Multi-agent Environment Evolution. In *Proc. of the 4th International Conference on Computing Anticipatory Systems CASYS 2000*, Liege, Belgium. This paper has received Best Paper Award with Belgium Crystal at Symposium I: Fuzzy Systems, Neural Nets, Genetic Algorithms and Anticipation.

Annexe

Cette annexe contient des exemples de la réalisation dans Prolog du modèle de pouvoir social et autonomie et son utilisation dans le scénario ADOMO :

```
%%%%%% WORLD DESCRIPTION - ACTIONS, GOALS, PLANS and associated predicates  %%%%%%%

action(neg1).
action(neg2).
action(exec_job).
action(pay).

plan(plg, gain_money, [negotiate], [exec_job]).
plan(plneg, negotiate, [], [neg1,neg2]).

conflict(plneg, neg1, neg2).
enables(plg, negotiate, exec_job).

goal(X):-plan(_,X,_,_).

%%
%% verifies if an element belongs to a plan
%%
belongs(El, Plan):-plan(Plan, _, L, _), member(El, L).
belongs(El, Plan):-plan(Plan, _, _, L), member(El, L).

%%
%% get a list of all elements of a plan and a list of all elements not in a plan,
but which enable elements from it
%%
getPlanElements(Plan, L):- findall(X, belongs(X, Plan), L).
getAllEnables(Plan, L):- findall(X, (enables(Plan, X, _), not(belongs(X,Plan))),
L).

%%
%% verifies if all the elements of a list satisfy a predicate with a given
parameter
%%
verifyList(_, [], _).
verifyList(Pred, [X|R], Param):-call(Pred,Param,X), verifyList(Pred,R,Param).

%%
%% verifies if two elements of a plan are in conflict with each other in the plan
%%
conflictElem(X, Y, Plan, L):-member(X,L), member(Y,L), X\=Y, conflict(Plan, X, Y).
conflictElem(X, Y, Plan):-getPlanElements(Plan, L), conflictElem(X, Y, Plan, L).



%%%%%% THE ORGANIZATION %%%%%%%

%knows_how(_,_).
knows_how(a1,neg1).
knows_how(a1,neg2).
knows_how(a2,neg2).

knows_how(a3,exec_job).

permission(a1,neg1).
permission(a2,neg2).
permission(a3,exec_job).

plays(ea,rExec).
plays(ba,rBroker).
plays(ca1,rContr).
```

```
plays(ca2,rContr).
plays(ma,rMarket).

authority(rExec,rBroker,negotiate).

obligation(rContr,pay,[Y,Id,Val]):-
plays(X,rContr),contract(_,X,Id,Val,active),plays(Y,rMarket).
obligation(rMarket,pay,[Y,Id,Val]):-contract(Y,_,Id,Val,fulfilled), plays(Y,rExec).
obligation(rExec,fulfill,[Id,Val]):-plays(X,rExec),contract(X,_,Id,Val,paid).
obligation(rContr,negotiate,[_,rBroker]).

permission(_,pay,_).
permission(rBroker,neg1,rContr).
permission(rContr,neg2,rBroker).
permission(rExec,exec_job,_).



%%%%%%% SOCIAL COMMITMENTS %%%%%%


%%% operations

isSComm(Id):-sc(Id,_,_,_,_,Cond), number(Cond).
isSPol(Id):-sc(Id,_,_,_,_,Cond), atom(Cond).
isRComm(Id):-sc(Id,_,_,_,_,true).

status(Id,S):-sc(Id,_,_,_,S,_).

changeStatus(Id,Stat):-sc(Id,X,Y,Obj,S,C),
retract(sc(Id,X,Y,Obj,S,C)),assertz(sc(Id,X,Y,Obj,Stat,C)).

fulfillAll([Ag,Action,T]):-sc(Id,Ag,_,Action,act,Cond), number(Cond), T=<Cond,
changeStatus(Id,ful), fail.
fulfillAll(_).

violateAll([T]):-write("v"),sc(Id,_,_,_,act,Cond), number(Cond), T > Cond,
changeStatus(Id,vio), fail.
violateAll([REQ,X,Y,Act,CR]):-plays(X,R),
sc(Id,Y,_,[[REQ,R,Act],[CR]],act,Cond),atom(Cond), changeStatus(Id,vio), fail.
violateAll(_).

plays(Ag,R):-sc(_,Ag,_,[R],act,true),!.

valid(Id,T):-isSComm(Id), sc(Id,_,_,_,act,Cond), T=<Cond.
valid(Id,_):-isSPol(Id), sc(Id,X,_,_,act,Cond), plays(X,Cond).
valid(Id,_):-isRComm(Id).



%%%%%% POWERS AND DEPENDENCIES %%%%%%%

%%%%%% INDIVIDUAL POWERS  %%%%%%%

can(Ag, A):-action(A), knows_how(Ag, A).

can(Ag, Goal):- goal(Goal), plan(Pl, Goal, _, _), can(Ag, Pl).

can(Ag, Plan):- getPlanElements(Plan, L), L\=[], verifyList(can, L, Ag),
not(conflictElem(_,_,Plan,L)),
  getAllEnables(Plan, L2), verifyList(can, L2, Ag).


may(Ag, A):-action(A), permission(Ag, A).

may(Ag, Goal):- goal(Goal), plan(Pl, Goal, _, _), may(Ag, Pl).
```

```
may(Ag, Plan):- getPlanElements(Plan, L), L\=[], verifyList(may, L, Ag),
getAllEnables(Plan, L2), verifyList(may, L2, Ag).


%%%%%% BASIC DEPENDENCIES  %%%%%%%

deo_depends(A1, A2, X, Plan, Goal):-plan(Plan, Goal, LG,LA),
(member(X,LG);member(X,LA)), not(may(A1,X)), may(A2,X),A1\=A2.
deo_depends(A1, A2, X, Plan, Goal):- plan(Plan, Goal, _, _), enables(Plan, X, _),
not(belongs(X,Plan)),
  not(may(A1,X)), may(A2,X),A1\=A2.


%%%%%% AGENT REASONING %%%%%%%

fulfillGoal(Ag,Goal):-retractall(plan(Ag,[Goal|_])).
payContract(X,Y,Id,Val):-retract(contract(X,Y,Id,Val,active)),
asserta(contract(X,Y,Id,Val,paid)).
fulfillContract(X,Y,Id,Val):-retract(contract(X,Y,Id,Val,paid)),
asserta(contract(X,Y,Id,Val,fulfilled)).

reason(Ag):- fulfillObligations(Ag), adopt_tasks(Ag), !, (pursue_plan(Ag), !;
generate_plan(Ag), pursue_plan(Ag)).

fulfillObligations(Ag):- findall(T, doObligation(Ag,T), _).

doObligation(Ag,Task):-plays(Ag,Role), obligation(Role, Task, Param),
obligationToGoal(Ag,Role,Task,Param).

obligationToGoal(Ag,Role, pay,[Y,Id,Val]):-contract(X,Ag,Id,Val,active),
permission(Role,pay,_),
  payContract(X,Ag,Id,Val), assertz(action(Ag,[pay,Y,Val])).
obligationToGoal(Ag,Role,pay,[Y,Id,Val]):-
contract(Y,_,Id,Val,fulfilled),permission(Role,pay,_),
  retract(contract(Y,_,Id,Val,fulfilled)), assertz(action(Ag,[pay,Y,Val])).
obligationToGoal(Ag,_,Task,_):-plan(Ag,[Task|_]).
obligationToGoal(Ag,_,Task,Param):-not(simpleTask(Task)), not(plan(Ag,[Task|_])),
asserta(plan(Ag,[Task|Param])).

adopt_tasks(Ag):- findall(T,adopt_deleg(Ag,T),_).

adopt_deleg(X,T):-delegate(Y,X,T,P),L=[T,Y,P], retract(delegate(Y,X,T,P)),
decideAdoption(X,L),
  (task(T,_),assertz(plan(X,L));simpleTask(T),assertz(action(X,L))).

decideAdoption(Ag, [Task, Whom, _]):- adoptBecauseInterested(Ag,Task);
adoptBecauseAuthority(Ag, Whom, Task).

adoptBecauseAuthority(X, Y, Task):- plays(X,R1), plays(Y,R2),
authority(R2,R1,Task).

adoptBecauseInterested(X, Task):- task(Task,_), plan(X, [P|_]), subgoal(Task, P).
adoptBecauseInterested(X, Task):- task(Task,_), plan(X,[Task|_]).
adoptBecauseInterested(X, Task):- simpleTask(Task), plan(X, [P|_]), task(P,L),
member(Task,L).

pursue_plan(Ag):- plan(Ag,[Goal|Params]), execute(Ag,Goal,Params).

generate_plan(X):- (goal(X,Goal); desire(X,D,STR),STR>0.0,satisfy(Goal,D)),
  findall(Sub, (subgoal(Sub,Goal), assertz(plan(X,[Sub])) ), _).


%%%% TASK EXECUTION %%%%

execute(X,Task,Params):-simpleTask(Task), fulfillGoal(X,Task),
asserta(action(X,[Task|Params])), propagateGoal(X,Task).
propagateGoal(X,Task):-plan(X,[P|_]), task(P,L), member(Task,L), fulfillGoal(X,P).
```

```
execute(X,wait,_):- contract(X,_,_,_,paid), !, fulfillGoal(X,wait), pursue_plan(X).
execute(_,wait,_).

execute(X,fulfill,_):- contract(X,Y,Id,Val,paid), task(fulfill, [Act]),
plays(X,Role), permission(Role,Act,_),
  fulfillContract(X,Y,Id,Val), fulfillGoal(X,fulfill),
assertz(action(X,[Act,Id,Val])).

execute(X,negotiate,Params):-task(negotiate,[T1,T2]), plays(X,R),
execute(negotiate,Params,T1,T2,X,R).

execute(negotiate,[_,Z],T1,T2,_,R):-not(permission(R,T1,_)), permission(R,T2,Z).
execute(negotiate,_,T1,T2,X,R):-not(permission(R,T1,_)), not(permission(R,T2,_)),
permission(R2,T1,R3),
  fulfillGoal(X,negotiate), asserta(plan(X,[wait])),
assertz(action(X,[delegate,R2,negotiate,R3])).
execute(negotiate,[Y,Z],T1,T2,X,R):-permission(R,T1,Z), not(permission(R,T2,_)),
permission(Z,T2,R),
  fulfillGoal(X,negotiate), assertz(action(X,[delegate,Z,T2,Y])),
assertz(plan(X,[T1,Z,Y])).
```