Contributions to an Anthropological Approach to the Cultural Adaptation of Migrant Agents

Rafael Heitor Bordini

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London.

Department of Computer Science
University College London

1999

Abstract

This thesis proposes the use of Cultural Anthropology as a source of inspiration for solutions to the problem of adaptation of autonomous, intelligent, computational agents that migrate to societies of agents with distinctive features from the ones of the society where those agents were originally conceived. This has implications for interoperation of disparate Multi-Agent Systems. In particular, the cognitive approach to anthropology is argued to be a suitable theoretical foundation for this topic. Fieldwork practice in social anthropology is also indicated as an useful source of ideas. A pragmatic theory of intensionality is incorporated in this anthropological approach, resulting in a mechanism that allows agents to ascribe intensional ontologies of terms to societies that use unfamiliar means of communication; also, taxonomical relations among the terms in such ontologies can be retrieved, by means of a process inspired by the counterpart activity of ethnographers. This is presented using the Z notation for formal specification of systems, and illustrated on a set of terms from the game of cricket. Subsequently, a simulation of a game of cricket is described where one of the players is unfamiliar with the game, and therefore needs to learn the game by observing the other players. A reasonable behaviour for such a player is obtained, and the simulation offers grounds for further anthropologically-based studies. Further, a study of theories of moral sentiments is presented, and the Iterated Prisoner's Dilemma is used in simulations based on those ideas. The results of the simulations show clearly the positive impact, on groups of agents, of altruistic behaviour; this can only be coherently obtained in autonomous agents by modelling emotions, which are relevant for this project as anthropologists recognise them as an essential cross-cultural link. Finally, the consequences of this project to conceptions of Distributed Artificial Intelligence are discussed.

© Édition de la Reunion des Musées Nationaux, see (Mathieu and Lacambe 1997). Reproduced by kind permission of the *Agence Photographique RMN*.

The Migrant Angel
by *Gustave Moreau*
Water-colour, 30 x 23 cm.
(cat. 441) Musée Gustave Moreau.

Musée Gustave Moreau
14 rue de la Rochefoucauld,
75009, Paris.

Dedication

I dedicate this thesis to my grandmother, Adelina Mazzini (in memoriam)*, who assisted in my upbringing (the fundamental* cultural adaptation*); and to my great-aunt Iris (Lina) Stendardi Mazzini* (in memoriam), *a brave and* truly altruistic migrant agent *as well as a delightful soul, for her always uplifting smile and the sweetest memories she left.*

This thesis was typeset in Monotype Baskerville 10.95 on 20.4 using LaTeX.

Acknowledgements

If only there was an adequate way of acknowledging Prof. John A. Campbell, my supervisor. I cannot think of a way of expressing all that he has done for this thesis to be in its present state; all the amazingly careful corrections to the texts, the comments, suggestions, discussions, etc. And even if there was such a way, how could one possibly acknowledge with words all the lessons about academia and life in general, as well as scientific matters (in particular, but not at all only, on DAI), that only a supervisor of the stature of Prof. John Campbell can provide. And why not thank him for "simple" things like being always kind, understanding and encouraging. Eliseo Reategui in his thesis already mentioned the delightful stories we have the privilege of hearing during the meetings, in what is a brilliant didactic style; he also referred to Prof. Campbell as "a remarkable man." I cannot find better words.

Many researchers have helped with this work, and I acknowledge their help here as well as I express my gratefulness for their inspiring collaboration. I am indebted to Renata Vieira and Ana L. C. Bazzan for their support on two of the areas we have drawn from: formal pragmatics (Chapter 5) and game theory (Chapter 7), respectively. I also acknowledge: Dr. Rosa M. Viccari, Dr. Antônio Carlos da Rocha Costa, my previous supervisors (also Renata Vieira and Jomi F. Hübner, main collaborators at the time); Prof. Yves Demazeu for the opportunity to present part of my work for his group and Dr. Jaime S. Sichman for some hints given at the time; Dr. Philip C. Burnham, from the Department of Anthropology at UCL, who gave some advice on sources in anthropology (as did Prof. Les Gasser, mentioned next); to Prof. Philip Cohen and Prof. Les Gasser, for the encouragement in my poster presentation at ICMAS'95; to Mark d'Inverno and Michael Luck for repeated support in my using of their Z-based framework; and finally to Ricardo C. Krug, Gerson G.H. Cavalheiro, Luciana P. Nedel, Soraia R. Musse, Giannis Koufakis and Jon Davidson for various help with the implementation of the cricket simulator. Thanks also to the some anonymous referees to our papers for many interesting comments. I also thank my brother Ricardo Mazzini Bordini for coining the term *artificial culture* and for suggesting the *Moiras* for association with the parts of my thesis. And Evangelos Grammenidis, a research student in the Classics Department at UCL (and thanks to Giannis Koufakis for the contact with him) for the meticulous suggestions and corrections to the form of the Greek citation. To the *Agence Photographique RMN* for the kind permission (and providing the computer file) for the illustration on Page 3.

I am very grateful to CAPES, a Brazilian government agency for postgraduate studies, for the grant for my PhD. Thanks a lot to my brother Rubens Salvador Bordini, who has been my proxy in Brazil, for all the troubles he had to deal with. Further, I acknowledge some colleagues who helped with daily life in this department: Eliseo Reategui (the "compatriot agent"); Paul Pal and Arif Iqbal (the "sympathetic agents"); Alberto Souza (to whom I functioned as compatriot agent), José Luis C.M. Borges, Antonio Liotta, Jorge Ortega-Arjona and Tom Quick, recent office-mates; and my office-mates for many years: Alfarez Abdul-Rahman, Giannis Koufakis and Jon Davidson. It is not possible to name them all, but thanks too all office-mates in the (many) rooms in which I have lived during these four years at UCL-CS: 208, 209, 227, 401; some other research students at UCL-CS have been very nice too, and again I mention only as representative Hugh Mallinson, a charming English gentleman who has recently done a great deal to socialise the department. Thanks to my second supervisor, Prof. Philip Treleaven. Thanks to Russel Winder, and later Mel Slater, who were the Research Student Tutors for the department; thanks also to Susan Pike who helped a lot since she started as PhD Student Administrator. To Russel Winder also for all I learnt about LaTeX with him (and Julia Schnabel, among the few in the department to be as fond of LaTeX as I was). Thanks to some administrative and (the very competent and sympathetic!) systems-group staff (as I cannot mention them all, I take Marion Sutcliffe and Barry Stein, respectively, as representatives); thanks are due even to the nice security guard who I met so many times when working late.

Finally, the most important type of acknowledgement is actually that of a personal nature. No one accomplishes a PhD programme (particularly abroad) without *a lot* of emotional support. Therefore, to acknowledge all of my family and friends is fundamental. However, I adore my parents (Idahyr Salvador Bordini and Maria Celestina Mazzini Bordini), brothers and sisters (Ricardo, Lizete, Roberto, Renato, Rubens, Liliane), sisters-in-law (Andréa, Sandra, Mai and Cristina), and my niece (Thays, Roberto and Sandra's daughter) so much that again I am left with no words. But I must at least acknowledge that, besides all emotional support, my parents also partly supported (particularly towards the end of the PhD) the heavy expenses of my living in London (and also some luxuries like the MT Baskerville fonts for typesetting this thesis). Also, I have some very good friends, and friends for me are (I dare say) even more important than they really are to everybody. They are: Arthur G.S. Bloise, Nara M. Bigolin, Denise B. da Silva, Renata Vieira, Alvaro F. Moreira, Debora A. Santos, Eliseo B. Reategui, Luciana P. Nedel, Otilia Werner-Kytölä, Ricardo C. Krug and Gerson G.H. Cavalheiro. Thanks to all of them and all others that I have failed to mention here.

Contents

I The Past (*When it Was Found that Autonomous Agents Were Fated to Migrate*) 14

Appendices

List of Figures

List of Tables

List of Algorithms

Preface

One thing that needs clarification from the very beginning is the apparently strange use of the word "anthropological" in the title of a computer science thesis, simply for the fact that the Greek word *anthrōpos* means *man*; while of course we are not dealing here with humans but with computer systems. However, cultural anthropology, in so far as it concerns the study of social institutions and relations, does have a say in computational problems in the area of Distributed Artificial Intelligence (DAI). To show that that is so is the main purpose of this thesis. If the reader, at the end, has a feeling that interoperation of disparate societies of agents is possible (i.e., it is possible to interoperate without standardisation), and that anthropology is relevant to open research issues in DAI and therefore that one should do justice to it and recognise its place alongside sociology, psychology, cognitive science, economics, and others as foundational disciplines in DAI, this thesis will have made its point.

The thesis is organised in three parts, reflecting whether the chapters therein are of a preliminary nature, constitute the basis of our anthropological approach to interoperation of Multi-Agent Systems (MAS), or point to future directions in MAS in the light of our approach. The thesis presents a collection of slightly independent works (the "contributions"), each directed towards a particular facet of the overall problem of cultural adaptation of migrant agents and exploring a particular point provided by such anthropological conception.

An unusually large number of abbreviations will be used in the text. However, we hope to have alleviated any difficulty in the reading of the thesis resulting from this by including all the abbreviations in the Index, and some of them are included in the Glossary as well (see pages 236 and 235, respectively).

The work reported in this thesis is original except where noted in the text by appropriate citations, and has not been presented for any other degree.

Part I

The Past

(*When it Was Found that Autonomous Agents Were Fated to Migrate*)

« ἄλλας δὲ καθημένας πέριξ δι’ ἴσου τρεῖς, ἐν θρόνῳ ἑκάστην, θυγατέρας τῆς Ἀνάγκης, Μοίρας, λευχειμονούσας, στέμματα ἐπὶ τῶν κεφαλῶν ἐχούσας, Λάχεσίν τε καὶ Κλωθὼ καὶ Ἄτροπον, ὑμνεῖν πρὸς τὴν τῶν Σειρήνων ἁρμονίαν, Λάχεσιν μὲν τὰ γεγονότα, ... »
Πλάτωνος, Πολιτεὶα

“*And there were other three who sat round about at equal intervals, each one on her throne, the Fates, daughters of Necessity, clad in white vestments with filleted heads, Lachesis, and Clotho, and Atropos, who sang in unison with the music of the Sirens, Lachesis singing the things that were, ...* ”
Plato, The Republic [617B–C]

(Shorey 1935)

Chapter 1

Introduction

1.1 Definition of the Problem and Basic Concepts: Migration of Agents and Their Process of Adaptation

We deal in this thesis with the problem of *migration of agents* among different (mutually unfamiliar) societies, in the context of *Multi-Agent Systems* (MAS). The basic idea is that certain individual agents should be able to interact in societies of agents which were designed using paradigms or theories of agents different from their own, or which had different histories of autonomous evolution, therefore allowing different systems to profit from occasional interactions. This problem was initially defined in (da Rocha Costa, Hübner, and Bordini 1994; Bordini 1994; Bordini *et al.* 1995). Our project considers basically the agent adaptation that takes place in such migrations. In particular, this thesis aims at presenting an "Anthropological Approach to the Cultural Adaptation of Migrant Agents;" that is, we present the idea of using cultural anthropology (also called social anthropology) as a source of insights for developing solutions to the problem of enabling agents to interact in societies of agents with whose "cultures" they are not familiar.

To make the notion of MAS clearer, at this early stage (more on this is provided in Chapter 2), we include below some definitions given by Gasser (1987). Research on Artificial Intelligence (AI) traditionally involves "a single intelligent agent reasoning and acting under a single thread of control or focus of attention. ... *Distributed Artificial Intelligence* (DAI) is the subfield of AI concerned with concurrency in AI computations, from the standpoint of multiple problem-solvers, multiple control loci, and multiple simultaneous problems." He then classifies DAI work in terms of three arenas: *Parallel AI* "is concerned with developing computer architectures, languages and algorithms for concurrent AI;" *Distributed Problem-Solving* "is the study of techniques for dividing the work of solving a particular problem among several modules that cooperate by interacting and sharing knowledge about the problem and the developing solution;" and

Multi-Agent Systems (MAS) "is concerned with coordinated intelligent behavior among a collection of (possibly pre-existing) autonomous intelligent 'agents:' how they can coordinate their knowledge, goals, skills, and plans jointly to take action or solve (possibly multiple, independent) problems." Agents in MAS also interact and share knowledge about the problems they are solving, but unlike agents in distributed problem solving they "must also reason about the *processes of coordination among the agents*." Gasser also mentions that the classification of a system within each of these areas can be established by considering "how much of the reasoning about problem solving and coordination is done by system developers, and how much is done by the system itself; how adaptive can the system itself be to changes in the problem or context." The systems deserving classification as MAS are the most adaptive and independent ones.

Before defining better the introductory aspects of agent migration and adaptation, we first need to introduce some of the terminology we shall use throughout the thesis. A *Migrant Agent* (MA) is an agent that is capable of leaving and entering societies "freely." If it is leaving society **S1** and entering society **S2**, then, from the MA's point of view, **S1** is the *society of origin* and **S2** is the *Target Society* (TS). Considering **S1**, the MA is an *emigrant agent* and considering **S2** it is an *immigrant agent*. We also refer, at times, to the terms *local* and *foreign*. An immigrant agent can also be called a *Foreign Agent* (FA), and the *local agents* are the ones originally present in a target society; that is, those with which an immigrant agent is not familiar.

As the idea of several levels of organisations or groups in MAS (Gasser 1991) or the intrinsic *recursion principle* of MAS (Demazeau 1995) is not particularly new, it is important to clarify what we mean by different societies. Considering contrasting designers' preferences, it is quite likely that systems designed by different people may come to have different features in communication language and social norms that govern the action of the agents in a society. In (da Rocha Costa, Hübner, and Bordini 1994) it was already remarked that distinct open societies do not necessarily agree on language and protocols for communication and knowledge interchange. Not to mention that DAI has a whole range of different approaches. Therefore, apart from this (cognitive) type of agents, societies may be formed of agents that behave in a completely different way (we make this more clear in Section 4.2, where we redefine the problem of migration in anthropological terms).

Another point, apart from the differences in the initial design, is that we think that *autonomous intelligent agents* should be capable of "hanging loose" from any standard or initial features that they receive when they are created (if they are *intelligent* and *autonomous*, there is no reason why they should not). This means that a group of agents may come to promote changes in their language, or their pre-defined ways of coordination (e.g. social norms (Conte, Castelfranchi, and Dignum 1999)), just as cultural changes occur with evolution in human societies; the problem of agents changing their communication protocols, for example (and see Chapter 3), was already foreseen in (Populaire *et al.* 1993). Further, it is plausible that each society of agents may have its own way of perceiving and organising material phenomena in its own environment, deliberating over its own social structure, etc. (see the discussion on *artificial cultures*, Section 4.2).

Therefore, when we mention interaction among different societies, we are speaking of a different level of contrast from that of societies of culturally identical groups (or organisations) of agents (or of sub-agents composing agents) discussed in the literature that refers to adaptation of agents. In brief, we are dealing with *interoperation* (Genesereth and Ketchpel 1994; Wiederhold 1994) of completely distinct MAS. One consequence of our view of allowing differences in societies rather than standardising them (which is the usual approach to interoperability) is that it is desirable to develop agents that can produce anthropologically-based formal descriptions of the cultures present in arbitrary MAS, in order to help migrating agents in their processes of cultural adaptation. Such formal descriptions must include various aspects of a society. Both issues mentioned above are discussed further in Section 4.2.

As stated in (Bordini 1994), these significantly different societies are quite likely to be in distinct geographical locations; they are independent both functionally and physically. Thus local communication among them is not possible and the actual environment that agents share is also different. Accordingly, agents cannot "be" (i.e., interact) in both societies at the same time. However, some liaison does exist between these sites so that an agent can move from one to the other (the ubiquitous, world-wide network infrastructure makes this trivial). That is the main reason why we refer to *migration of agents*: the agent leaves a group of partners and joins another (but it is no longer practicable to interact as before with the previous one).

There are two rationales supporting the idea of an agent migrating from one society to another, which are also very easy to understand by reference to the human analogue. As mentioned in (da Rocha Costa, Hübner, and Bordini 1994), an agent immigrates according two rationales: (i) the agents in the target society need the immigrant agent's specific abilities in order to accomplish their current social goals, and could, by whatever means, persuade it to join them; or (ii) the immigrant agent needs the help of the local agents in the target society for accomplishing its individual goals, given that this is not possible in the society of origin (e.g., due to the lack of abilities of its current partners), and the immigrant agent believes it can manage to elicit help from the local agents.

Therefore, considering all the observations above, we conclude that if a world-wide variety of culturally distinct MAS is to be regarded, which, as we have mentioned, means a concern for interoperability, it is not realistic to expect the existence of foolproof standard means that would allow an agent to interact in every different society, as assumed in the aforementioned standardising approaches, e.g., the Agent Communication Language (ACL) produced by researchers in the ARPA Knowledge Sharing Effort (KSE) (Genesereth and Ketchpel 1994). Furthermore, we think it is interesting from the point of view of social simulation (Castelfranchi 1990) and agent cognition in general (Conte and Castelfranchi 1995) to have different societies using different languages and protocols, social norms, coordination mechanisms and all other aspects of agency. Accordingly, it is necessary that agents be able to adapt themselves to different "cultural" contexts, as observed in human societies. We argue that the means to do so is by having agents developing "anthropological" descriptions that can be used by migrant agents in their process of adaptation, as mentioned above. This leads to the idea of anthropology as a further discipline to be added among the existing subjects that contribute to the interdisciplinary field of Distributed Artificial Intelligence (DAI) or MAS; this suggestion will be made more specific throughout the thesis. Interestingly, anthropology is one of the disciplines where social simulation based on MAS techniques is used (see, e.g., (Doran *et al.* 1994)); we believe that, as for other disciplines that are relevant for the foundations of DAI (e.g., sociology), anthropology too can have a reciprocal relation with DAI, in the sense that DAI can help solving problems in anthropology—just as anthropology, in its turn, can help with DAI problems (related to interoperation of culturally distinct MAS) that have been neglected so far.

1.2 Overview of the Approach

The task we have set up for ourselves in this thesis—i.e., to show the relationship between cultural anthropology and DAI, as we mentioned in the Preface—is rather unusual for a computer science thesis. Therefore, the form in which we shall express our views on the matter, or indeed the method used to build up the argumentation of our thesis is, as a consequence, also a little unusual. We deemed that, e.g., building a system based on anthropological ideas and showing an improved performance in comparison to a 'non-anthropological' (i.e., traditional) approach, which is the usual method in computer science, would not suffice for such a general point as we aim to make here, besides being an impractical task in the present state of the art in MAS.

We have therefore opted for a multidisciplinary thesis in which the very method of demonstration is "mixed," in the sense that it is influenced by, or is borrowed from, the several sciences involved. The reader will not fail to notice that part of the thesis is merely "philosophical"; that is, some material from the humanistic sciences is discussed at an abstract level and compared to the field of computer science we are concerned with, namely DAI (this style in used mainly in Chapter 4)[1]. Also, we use a formal approach to develop some ideas inspired by anthropology to solve a particular problem in adaptation to communication features (Chapter 5). The formalisation is presented in Z, which in turn is based on first-order logic and set theory (as usual in theoretical AI). Further, we use an engineering (of software) approach and work in an experimental style (that is, using computer simulation and analysis of generated data)—these approaches apply to Chapters 6 and 7 respectively.

All these methods give particular contributions in grounding the discussion we present towards the end of the thesis, where we reconsider the many points involved in the project. This variety of contributions to our approach is an important asset for our enterprise, even if working on all of them simultaneously has meant, at times, leaving several unresolved issues in each of the contributions.

1.3 Structure of the Thesis

This thesis is structured in three parts. We have figuratively associated them with Past, Present and Future. The "past" (Part I) deals with the preliminaries, background (with a brief historical perspective on the area) and a review of our previous work on adaptation at the level of language and communication, which is included here as it helps to ground some of the argumentation we set forth throughout the thesis (particularly towards the end). The "present" (Part II) deals with contributions we make towards an anthropological approach to MAS (i.e., it concerns our present claim about the relevance of anthropology in a DAI context). Last, the "future" (Part III) deals with our views on the issues that an anthropological approach raises for the future of MAS (including some initial experimental work that helps to support some of those views), as well as discussion, future work, and concluding remarks.

In Part I, after this introductory chapter, we provide, in Chapter 2, a very brief review of the history of major work on DAI from our particular perspective, brief descriptions of others that are instrumental to our own work, and a picture of the several types of research on DAI. The last chapter of this part, Chapter 3, presents briefly our previous work on adaptation of agents to target societies in the level of language and communication (regarding, in particular, communication protocols).

We redefine the problem we are dealing with (namely, migration of agents) in anthropolog-

---

[1] The purpose here is to comment on the method used in each of the contributions that is part this thesis. A description of the structure of the thesis, with comments on each chapter, is given in the next section.

ical terms in Chapter 4, and the three chapters in Part II either contribute to or profit from the anthropological metaphor. This anthropological metaphor is established in Chapter 4, where most of the material taken directly from anthropology is presented. It includes the issues in anthropology that we found to be related, at least at an abstract level, to research problems in DAI. We present a particular contribution to the problem of adaptation to target societies in Chapter 5; it is concerned with ontological descriptions of terms used in agent communication languages, as well as with their taxonomical relations, and is inspired directly by anthropological methods. An exercise derived from the game of cricket and presented in Chapter 6 is a practical example for the anthropological metaphor, and is suitable for experimentation on several ideas drawn from studies in anthropology.

Finally, Part III deals with prospects for our approach and some views on prospects for the area in general. The contribution in Chapter 7 is based on the idea that emotions are, as discussed by anthropologists, a cross-cultural link, i.e., they are universally comprehensible. This is an important lesson from anthropology that is also relevant to the future of interoperability of MAS. We present in that chapter initial simulation results in support of an "emotional stance" for MAS. The "future" figure of speech has two aspects here: that chapter points to future directions in DAI and is itself a part of our enterprise on interoperation which we initially intended as a future project, but which we were persuaded to start pursuing, to a certain extent, in the process of developing this thesis, because of its connections and relevance to the subject-matter of this thesis, as well as its intrinsic interest. All the discussion related to the main chapters of the thesis (i.e., Chapters 3–7) is set out in Chapter 8 where, based on our anthropological approach, we indicate possible future directions in DAI, in particular concerning interoperability of MAS, consider some of the limitations of the thesis, and make contrasts with differing points of view that have been presented in the literature. The final chapter reviews the main points made throughout the thesis, comments on future work for each of the "contributions," and provides final remarks.

Chapter 2

Background

This chapter is not intended as a traditional background chapter of a PhD thesis, where one presents the major concepts of the area and surveys all the work related to each of them. We have chosen to use a rather short chapter mentioning the papers we appreciate the most or the ones that are particularly relevant to the development of the thesis or for the discussions to be presented in it (however, in addition, we include towards the end of this chapter a large collection of references, which represent the many different approaches or areas of work on DAI). The reason we thought it right to make this choice follows from the peculiarities of the main area of the thesis (as we have mentioned, it is a multidisciplinary one). The field of agent-based systems is both a relatively young field of research and one that has had recent mass attention, the result being the existence of numerous easily-available surveys and introductory papers for the area.

Examples of large surveys are (Wooldridge and Jennings 1995; Green *et al.* 1997)[1]; introductory papers with plenty of references can be found in (Wooldridge 1998; Sycara 1998; Jennings, Sycara, and Wooldridge 1998); there is a also an extensive DAI bibliography in (Wooldridge, Müller, and Tambe 1996)[2]. Another important class of material in this respect comprises books with collections of selected (published) papers; see, e.g., (Huhns and Singh 1997; Bond and Gasser 1988; Avouris and Gasser 1992; O'Hare and Jennings 1996)[3]. A first introductory book for the area is to be released (Weiß 1999), from which we are aware of two chapters: one that reviews formal methods used in DAI (Singh, Rao, and Georgeff 1999) and

---

[1] An electronic repository of agents material can be found at the AgentLink WWW page: `http://www.AgentLink.org/`.

[2] Some DAI bibliographies are also available in the WWW. See, e.g., URLs <`http://www.cs.twsu.edu/~haynes/dai.bib`> and <`http://www.cs.helsinki.fi/~hhelin/agents/agent-bib.html`>; for bibliographies in all areas of Computer Science, helpful addresses are <`http://theory.lcs.mit.edu/~dmjones/hbp/bibsearch.html`>, <`http://liinwww.ira.uka.de/bibliography/index.html`>, and <`http://www.dfki.uni-sb.de/imedia/lidos/bibtex/`>.

[3] For a more in-depth study of the area it is important to refer also to edited books with papers from important conferences, see (Huhns 1987; Gasser and Huhns 1989a), and in particular the ICMAS proceedings (Lesser and Gasser 1995; Durfee 1996; Demazeau 1998).

another on industrial applications of DAI (Parunak 1999). There are also specialised surveys, e.g., (Müller 1999) on agent architectures and (Iglesias, Garijo, and González 1999) on agent-oriented methodologies of software engineering. In the area of *Social Simulation* and *Artificial Societies*, major collections of papers are (Gilbert and Doran 1994; Gilbert and Conte 1995; Sichman, Conte, and Gilbert 1998) (for an introduction to the area see (Doran and Gilbert 1994)); and a collection of papers in the area of simulation of institutions, organisations and groups is (Prietula, Carley, and Gasser 1998).

We emphasise that the papers cited in this chapter are from the area of Multi-Agent Systems only. This being a multidisciplinary thesis, one would expect to see work in other related areas alluded to in this chapter too. Nevertheless, we considered it more appropriate to introduce the relevant material from other areas in the chapters to which they are instrumental. It is important to remark, for the reader expecting to see mentions of the literature on *mobile agents* (Chess *et al.* 1995) in this thesis, that such a class of agents is not significantly related to this thesis, for that type of work is concerned with details of network infrastructure for transportation of software from one site to another (including issues such as security), aiming in particular at application in mobile computation. Further, that area assumes that agents find standard environments in every site they visit: not only standard script interpreters for the execution of the code ("host-independent execution environment for itinerant agent programs" (Chess *et al.* 1995)), but also for communication (if needed), e.g. the standard for agent communication that we describe in Section 2.2.3, which is what is relevant for our discussion. Therefore, they face no such problems as autonomous agents migrating among MAS that display different "cultures", which this thesis aims at addressing. Also the word *adaptation*, which we use and which is sometimes found in the MAS literature, can be misleading; in general it relates to (machine) learning in general in a MAS context (see references in Section 2.3), as opposed to adjustment to cultural variety as intended here.

This chapter is organised as follows. The first section mentions the papers that we considered to have had a major impact in the line of evolution of the work in the area, or the ones we considered might have such an impact in the future. That is, that section provides an unashamedly biased collection of references. Due to the notorious vagueness of the word *agent* (cf. (Wooldridge 1998; Franklin and Graesser 1997)), which is presently used in many areas of Computer Science, a personal account of important or historical papers at least helps to give a more precise idea of the notion of agent employed in the thesis, or at any rate define the area of agent work to which we refer (by determining the research community with which the thesis is identified). Next, in Section 2.2, we present brief descriptions of relevant work. Finally Section 2.3 contains a list of references to somewhat influential papers in the area (representing the several different paradigms of agent research), or papers that are also relevant—but to a lesser extent—to the thesis (that very list of (types of) work in DAI is useful for our formulation of the idea of cultural variety of MAS in Section 4.2).

2.1 Papers of Historical Importance

The theory of *speech acts* (Cohen and Levesque 1990b; Cohen and Perrault 1979; Allen and Perrault 1980) (see also (Allen 1987)), which relies on the work of the philosophers of language J.R.Searle and J.L.Austin (Searle 1969; Austin 1962), permeates all the existing work on agent communication. The work of Cohen and Levesque on *joint intentions* in (1990a) is also influential for all researchers working on notions of commitment.

The work in (Campbell and d'Inverno 1990) was also based on speech acts (in the sense that the intentionality of an exchanged message is explicitly represented in it) and proposed a taxonomy of twenty-one *tones of communication*, each one to be associated with a *knowledge interchange protocol* which states an efficient organisation of interactions between agents so as to fulfill the expectations of that particular communication tone.

Also of considerable importance has been the work in (Werner 1989), with its first attempt to give a pragmatic interpretation for a communication language where again an *illocutionary force*[4] is made explicit, the *propositional content* being expressed in a propositional temporal language. This pragmatic interpretation entails an account of how intentional states, which are represented formally, are changed by communicative exchange. This work is also pioneering with respect to its allusion to *social roles.*

In (Castelfranchi 1990) a *social simulation* approach to MAS was proposed, in which it is important to considered (i) why autonomous agents would engage in social interaction and (ii) how agents could turn their own goals into social ones (i.e., how they could persuade other agents to adopt their goals). Castelfranchi, accordingly, first discusses notions such as *social dependence*, in connection with (i), and *social power*, in connection with (ii), at a time where benevolence was taken for granted (that is, typical agents considered in research at the time were not autonomous, in the sense that they were supposed to accept whatever tasks were asked of them). In a series of

---

[4] To understand what *illocutionary* means, one must bear in mind that not all declarative utterances state facts; some (called *performatives*) constitute the performance of an action, so they are not "true" or "false," but rather are *felicitous* or not. The acts performed *in* making those utterances are *illocutionary acts* (other types of *speech acts* are: *locutionary acts*—the utterances themselves, and *perlocutionary acts*—the ones performed *by* making the utterances). "For instance, stating, requesting, warning, ordering, apologizing, are claimed to be different types of illocutionary acts, each of which is said to have a unique *illocutionary force* that somehow characterizes the nature of the act. Each illocutionary act contains *propositional content* that specifies what is being requested, warned about, ordered, etc." (Cohen and Perrault 1979).

papers (some of which are mentioned in Section 2.3), they developed a line of work that culminated with (Conte and Castelfranchi 1995), where they elaborate on the *micro-macro link problem*, a fundamental issue in the social sciences which is paramount also in MAS: explaining social (macro level) phenomena from individual (micro level) traits is notoriously difficult (especially in systems of cognitive agents, but also in reactive ones, as mentioned in the next paragraph), as much as it is difficult to understand precisely what influences the social level has on individual cognition.

An important class of work, in our opinion, consists of papers that seek inspiration in varied sources: the decentralised aspect of DAI allows for a diversity of analogies. Inspiration to solve problems in DAI can come from anywhere in the sciences: from physics to anthropology[5]. In (Perram and Demazeau 1997) a type of *reactive* multi-agent system is presented, and shown to be a generalisation of statistical mechanical systems, which are employed by certain physicists. With this they were able to confirm the difficulty of determining analytically the overall behaviour of a MAS from knowledge of the individual agents and their interactions, the alternative being simulation. Nevertheless, they were able to alleviate the problem by proposing a methodology for studying reactive MAS which resembles the Monte Carlo computer simulation technique used by physicists for statistical mechanics.

A basic architecture for cognitive autonomous agents was presented in (Demazeau and Müller 1990), which proposed a "Decentralized AI" approach (one that starts from the autonomous entities rather than the global system) in contrast with the established view of DAI. In (Demazeau 1995) the whole MAGMA approach to MAS is summarised, including the structure of a MAS in that framework and the communication language and protocols they propose. The micro-macro link mentioned above is also discussed there.

A recent work which is quite revolutionary, in the sense that it presents ideas that not only are new ones but are in opposition to those that are commonly agreed upon in the area, is the one presented in (Doran 1998a) (see also (Doran 1998b)). The most interesting facet of that work, which is quite in resonance with this thesis, is that it is a result of a long experience in building computer simulations based on multi-agent systems for studies in archaeology and anthropology (Doran *et al.* 1994; Doran and Palmer 1995). Doran claims that *collective misbelief* can be beneficial to a society of agents, where "misbelief" is intended to mean a belief in a proposition that does not correspond to the actual state of affairs of the world (i.e., environment) in which the agents are embedded. One of the types of misbelief studied by Doran, and shown

---

[5] Indeed, inspiration to solve DAI problems can come not only from the sciences but any area of human knowledge or activity: e.g., the work on "believable agents" mentioned later draws from dramatic theory.

by simulation to have beneficial effects in societies of agents, is that of *cults*; that is, the belief in non-existing agents or, better said, "agentifying" resources (very much in the way "primitive"[6] tribes tend to personify their resources, e.g., mountains or fountains). In his presentation for (Doran 1998b) he also mentions that emotions in agents may be a suitable way for designers of MAS to control their agents by means of attachment to cults (but see our discussion on this idea in Section 8.3).

We now only allude to further important works in the history of DAI, without going into detail. The work on emergence of social laws in (Shoham and Tennenholtz 1992) is also pioneering and influential for the more prolific work on *social norms* by Conte and Castelfranchi (see Section 2.3 for references). One of the first researchers to initiate work on DAI is Lesser, he and his colleagues producing substantial work on coordination mechanisms: for the most recent paper with the bulk of their work see (Lesser 1998); see also (Durfee and Lesser 1991; Decker and Lesser 1995). Pollack is an exponent of this area with respect to planning (1992) (see also (Pollack 1998)). Huhns's Carnot project has first used agents as mediators to legacy systems (Collet, Huhns, and Shen 1991; Huhns 1995). The *Contract Net Protocol* (Davis and Smith 1983; Smith 1980) is undoubtedly the most well-known negotiation protocol. An early work by Georgeff (see references to his now well-known work on BDI agents in Section 2.3) is (Georgeff 1987).

Finally, we mention two papers that were important in setting the foundations of DAI in sociological terms: these are (Gasser 1991) and (Hewitt 1991). In particular, the work on introducing the principles underlying an intrinsically social conception of knowledge and action that are instrumental for the foundations of DAI, presented in (Gasser 1991), is relevant for this thesis insofar as it has similar[7] purposes to the ones of our work in Chapter 4: there, we seek principles from anthropology rather than sociology for contributing towards a foundation of DAI with the added complication of migration between different societies (i.e., interoperability of MAS) which we set about to understand in this thesis.

2.2 A Collection of Relevant Work

This section contains three brief descriptions of important work in DAI. Basic understanding of this material is useful for some of the work or discussion in this thesis.

---

[6] The reason for the quotes around that word is to mark the frequently-encountered arrogance of Western culture in referring to others which is implicit in that word; it is, however, unavoidable to use it.

[7] Only *similar* purposes, because Chapter 4 is, clearly, a modest contribution towards that direction.

2.2.1 Luck and d'Inverno's Formal Framework for Agency

A formal framework for the specification of agent theories was introduced by Luck and d'Inverno (1995); it is based on the Z formal specification language (Spivey 1992; Potter, Sinclair, and Till 1996); see Appendix B, Section B.1, for a brief account of the basics of Z. In that paper they aim at providing a principled theory of agency and autonomy by describing a three-tiered hierarchy of entities where *agents* are *objects* with *goals* and *autonomous agents* are *agents* with *motivations*. In Z, this is formalised quite simply:

$$[Action, Attribute, Motivation]$$
$$Goal == \mathbb{P}\, Attribute$$

$$\begin{array}{|l}
\hline
\textit{Object} \\
capableof : \mathbb{P}\, Action \\
attributes : \mathbb{P}\, Attribute \\
\hline
\end{array}$$

$$\begin{array}{|l}
\hline
\textit{Agent} \\
Object \\
goals : \mathbb{P}\, Goals \\
\hline
goals \neq \{\ \} \\
\hline
\end{array}$$

$$\begin{array}{|l}
\hline
\textit{AutonomousAgent} \\
Agentmotivations : \mathbb{P}\, Motivation \\
\hline
motivations \neq \{\ \} \\
\hline
\end{array}$$

They presented further notions of multi-agent systems in that paper (see also (d'Inverno 1998) and, for issues of developing simulations from their formal framework, see (Luck, Griffiths, and d'Inverno 1997)). This framework was subsequently extended and used to formalise a varied collection of techniques or approaches to MAS in a series of papers. We list those we are aware of below:

- Davis and Smith's Contract Net Protocol (1983) in (Luck and d'Inverno 1996);
- Sichman's Social Reasoning Mechanism (1998) in (d'Inverno and Luck 1996a);
- Kinny's **Agentis** Agent Interaction Model (1999) in (d'Inverno, Kinny, and Luck 1998);
- Rao's AgentSpeak(L) Programming Language (1996a) in (d'Inverno and Luck 1998).

We have profited from Luck and d'Inverno's framework in the specification of our notions of ascription of ontological descriptions and retrieval of taxonomical relations from them, which we present in Chapter 5.

2.2.2 Sichman's Social Reasoning Mechanism

One of the agent theories that has been formalised according to Luck and d'Inverno's formal framework (presented above) is Sichman's *social reasoning* mechanism (Sichman 1998; Sichman and Demazeau 1995; Sichman *et al.* 1994; Sichman and Demazeau 1994), which is based on Castelfranchi's et al. concept of *social dependence* (Castelfranchi 1990; Castelfranchi, Miceli, and Cesta 1992). That formalisation can be found in (d'Inverno and Luck 1996a).

Social reasoning is a term that refers to an agent reasoning about the other agents in the system. Sichman's social reasoning mechanism is included in his DEPINT (1998) system, a multi-agent simulation tool. Sichman's work is relevant to this thesis insofar as it is one practical means of allowing for open societies; that is, it is a mechanism to be used in societies where autonomous agents can leave and enter "freely." Although relevant here, that work does not propose a solution to the problems we tackle in this thesis. It is an example of the case where agents who enter the society necessarily share that "culture." This is discussed in Section 4.2, therefore it is useful to include the main ideas of Sichman's work, which we do next.

That social reasoning mechanism considers an agent's plans and goals as well as those of the other agents and uses them to calculate a network of *dependence relations*. These allow the classification of an agent's situation in regards to a particular goal into a taxonomy of "goal situations" and "dependence situations," basically by considering whether the agent can achieve that goal by itself or only with the help of others, and in the latter case, whether those other agents depend on itself to achieve that goal or some other goal, and whether these findings are based on the agent's own plans or those of the others.

Reckoned dependence situations help an agent adapt to particular situations in a society at a certain time to the extent that they produce knowledge of which plans are *feasible* at that time (those whose every action can be performed by at least one of the agents), and consequently which goals are *achievable* (i.e., those for which there is a feasible plan). Further, they provide for coalition formation, given that agents can use the inferred dependence situations to decide the most suitable partners with whom to cooperate. They also help with the process of belief revision: when receiving proposals for cooperation, the receiving agent can inform the others, if that is the case, of their wrongly-inferred dependence relations regarding itself, which entails that they have false or incomplete beliefs about that agent's plans or goals.

2.2.3 KSE's Agent Communication Language

The Agent Communication Language (ACL) from the ARPA Knowledge Sharing Effort (KSE) (Genesereth and Ketchpel 1994; Genesereth and Singh 1994; Singh, Genesereth, and Syed

1995) is perhaps the first widespread technology for agent communication to include some of the complex concepts for high-level communication in the DAI literature (to which we alluded earlier in this chapter). It is intended as a standard for agent interaction: an opposite approach from that of this thesis to interoperability (Wiederhold 1994) of multi-agent systems (it is also intended to deal with legacy systems through *mediation* (Wiederhold and Genesereth 1995; Collet, Huhns, and Shen 1991)). This is precisely why it is important to introduce this approach here: we refer to it in the discussion in Section 8.1.

It is not our aim to present ACL in detail here. It is sufficient to recall that ACL is composed[8] of: i) ontologies of vocabularies (dictionaries) (Gruber 1993b; 1993a; Wiederhold 1994); ii) a language for the codification of the actual messages to be exchanged by agents, called Knowledge Interchange Format (KIF)—basically, a (prefix version of) first-order logical language with extensions (Genesereth, Fikes, and others 1992); and iii) a language for adding context to the messages called Knowledge Query and Manipulation Language (KQML) (Finin *et al.* 1994b; 1994a; Finin, Fritzson, and McKay 1992; Labrou 1996; Labrou and Finin 1994; Mayfield, Labrou, and Finin 1995; 1996). A proper ACL message is a KQML expression in which the arguments are KIF sentences formed of words from a particular vocabulary (ontology) (Genesereth and Ketchpel 1994). KQML is based on the theory of speech acts (we have given references to this in Section 2.1); it is formed of an extensible set of *performatives* which determine the sorts of interactions agents can have (Finin *et al.* 1994a).

We now exemplify the use of some of KQML's *reserved performatives* and the KIF notation. We do this by specifying with KQML performatives—with message contents coded in KIF—the main ideas in this very short (arbitrary) extract of the dialogue[9] in the *Symposium* of Plato (Lamb 1925; 214d–215a) (some classical texts, including this, can also be found in the WWW: see (Crane nd)):

> ERYXIMACHUS: Well, do that if you like, praise Socrates.
>
> ALCIBIADES: You mean it? You think I had better, Eryximachus? Am I to set upon the fellow and have my revenge before you all?
>
> SOCRATES: Here, what are you about,—to make fun of me with your praises, or what?
>
> ALCIBIADES: I shall speak the truth; now, will you permit me?
>
> SOCRATES: Ah well, so long as it is the truth, I permit you and command you to speak.

---

[8] There are links to all part of the KSE's ACL project, as well as a general repository of agent-related material, at the UMBC AgentWeb, URL **`<http://www.cs.umbc.edu/agents/>`**.

[9] Changed by us to a script format, as is the case in other editions of that dialogue. We found that particular translation more suitable for our purposes, but a script form is also helpful here.

The practical issues in this dialogue (i.e., removing all the poetic effects in the conversation) can be "translated" into KQML as shown in Figure 2.1 (note that requesting and granting permission is not very clearly expressible in KQML). In the figure, **symposium** is a hypothetical ontology giving definitions to the constants used in the KIF messages[10]. The agent called **facilitator** that receives some of the messages below is part of the *federated system* used by KQML (Genesereth and Ketchpel 1994); facilitators derive from and generalise the concept of mediator (see references for mediation above). The performative **achieve** means that the **:sender** agent requests the **:receiver** agent to perform all necessary actions so that **:content** become true in the "world;" the performative **advertise** means that the **:sender** agent is committing itself to perform the performative in the **:content** field when receiving it from the **:receiver** agent; the **untell** performative informs that the **:content** is not present in (or is not provable from) the knowledge base of the **:sender** agent (whereas **deny**, which is not in the example provided, is used when the negation of the **:content** is in the knowledge base); the remaining performatives used in the example in Figure 2.1 (**ask-if**, **tell** and **broadcast**) are self-explanatory.

This example is by no means an extensive selection of KQML reserved performatives (or the KIF notation): the intent in providing the example was just to give a flavour of KQML/KIF communications; further details can be found in the literature we have provided in this section. Note that it is rather unusual in a KQML-speaking agent scenario to have an agent **broadcast**ing an **advertise** of a performative immediately after being sent precisely that performative (in the case an **achieve**). (In humans, this type of behaviour is probably more irritating than unusual, though.) Nonetheless, that is not completely inconceivable in a MAS; it is possible that the agent was not prepared to perform that **achieve** at the time of receiving it, and only later decided that if anyone requested that same thing, it would perform accordingly.

There is another more recent enterprise in line with the "standardising approach." This is a non-profitable organisation called Foundation for Intelligent Physical Agents (FIPA) (FIPA 1998). They have specified several aspects of agent technology; for their communication language particularly, see (FIPA 1997). Their specifications include a speech-act-based language which is very similar to KQML; FIPA actually designed their language purposely to be compatible with KQML (e.g., to allow moving from one to the other easily) (Steiner 1998). Even more recently The Agent Society was created, with similar aims (see their Reference Model for Open Agent Systems in (The Agent Society 1998)).

---

[10] Conversation identifiers in KQML examples usually start with id0. We have started with id1 because the extract of the dialogue above follows closely a thread of conversation that was not included here.

```
(achieve
     :sender        Eryximachus
     :receiver      Alcibiades
     :in-reply-to   id1
     :reply-with    id2
     :language      KIF
     :ontology      symposium
     :content       "(praise socrates)" )
(broadcast
     :sender        Alcibiades
     :receiver      facilitator
     :reply-with    id3
     :language      KQML
     :ontology      kqml-ontology
     :content       (advertise
                         :sender        Alcibiades
                         :reply-with    id4
                         :language      KQML
                         :ontology      kqml-ontology
                         :content       (achieve
                                             :receiver      Alcibiades
                                             :reply-with    id4
                                             :language      KIF
                                             :ontology      symposium
                                             :content       "(praise socrates)" ) ) )
(ask-if
     :sender        Socrates
     :receiver      Alcibiades
     :in-reply-to   id4
     :reply-with    id5
     :language      KIF
     :ontology      symposium
     :content       "(=> (praise socrates) (make-fun socrates))" )
(untell
     :sender        Alcibiades
     :receiver      Socrates
     :in-reply-to   id5
     :reply-with    id5
     :language      KIF
     :ontology      symposium
     :content       "(=> (praise socrates) (make-fun socrates) )" )
(tell
     :sender        Alcibiades
     :receiver      Socrates
     :in-reply-to   id5
     :reply-with    id5
     :language      KIF
     :ontology      symposium
     :content       "(=> (praise socrates) (speak-truth-about socrates)) ")
(achieve
     :sender        Socrates
     :receiver      Alcibiades
     :in-reply-to   id4
     :reply-with    id6
     :language      KIF
     :ontology      symposium
     :content       "(praise socrates)" )
```

Figure 2.1: A KQML Equivalent of an Extract of a Platonic Dialogue

2.3 Further Papers in the General Area

Another undeniable characteristic of research on MAS, besides those we mentioned in the beginning of the chapter, is that the area encompasses work of drastically contrasting natures. This is one of the reasons why the problem of interoperability is particularly salient in MAS: it implies the existence of several "cultures" in MAS (the other reason for it being autonomous evolution; see Section 4.2). Unlike the papers presented earlier in this chapter, describing the ones below would not at any rate be essential to the working-out of our thesis[11]; all we need to show here is that such a variety of approaches exists in the area. We therefore only provide a list of major references for each of these different conceptions of or approaches to MAS (or different areas of work within the cognitive approach to agents, which is most relevant to our work):

- Game-Theoretic approaches to coordination of self-interested agents, the most renowned being that of Rosenschein and colleagues: (Ephrati and Rosenschein 1994; Fenster, Kraus, and Rosenschein 1995; Rosenschein and Genesereth 1985; Rosenschein and Zlotkin 1994; Zlotkin and Rosenschein 1994) and Durfee and colleagues' "Recursive Modeling Method" (Durfee 1995; Gmytrasiewicz and Durfee 1993; Gmytrasiewicz 1996; Vidal and Durfee 1996a; 1996b);
- Market-Oriented approaches, notably the work by Wellman: (Wellman 1993; 1994; 1995; Mullen and Wellman 1996);
- BDI logics and abstract agent architectures based on the BDI model: (Rao and Georgeff 1992; 1995; Rao 1996b)—as a reference on belief logics and possible worlds semantics, see (Fagin and Halpern 1988);
- Practical architectures based on the BDI model:
  - IRMA (Bratman, Israel, and Pollack 1988; 1987; Pollack 1992)—which includes a means-ends reasoner that computes plans from first principles (Fikes and Nilsson 1971);
  - PRS (Georgeff and Lansky 1990);
  - dMARS (d'Inverno *et al.* 1998);
  - AgentSpeak(L) (Rao 1996a)—a programming language based on these previous experiences on BDI architectures;

---

[11] It is worth mentioning here that the work that is closely related to this thesis will be described or discussed properly at various points throughout the thesis. These include a subset of the works referred to in this Chapter and many others which are not mentioned here as they are from other sciences or do not help to characterise the area of MAS in our particular perspective, which is the purpose of this chapter.

- Work on agent communication based on speech acts (references to the speech act theory in general were given in Section 2.1): (Werner 1988; Vanderveken 1990; Haddadi 1995; Verharen 1997);

- In (Cohen and Levesque 1995), issues of agent communication language semantics are presented (in the form of criticism referring to semantic problems of KQML), and Smith *et al.* (1998) give semantic (in the form they propose in (Smith and Cohen 1996; 1995)) for their speech-acts based *conversation policies* drawing on Cohen and Levesque's theory of joint intentions (references in Section 2.1);

- Further references to the work of Castelfranchi and colleagues on social concepts as commitment, norms, values, trust, personality, micro-macro link, etc., are: (Castelfranchi and Falcone 1998; Castelfranchi, Conte, and Paolucci 1998; Castelfranchi, de Rosis, and Falcone 1997; Castelfranchi, Miceli, and Cesta 1992; Castelfranchi 1995; 1998; Conte and Castelfranchi 1995; 1994; Conte, Castelfranchi, and Dignum 1999);

- One of the exponents of formal work on DAI is Singh, in particular regarding *intentions* and *know-how*: (Singh and Asher 1991; Singh, Huhns, and Stephens 1993; Singh 1994); see also his work on agent specification (Singh 1996; 1998);

- One of the most cited papers in the area in recent years is (Shoham 1993), where AGENT0 is presented, the first attempt to create a programming language that is "*agent-oriented*;"

- Specification of agents in temporal logics, in particular the language Concurrent METATEM: (Fisher 1995; Fisher and Wooldridge 1994; Fisher 1999);

- Further references on computational models of organisations: (Bond 1990; Carley 1998)—recall that we have also provided general references to work on application of MAS techniques for simulations in the social sciences in the beginning of the chapter;

- Learning in MAS settings, usually employing some sort of reinforcement learning: (Weiß and Sen 1996; Weiß 1997; Sandholm and Crites 1995); for variations on Q-learning, see (Pebody 1997);

- Work on cultural evolution and "memetics" (relevant to our work in Chapter 7): (Reynolds 1994; Hales 1998a; 1998b; 1998c); for applications of genetic algorithms to MAS, see (Manela 1993);

- Information agents, i.e., those that are mediators to large databases or those that gather information in the Internet: (Klusch and Shehory 1996; Decker *et al.* 1997), and see also the references for mediation in Section 2.2.3;
- Architectures for agents with *motivations* (Norman and Long 1996; Norman 1997) which include *emotional* ones (relevant to our work in Chapter 7): (Aubé and Senteni 1996; 1995; Sloman and Poli 1996; Rizzo *et al.* 1997; Bates 1994; Hayes-Roth, Brownston, and van Gent 1995);
- Some other isolated papers with noteworthy work: Galliers first pointed out the positive role of conflicts (1990), a subject that has only recently attracted further attention; and a very interesting, recent work on commitments and roles is (Cavedon and Sonenberg 1998).

The present expansion of research on "agents" is such that this list could go on indefinitely. Nevertheless, the sample we have given is, to the best of our knowledge, representative of the main areas of research in MAS and sufficient for our consideration of "cultural variety" in Section 4.2.

Chapter 3

Adaptation at the Level of Language and Interaction

In this chapter we describe some previous work we have carried out on the linguistic features needed to support the migration of agents between societies that have different communication mechanisms (that is, we have dealt with the problem of an agent entering an existing society whose conventions of communication are unknown). The ability of agents to learn different communication conventions is fundamental for the migration problem as we have posed it, in virtue of communication being a crucial aspect of agent interaction (especially so for cognitive agents). This particular aspect of adaptation is also very important for our critique of current major approaches to interoperability, presented in Section 8.1; that is why it is important to present a compact version of our previous work here.

3.1 Introduction

We provide here an initial treatment of the problem of how an agent can become acquainted with the communication aspects (in particular, concerning communication protocols) of varied communicating societies, i.e., societies that do not use the same protocol set and not even the same languages used to deal with protocols. This work was first described in detail in (Bordini 1994) (see also (Bordini *et al.* 1995)), where we noted that migrant agents can use communication protocols specific to societies to which they migrate if they are able to interpret any *meta-language* in which the syntax and semantics of the protocol languages have been expressed in a formal way. These languages are the ones used for protocol description and for protocol execution which are employed in a particular target society. We have also noted that an *immigrant agent* must have access to some basic *meta-protocols* allowing them to perform minimal interactions which lead them to learning the actual protocols used in that society. The target societies, thus, must make available what we have called *public information* on their communication features. Such public information includes the syntax and formal semantics, expressed in a meta-language, of languages used to describe and execute protocols in that society, as well as the meta-protocols.

We state the general ideas about the architecture of the sub-agent responsible for adaptation at the communication level in Section 3.2. Section 3.3 includes comments on issues of the public information on communication features that a target society must offer for an immigrant agent to adapt to it, including protocol languages and meta-protocols of a sample target society (most of the detailed material is in Appendix A). Section 3.4 discusses the formal semantics given to the example protocol languages we provide, using two different formal frameworks that can serve as meta-languages; it also describes concisely some aspects of an implementation of these ideas.

In order to keep this section brief, given that it is only supplementary to the main thesis to be presented here, we have placed most of the details in Appendix A: the syntax of the protocol description language of our sample target society is in Section A.1 (the protocol execution language is below in Section 3.3.2), the formal semantics of these two languages in Section A.2, and the meta-protocols in Section A.3.

3.2 Overview of the Architecture

This section covers the architecture and functioning of the *Communication Sub-Agent* (CSA). We consider here that agents are composed of **sub-agents**, which work together to build up the agent's behaviour (resembling the outlook of (Minsky 1986)). This has been called the recursive principle of MAS by Demazeau (1995), and also discussed by Gasser (1995). A sub-agent is an autonomous, parallel part of an agent that is in charge of a specific task in the functioning of the agent (e.g., planning, belief revision, perception, and so on). We view these sub-agents in a similar way to the parallel components of an agent in (Corrêa and Coelho 1995); also in the approach of (Fisher 1999) agents are composed of sub-agents. The CSA is the sub-agent responsible for the adaptation of an immigrant agent to a new society at the level of language and interaction, and then for controlling the conversations the agent will carry on in that society.

Section 3.2.1 presents our distinction of protocol description language and protocol execution language. The remaining part of this section deals with the architecture of the CSA particularly. In Section 3.2.2, we consider the adaptation phase of an immigrant agent (at the level of communication protocols), and Section 3.2.3 explains the functioning of the CSA while the agent is interacting in the new society.

3.2.1 Protocol Languages

We first need to clarify the distinction we have made between protocol description language and protocol execution language. The *Protocol Description Language* (PDL) we have used as an example (to be commented on in Section 3.3.1) is basically the language presented in (Populaire *et al.* 1993). That work expounds a universal language that can be used to describe a set of protocols. It has been shown to be generic enough to describe a number of protocols that have been presented in the literature. Notwithstanding, that language may not be generic enough to describe all sorts of interactions among agents (e.g., the ones we will consider below). Thus, different societies may have intrinsically different communication features due to the different functionalities they have. Furthermore, if the agents are capable of discussing the protocols they use, they may promote changes to those protocols and consequently to the PDL as well. Therefore, different societies starting from the same features may arrive at different versions of the protocol languages they use. Dealing with this is in tune with our goal of allowing interoperation despite the fact that societies of agents can follow different paths of historical evolution.

The PDL in (Populaire *et al.* 1993) also allows the coexistence of different versions of the same protocol. This introduces dynamics to the protocols used in the society, since they can be improved (even though it is still an open research issue how an agent can itself propose modifications to the protocols it uses). However, we have found that the form for agents controlling the changes in the conversation state during the execution of protocol transitions proposed in that work was unsuitable for the existence of multiple versions of a protocol. We do not discuss our alternative approach in detail here (see (Bordini *et al.* 1995)), but one of the differences in our approach is that both sending and receiving a message causes a state change in the conversational structure of the agent: this allows agents to keep track of every transition in the protocol, which is the only way to make it absolutely sure that there is no mismatch of protocol versions between the communicating agents. Also, this has been a crucial aspect in the specification of the semantics of the protocol execution language (see Section 3.4).

Another relevant point for our notion of protocol languages is the kind of interaction structures we are interested in providing to the agents. Based on the account of complex structures for agent interaction proposed in (da Rocha Costa and Coelho nd), we consider that agents may be able to decide to have parallel conversations, as well as to decide to interrupt a conversation and start a new one. The latter may occur whenever the agent realises it cannot fulfill all requirements needed to proceed with a conversation (e.g., lack of information). Thus, it needs to start a new conversation to reach those requirements and then resume the previous conversation from the state in which it had been interrupted. Parallel conversations can be both independent conversations or different alternatives for a conversation (many attempts to achieve a certain purpose). A *Protocol Execution Language* (PEL) makes explicit the forms of handling protocols (which determine how complex a conversation using them can be) that are specific to the society using it.

3.2.2 CSA in a Migration Process

We show in Figure 3.1 the overall functioning of the CSA while trying to adapt to a new society in a migration process. The CSA of the *Immigrant Agent* (**ImAg**) must access the *Public Information on Communication* (**PIC**) contained in a global area (i.e., a public one) of the Target Society (**TS**). The part of **TS** that appears in the figure shows this global information area and two agents (**Ag1** and **Ag2**) along with their CSAs (although it is not mandatory that they have such architecture, unless they too are migrant agents).

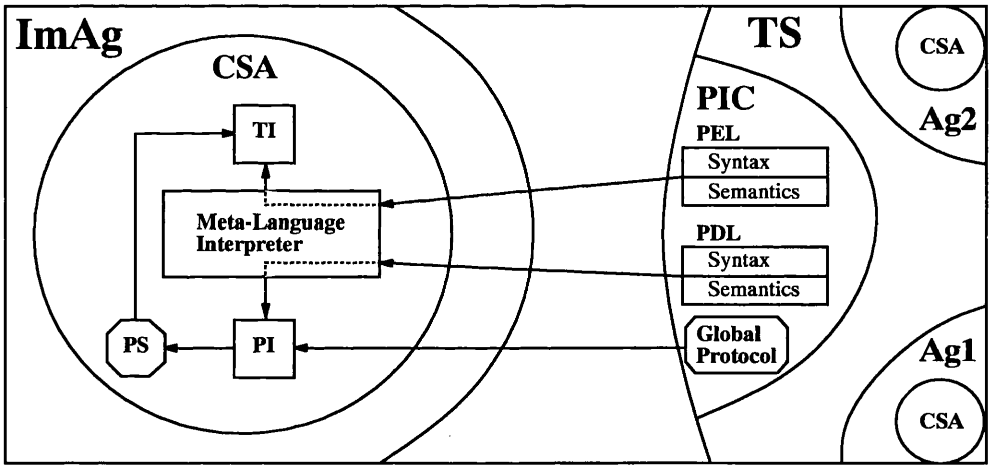


Figure 3.1: The Functioning of the CSA in a Migration Process

The CSA of the immigrant agent starts the adaptation process by reading the specification of the syntax and the semantics of the Protocol Description Language (**PDL**). This specification is written using the meta-language the CSA can interpret (see the **Meta-Language Interpreter** in the Figure). From this specification, the CSA generates a Protocol Interpreter (**PI**, in the figure) which is capable of translating any protocol specification expressed in the way that is typical to the target society into the CSA's *Protocol Internal Structure* (**PS**, in the figure). This is the structure of protocols to be used by the *Transition Interpreter* (**TI**), which executes the protocol transitions when the immigrant agent communicates. The **TI** is generated by the CSA from the formal specification of the Protocol Execution Language (**PEL**), also via meta-language interpreter. We refer to "execution of transitions" because, as in (Populaire *et al.* 1993), we also assume that transmissions between two agents include both the message and a reference to the next state (thus determining the possible transitions for continuing the conversation). That is, the whole protocol transition is transmitted. The details of the functioning of the transition interpreter are given in the next section.

The global area in the target society containing the information about communication conventions (i.e., **PIC**) includes the **Global Protocol**, which is a set of *meta-protocols*. We call meta-protocol a protocol that deals with other protocols; that is, their messages contain protocols (or parts of protocols). In section 3.3.3 we show two situations in which it is necessary for the agents to exchange (parts of) protocols, in such a way that the messages exchanged contain protocols. These meta-protocols must be described in **PIC** using the protocol description language whose formal specification is also there, allowing the CSA of the immigrant agent to interpret them using the **PI** just created, and adding them to the protocols it knows already (**PS**, in the figure). This is, in fact, straightforward since meta-protocols are also protocols, so they can be described as the actual protocols of that society.

3.2.3 CSA Executing Protocols

We now concentrate on the functioning of CSA when an immigrant agent is interacting in the society it has joined most recently (after the adaptation phase accomplished during the migration process, described above). This is shown in Figure 3.2, where **TI**, **PI** and **PS** are the same as in Figure 3.1. However, the principal functional element in the architecture now is **TI** instead of the meta-language interpreter (used only in the process of adaptation at the level of Language and Interaction). The transition interpreter is responsible for all transition exchanges between that agent (**ImAg**) and the other ones in the **Environment**. Each transition executed by **TI** causes the updating of the **Conversational Structure**.

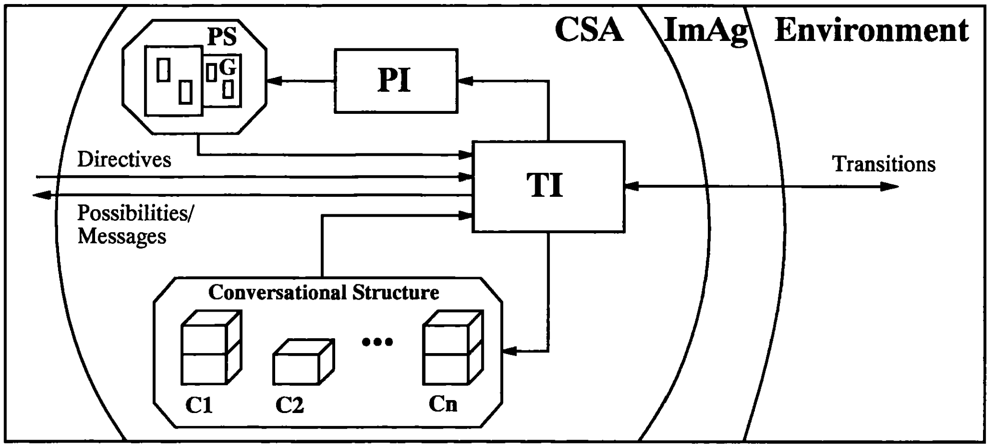


Figure 3.2: The Functioning of the CSA while Executing Protocols

If a transition is arriving at **ImAg**, **TI** sends the message content of that transition for the other sub-agents to interpret it pragmatically (this is the meaning of the arrow directed from **TI** towards the inside of the agent—the space to the left in the figure). Considering the idea that message interpretation is beyond communication (Castelfranchi 1990), it is reasonable here to suppose that this is accomplished by other sub-agents[1] in a conventional agent architecture (e.g., those responsible for belief management, planning, reasoning about the others, etc.). One concrete possibility is for the pragmatic interpretation of messages to be accomplished using the approach in (Werner 1989).

In the case when it is **ImAg**'s turn to send a message, the **TI** checks the current state of that conversation in the Conversational Structure (see figure), consults that protocol state in **PS** and finally provides the other sub-agents with the alternatives for continuing the conversation (these are referred to by the term "possibilities" above the arrow pointing to inside the agent in the figure). Again it is an outside decision[2]: possibly a sub-agent responsible for planning will decide which is the best transition among the available ones. The CSA then receives, from the other sub-agents, the communication directives (see the arrow from the inside of the agent towards the CSA). These are instructions concerning the structures of interactions (given by the conversational structure), and can prefix the chosen transition (see Section 3.3.2 for examples).

In the Conversational Structure, parallel conversations may exist (in the figure: $\mathbf{C1}, \mathbf{C2}, \ldots, \mathbf{Cn}$). The CSA represents a conversation by recording its current state (denoted by a box in the figure). One can notice that the structure of each of the conversations occurring in parallel is a "stack" of states (denoted by a stack of boxes in the figure). This allows for interrupted conversations: the stack keeps track of the current states in which each conversation was interrupted. When a conversation reaches its end, the one below it (if there is any) in the stack is resumed. Thus, the conversations taking place at any one time are the ones whose current states are at the top of each (parallel) conversation stack. These are the features specific to the PEL given as example in Section 3.3.2.

Another case to be handled by the **TI** is when the received message is a message from a meta-protocol (in other words, it may contain a whole, or part of, a protocol). In that case,

---

[1] It is arguable that the interpretation of an incoming message is accomplished by a collection of sub-agents, rather in a "multi-agent" fashion. This resembles the idea of an agent itself being implemented as a multi-agent system (this refers to Demazeau's recursive principle (1995) that we mentioned earlier). However, it is also possible that the architecture includes an interpreter sub-agent, which decodes the message and determines which are the sub-agents that should deal with each aspect of the message. In any case, the details of the architecture of the whole agent is a general task of the MAS area, not the subject of a particular research exercise; we concentrate here on the "communication sub-agent" which adapts the agent to the communication conventions of target societies. We therefore do not discuss message interpretation any further, but the contribution in Chapter 5 refers to one particular aspect of the understanding of contents of messages: that of ontological agreement.

[2] "Outside" in terms of the CSA, on which we focus here, but internal to the immigrant agent; that is, it is dealt with by other sub-agents.

the **TI** sends that message to the protocol interpreter (**PI**) which will update the CSA internal structure for protocols (**PS**). Recall that the protocol that is included in the incoming message is described using the PDL the immigrant agent has just learned (and is interpreted by its **PI**). Also notice that meta-protocols are represented in a specific *Global Protocol Area* (**G**, inside **PS**, in the figure) alongside the other protocols, so that they too can be consulted whenever necessary (i.e., when a conversation involving protocols is taking place).

3.3 Sample Public Information on Communication

An important assumption of this work on learning of protocol languages is the availability of Public Information on Communication in the target societies. Our model in (Bordini 1994) assumed that all societies provide the information needed for any immigrant agent's adaptation (but in the more general terms to be presented later in this thesis, this information can be obtained through "informant agents"). Below we present briefly the communication features of a sample, hypothetical society. We stress that these are the communication features of one possible society an immigrant agent (with the abilities recorded in the previous section) can "learn" (i.e., become able to understand and then interact with via meta-language interpretation).

3.3.1 Protocol Description Language

As noted previously, we have used basically the PDL in (Populaire *et al.* 1993), which exploits the several tones of communication presented in (Campbell and d'Inverno 1990). We have made some improvements in the production rules of its grammar; e.g., we have changed them so that a unique date is assigned to each protocol version. (Protocol versions, in that work, are identified by their creation date.)

We also introduce the possibility of *empty transitions*. Empty transitions are transitions between protocol states that occur when there is no message (i.e., a protocol transition without a message content). This corresponds to the agent allowing its interlocutor to continue sending information without its intervention. In analogy to human conversations, this is similar to one's interlocutor having nothing to say at a certain moment when one does not know whether that person has something to say or not. This is also one of the possibilities of interaction structures presented in (da Rocha Costa and Coelho nd).

Our version of that PDL is set out in Appendix A, Section A.1, and its formal semantics (as well as of the PEL, which we present next) is also given in Appendix A, both using the structural approach to operational semantics (Section A.2.1) and the denotational framework

(Section A.2.2).

3.3.2 Protocol Execution Language

In some societies there may exist a language used for agents to control the complex execution of the protocols they know, which are specified using a PDL type of language described previously. This language for the controlling of protocol execution makes it possible for agents to have the kind of complex interaction we have discussed in Section 3.2.1. We present its syntax (with informal semantic comments) below, and two specifications of its formal semantics are in Appendix A (Sections A.2.1 and A.2.2) as mentioned above. We have used the following conventions: terminal symbols are underlined and non-terminal symbols are enclosed in "<" and ">". Below, <KTransition> is the syntactic category related to the kinds of transitions permitted in the PDL, and <Name> uniquely identifies the conversation in which the transition occurs. The BNF for the PEL is:

<Name> : new <KTransition> this directive suggests that an agent wants to start a new conversation with the agent that will receive the message. Every new conversation, which may be in parallel with others, starts by the use of this directive. The very first conversation of an agent is created in parallel with no other. Our approach makes provision for special cases in the control of conversations. It may happen that another similar conversation (i.e., using the same protocol) is currently going on with other agents. Also, an agent can have different conversations with the same partner. This is possible because an agent can identify its conversations by their names and not by the state only (note that whenever the directive new is used, <Name> must be different from all others).

<Name> : nested <KTransition> should be used whenever the agent realises it is necessary to interrupt a conversation; another one, which starts with transition <KTransition>, will proceed, and when finished the previous one will be continued from exactly the state in which it was interrupted.

<Name> : resume is used to terminate prematurely a nested conversation that is going on, and to resume the previous one; this is needed in case it becomes possible to continue the interrupted conversation before the nested (auxiliary) one naturally reaches its end; i.e., the former grounds to engage in a nested conversation no longer hold.

<Name> : <KTransition> when no directive is added, the transition simply causes a change in the current state (i.e., the one on the top of the stack) of the conversation to which it refers, not in the conversational structure itself.

An actual message to be transmitted among agents in possible implementations of this communication mechanism is composed of: (i) the transition to be executed by the receiver agent, expressed using the PDL; (ii) possibly one of the communication directives of the PEL described above; (iii) the name of the conversation in which the transition occurs; and (iv) the identification of sender and receiver agents.

It is important to note that the protocols specified by means of the PDL may be nondeterministic. In other words, they may have transitions leading to different states on receipt of the same message; clearly, this is a quite important feature in communication mechanisms for autonomous agents. However, this nondeterminism does not exist during the execution of protocols since the PEL determines a specific transition to be executed. Recall that the transition to be executed has been chosen by the other sub-agents (as we mentioned in Section 3.2.3), which means that it is a general decision-making problem to be solved by the other (traditional) components in an agent[3] architecture, without any concern for the communication mechanism discussed here.

3.3.3 Meta-Protocols

Further, the immigrant agent needs some minimum meta-protocols that are also specific to each society. In the communication conventions of the sample society mentioned here, two meta-protocols are necessary, which we describe informally in this section (one deals with protocol mismatch failure and the other with protocol exchange). These two (meta) protocols are also specified using the PDL we mentioned in Section 3.3.1; the specification can be found in Appendix A, Section A.3.

The first proposed meta-protocol is there to be used in case a protocol version mismatch occurs and the other one to allow the immigrant agent (or indeed any agents) to ask others for the protocols being used in that society. We assume that our sample society adopts a decentralized management of protocols (as in (Populaire *et al.* 1993)); that is, each agent has its own copies of the protocols it uses explicitly represented in its internal structure. This is the reason why agents may happen to have different versions of a protocol, which introduces the possibility of protocol failures. Informally, this meta-protocol determines the following. Suppose agent A receives a transition from agent B that is not foreseen in A's version of the protocol they are using. In that case, agent A informs B that a version mismatch was detected and B replies informing A of the version it is using (identified by the creation date). If it happens that A has

---

[3] The agent that carries out that decision-making is necessarily the agent sending the message because of Populaire *et al.*'s approach of sending the next protocol state together with transmitted messages.

the older version, it asks *B* to send again the transition that caused the failure, so that *A* can amend its internal structure of protocols. In the opposite case, *A* informs *B* that the transition *B* previously sent is no longer valid. Then *A* transmits, on *B*'s request, the whole protocol state (a set of transitions) in which they are both currently engaged, so that *B* can choose another transition among the valid ones to progress with the conversation, after amending its own internal structure of protocols.

The second meta-protocol is rather simple, but not less important. The agent that starts a conversation according to this meta-protocol simply asks for a protocol *P* used in the society for a specific purpose. The agent receiving this message sends the whole protocol *P*, if it has any version of *P*, or otherwise replies saying that neither does it know any such protocol *P*. This is the meta-protocol that allows an agent to perform minimal interactions which eventually lead the agent to learning the actual protocols used in that target society. Once the agent takes hold of those protocols, it becomes able to engage in the relevant interactions necessary for the problem solving (or, more generally, social activity) that is being carried out by the target society.

3.4 Formal Semantics

In this work, as we have mentioned, for an immigrant agent's adaptation at the level of language and interaction to occur, the target society must provide the formal semantics of the two protocol languages described in the previous section. The language used for expressing the syntax and the semantics of such languages is what we refer to as *meta-language*; it can be any framework for giving formal semantics, as far as a computational interpreter exists for it. We presented, in (Bordini 1994), the formal semantics of the PDL and of the PEL (recall that this semantic specification would be part of the PIC, i.e., the area in a possible target society where the information about communication conventions is). Those specifications can be found in Appendix A, Section A.2. We have used both the denotational framework (the notation used was the one presented in (Watt 1991))—see Section A.2.1—and the structural approach to operational semantics (Plotkin 1981)—see Section A.2.2. The structural-operational version of the semantics is very elegant, while the denotational version is useful for the implementation of these ideas (discussed briefly below).

The semantics of a protocol described by means of the PDL is a set of state definitions (its name plus its transition relation) for each protocol. This is similar to the semantics of a variable-declaration section of an imperative programming language; in this case, it generates a mathematical structure representing a set of protocols that will guide the execution of transitions.

This is the internal structure of protocols **PS**, shown in Figure 3.2. Then the semantics of the PEL is, as expected, the changing of states and structures of interaction in the conversational structure, according to the protocols already interpreted using the semantics of the PDL.

We have also implemented a prototype of the CSA of a migrant agent using the Standard ML (SML) programming language (Paulson 1991) (for details of the implementation and its use, refer to (Bordini 1994)). Further, we have used ML-Yacc (Tarditi and Appel 1991) for the specification of the syntax and semantics of the languages that must be interpreted by the CSA. The semantics is expressed as SML functions in the semantic actions of the ML-Yacc specifications. This is the reason why we have used the notation in (Watt 1991) for giving the denotational semantics: it is very similar to SML notation. Because of this approach, we have been able to use our denotational semantics (in Appendix A, Section A.2.2) directly in the implementation. This is in accordance to the idea that the complete denotational semantics of a language can be viewed as an interpreter for it when used in a suitable functional programming language (Watt 1991; 1986).

That implementation can also be viewed as a testbed for different protocol languages, since the result of our CSA implementation reading the formal specifications is the transition interpreter and the protocol interpreter for those languages. The languages to be tested can than be seen to work in practice, which can be helpful in evaluating their expressiveness and suitability for the purposes of their designers. We do not provide the details of this implementation here; a complete account can be found in (Bordini 1994).

3.5 Concluding Remarks

Given that the material presented here is only auxiliary to our thesis, we have omitted the details, in particular on the implementation, which have been given in (Bordini 1994). One can also find there further discussion, e.g., on the feature of evolving complexity of interaction that is inherent to the approach.

We now briefly discuss the advantages of the use of the meta-language. We claim that with the proposal of the ideas in this chapter we introduce another specificity (or facet of autonomy) to agents, which is a very important one in comparison to real living societies: the communication languages used by individuals. Nevertheless, it is worth remarking that only agents able to migrate need to incorporate the meta-language interpreter. The approach does not require the meta-language to be universally interpretable, as for the PDL in (Populaire *et al.* 1993), or for the complete standard of communication proposed by the KSE (see Section 2.2.3). However, it is necessary that all specifications on the meta-level are done using a meta-language for which a computational interpreter is accessible to the migrant agents. We discuss this work further when we relate it to our ideas on interoperability of MAS, in Section 8.1.

After an immigrant agent is able to use the communication protocols of a target society, the adaptation process is far from complete. There still remains, for example, the learning of the actual language in which messages transmitted within protocol transitions are coded. In Chapter 5 we deal particularly with the ontological aspect of such languages. Also, aspects of communication are not all there is to cultural adaptation. Some societies may have important social norms which need to be learnt, and indeed some societies may not even have communication needs (e.g., in some game-theoretic approaches). Another important activity for a new agent is to become acquainted with other agents in the target society. The approach proposed in (Hübner 1995; da Rocha Costa, Hübner, and Bordini 1994; Hübner, da Rocha Costa, and Bordini 1995) includes both explicit presentation and inference of social roles by observation. It also suggests that both the adaptation at the level of *language and interaction* and at the level of *knowledge and performance* can be carried out in a uniform way using a functionalist approach devised in (da Rocha Costa, de Castilho, and Claudio 1995). A well-known approach to this problem is the work of Sichman (1998) (see further references in Section 2.2.2). The most straightforward solution to this problem would be the use of a presentation protocol, e.g. the one in (Berthet, Demazeau, and Olivier 1992). If such a protocol for agents introducing themselves is available in the target society, all the immigrant agent has to do is use the meta-protocol for protocol exchange that it already knows in order to ask for the presentation protocol from its potential new partners, and then use that protocol to become acquainted with them.

We note, *en passant*, that this part of our work is of interdisciplinary interest regarding its potential in building simulations concerning the process of acquisition of communicative skill (an important issue in Cognitive Science (Green and others 1996)). Furthermore, it is an initial but arguably useful contribution to a research area that is recognisably open, i.e., learning of communication languages in MAS (cf. (Durfee *et al.* 1997)).

As this discussion indicates, there are many problems involved in cultural adaptation apart from communication, which is undoubtedly a complicated problem in itself. The rest of the thesis seeks to deal with several aspects of cultural adaptation (including communication) from an anthropological point of view.

Part II

The Present

(*When Anthropologist Agents Come to the Aid of Migrant Agents*)

« ... Κλωθὼ δὲ τὰ ὄντα, ... »
Πλάτωνος, Πολιτεὶα

" ... *Clotho the things that are,* ... "
Plato, The Republic [617C]

(Shorey 1935)

Chapter 4

Cultural Anthropology in Multi-Agent Systems

This chapter focuses on raising theoretical questions from cultural anthropology, and from modern approaches in social sciences working on the limits of anthropology, sociology and psychology (in regard to cognition), in an attempt to establish grounds for an anthropological conception of MAS. The chapter also concentrates on a search for insights from field work practice in social anthropology and similar areas in making designs that are instrumental for a foreign agent to evolve cooperation with an unfamiliar community. It provides the general concepts and the theoretical basis that underly the whole of the thesis.

4.1 Introduction

We believe that an anthropological perspective can contribute to complex social simulations and have a significant role in MAS interoperability. We therefore suggest that this should be considered and investigated actively for the next generations of MAS: influencing the creation of DAI learning algorithms (surely, DAI settings provide territory for the creation of types of learning mechanisms not previously feasible, cf. (Durfee *et al.* 1997)), and producing new recommendations for designers of agent architectures and societies (if interoperability is an issue of theirs) as well as researchers concerned with the creation of software development methods that are agent-based (Iglesias, Garijo, and González 1999).

In Section 4.2 we extend the concept of migration of agents among different societies based on anthropological metaphors[1], and argue further in favour of the need for agents' thorough adaptation to a target society in order for them to establish fruitful interactions. That

---

[1] It is well agreed among MAS researchers that the use of anthropomorphic metaphors is one of the greatest advantages in future approaches to (agent-based) software development methodology (see, e.g., (Wooldridge 1997)). Our anthropological metaphors for the problem of migration have similar purposes.

same section also discusses the cognitive approach to cultural anthropology as the right theoretical/philosophical approach in anthropology on which to base the rationales for a proposal of using anthropological methods in DAI problems. We mention some of the anthropologists' field work techniques that can be useful in this computational counterpart situation in Section 4.3. The material discussed in Sections 4.2 and 4.3 has been presented in (Bordini and Campbell 1995). Section 4.4 then introduces the idea of legitimate peripheral participation (a specific approach to situated learning) and speculates on its relations to DAI in support of an anthropological conception; it is then followed by some concluding remarks are given in Section 4.5.

4.2 Cognitive Anthropology as a Foundational Discipline in Multi-Agent Systems

The "anthropological approach" to MAS is based mainly on two types of agent functionalities (two types of "foreign agents", in the view of a target society). First, there are the *Anthropologist Agents* (AA), which are agents that migrate to societies with the explicit intention of producing a cultural description of them. The outcome of the "field work" of such an agent in a society can be very valuable for easing the process of adaptation of other agents that migrate to that society because they have some actual needs or interests to participate in it. However, the cultural description generated by the anthropologist agent is not to be regarded as something that can be valid indefinitely, as cultural changes or evolution based on changing experiences are quite likely to happen. Second, there are the *Migrant Agents* (MA) themselves, i.e. agents that migrate to a society for the sake of accomplishing their own goals, or the social goals of the target society (as mentioned in Section 1.1). Such agents are the ones that accomplish interoperability of MAS in the scenario we have described (see further discussion in Section 8.1). Another important concept is that of artificial culture, discussed below.

It is important to emphasise what is new concerning the idea of "migration" of agents (further from what we mentioned in Section 1.1). There are several publications in DAI dealing with the problem of agents that do not know each other very well (Berthet, Demazeau, and Olivier 1992; Fenster, Kraus, and Rosenschein 1995; Durfee 1995; Sichman *et al.* 1994; Hübner 1995). Inasmuch as we consider the world-wide variety of societies of agents that may come to be created and the possible evolution of such systems, and as we consider complex societies, e.g. following the ideas in (Conte and Castelfranchi 1995), the problem here is that not only do agents not know each other well[2], but they also have limitations on the appreciation of the

---

[2] Agents that have different knowledge, plans, goals, etc., and simply are not aware of each other's particular

very structures used to achieve agreement among individuals, namely language—the subject of (Bordini 1994), briefly discussed in Chapter 3, with further contribution in Chapter 5—and *culture* (see below). In other words, we are considering here a sort of agent organisation that is far more complex than those focused on in (Ishida, Gasser, and Yokoo 1992); they are more akin to the idea of organisations (groups of socially committed agents) in the sense of (Castelfranchi 1995), and accordingly this sort of agent restructuring is of a higher degree of complexity.

The idea of *artificial cultures* (based on Cognitive Anthropology's view of culture discussed below, in Section 4.2.2) refers to the organising principles of behaviour, perception and interaction of agents in each particular society. As mentioned before, in MAS these particularities can be determined by the society's "creator" (designer) or as a result of that society's autonomous evolution. In terms of the current state of the art of MAS, describing the artificial culture of each society of agents would involve answering questions such as: "Do these agents behave like consumers and producers in an economic market?" (Wellman 1995), or "Are their interactions based on the notion of power and beliefs about social dependencies?" (Sichman *et al.* 1994; Castelfranchi 1995). Do they use protocols as proposed in (Campbell and d'Inverno 1990), do they use a language like the one in (Demazeau 1995), or do they use KSE's ACL (Genesereth and Ketchpel 1994)? Are these agents "believable" (Bates 1994) in such a sense that emotion is something to be taken into account?[3] How are material phenomena from the environment logically organised and referred to, and what are their important features from the point of view of this particular culture? (We return to this subject below, in Section 4.2.2.) Invention of a means for a foreign agent to be able to enter any society, find out answers to these questions, and then become thoroughly adapted (being able to interact fruitfully in that society), is where cultural anthropology has a stake in the foundations of DAI.

An important contribution to the social conception that underlies MAS was made by Gasser (Gasser 1991) in recognising the importance of current work in certain social sciences. In particular, the current conceptions from the social sciences mentioned in that papers help to give a reasonable foundation to MAS insofar as they concern the indissoluble social and situated character of cognitive phenomena. However, it seems that there is more to the relation of such works to MAS than this when we take into account the problem of migration among different artificial cultures. The ideas of portability of knowledge and the importance of language (both derived from conceiving learning as participation in communities of practice (Lave 1988;

---

details.

[3] We stress that this is not intended to be a thorough account of current works in MAS, but a (subjective) sample of major (significant) ones. Recall that a longer list of examples of different approaches to MAS was given in Section 2.3, supporting further our point on the potential cultural variety of MAS.

Lave and Wenger 1991), which we discuss in Section 4.4) are a plausible foundation for our work on adaptation of agents in a foreign artificial culture and our quest for justifying the placement of anthropology among the disciplines of interest to DAI.

4.2.1 The Cognitive Approach to Anthropology

If we are to use anthropology as a support for this new feature of MAS, and add it to DAI disciplines as a source of theoretical inspiration for solving the problems derived from considering that feature, a concern at this point should be: Which specific kind of work from the broad discipline of anthropology is: (i) likely to provide the most useful techniques for the conception and implementation of migrating agents, and (ii) which approach to anthropology has the right theoretical and philosophical basis to serve as a conceptual bridge for anthropology-inspired work in DAI? (Some of those techniques can be found in the next section.) For the latter, we propose a specific approach to cultural anthropology, namely *Cognitive Anthropology*, despite the recent critiques of that approach[4], mainly concerning *intracultural variation* (Lave 1988). This section discusses some points of cognitive anthropology, from the perspective of (Tyler 1969).

Cognitive Anthropology originated as a change in the theoretical orientation of building a general theory of culture. The aim was to use fairly formal methods to describe particular cultures, with each description being a theory for the culture it describes, "for it represents the conceptual model of organization used by its members" (Tyler 1969). Describing a culture means answering questions such as: "How would the people of some other culture expect me to behave if I were a member of their culture; and what are the rules of appropriate behaviour in their culture?" (Tyler 1969), which are basically the questions an immigrant agent faces. The aspect of generating formal descriptions is one point that makes this approach to cultural anthropology particularly relevant to the computational counterpart problem of understanding societies' particularities, even if social scientists can criticise its relative lack of "humanistic" factors.

More explicitly, these ideas from Cognitive Anthropology are important for the problem of migration of agents we have posed, as

> It [Cognitive Anthropology] focuses on *discovering* how different peoples organize and use their cultures. ... it is an attempt to understand the *organizing principles underlying* behaviour. It is assumed that each people has a unique system for perceiving and organizing material phenomena—things, events, behaviour, and emotions.

---

[4] Note that the issue of individuals' free will in not conforming to cultural uniformity is hardly relevant to non-human agents, and even less relevant for the purposes of cultural adaptation of agents on which we are interested here. Therefore, the critique of the cognitive approach to cultural anthropology implied by the nature of current trends in the social sciences should not prevent our using it as a basis for the work presented and proposed in this thesis.

(Tyler 1969)

That is, the interest is not about the material phenomena *per se*, but towards their relevance and cognitive organisation in the human mind, which are both culture-dependent. Classical cases of field work projects using this approach can be found, for example, in (Frake 1961) and (Conklin 1955).

4.2.2 Arrangements

An important change of approach that has taken place with the emergence of Cognitive Anthropology has led to seeking categories of description in the language of the natives rather than in the anthropologists' language. This is seen as a window into the mental processes of other peoples. Cognitive anthropologists study "how other people 'name' the 'things' in the environment and how these things are organized into larger groupings." (Tyler 1969). Their task is to realise how other people organise what seems chaotic, for everything in that environment is in principle unknown to the anthropologists. What is significant in the environment and how perception is organised are studied through the way natives label things: "Naming is seen as one of the chief methods for imposing order on perceptions." (Tyler 1969).

A *culture* is regarded as consisting of many *semantic domains* (i.e., subjects) organised through *semantic features* (also called *components* or *features* of *meaning*), which are the dimensions considered important or distinctive by natives. Each of these semantic domains may be ordered by one or more *arrangements* such as:

Taxonomies, used for semantic domains where perceptual phenomena are assigned to named classes that are organised into larger groupings. Items at the same level in a taxonomy contrast with one another, and items at different levels are related by inclusion. (The cognitive approach to discovering taxonomies, and further details on this specific type of arrangement, are presented in Section 5.2.3; because that material is part of the groundwork for Chapter 5, it is important that it is placed there.)

Paradigms, which have two major features that intersect one another (presented in the form of a table).

Trees, which are ordered by sequential contrast on only one feature at a time; they are based on successive choices among two or more alternatives.

Tyler (Tyler 1969) gives further details. Different cultures never share all the same semantic domains and features of meaning, and they do not organise all the features in the same way:

"The problem for the anthropologist is to discover the semantic domains and their features... He must attempt to discover the semantic world in which these people live." (Tyler 1969).

Identifying *levels of contrast* in taxonomies used in particular cultures is a very important point in the cognitive approach to anthropological field work (as in (Frake 1961) and (Conklin 1955), for example). The cognitive approach is clear in those papers (they were pioneer works in using that approach): "Folk taxonomies are cultural phenomena. Their structural variation within and between cultures must be explained by cultural uses to which a taxonomy is put, and not by appeal to differences in the cognitive powers of individual minds." (Frake 1961). In our multi-agent world, we can say that agents with similar architectures which happened to evolve in different cultures may have, say, different organisations of their perceptions, and different languages, among other disparities.

The structures (arrangements) listed above are rather familiar in computer science: as in anthropology, specialists in knowledge-based computing systems are concerned with elucidating how an intelligence perceives the world, so the correspondence of their tools of description is no surprise. It seems that this correspondence (between computer science and the anthropological work on ethnography of language) is stronger than it may at first appear; consider, e.g., the work on ontologies in the KSE (Gruber 1993b). One part of this thesis involves following up the idea that an immigrant agent must be able to *discover* the appropriate arrangements for the semantic domains in its target society, given the dynamics of MAS seen as autonomous, cognitive and evolutionary[5] systems; and discovering an ontology is a first step before the arrangements can be worked out. This is presented in Chapter 5, where we also expand the views on taxonomies from the cognitive approach to anthropology. (The complementary material on the cognitive approach is presented in that chapter (Section 5.2.3), near our work inspired by it (in particular Section 5.3.3), that being the most natural way of organising that material in the thesis.)

4.3 Research Methods in Anthropology and Design Methods for Multi-Agent Systems

Apart from choosing suitable philosophical concepts of anthropology, we also appreciate that there is much to gain from the practical experience of anthropologists in field work (Bernard 1994). This comprehensive textbook in anthropological methods confirms that some of the methods that can clearly be used for easing an immigrant agent's adaptation are standard and validated in anthropologists' experience in field work practice with foreign tribes. We describe

---

[5] On the question of the (false) controversy between emergence and cognition in agents, see the very interesting thesis put forward by Castelfranchi (1997).

briefly the anthropological research methods (from the ones we have studied) that we have found to bear the greatest relation to research problems in MAS in the subsections that follow.

4.3.1 Participant Observation

An important concept in cultural anthropology is *Participant Observation*, which is a special feature of ethnographic field work. It is a means for an anthropologist to be accepted in a community so that s/he can observe and record information about the culture of that people. Only if an anthropologist is successful in establishing rapport in a new community will people not change their behaviour because of his/her presence. The basic idea is that the anthropologist should participate in the normal activities of the community being studied (Bernard 1994).

The counterpart of this proposition for autonomous agents is that an agent in process of adaptation to a target society should, as much as possible, take part in the general "problem solving" that is occurring in that society. This should encourage agents to act as *informant agents*, which is crucial for a good analysis of the target society. The need to contribute to the general work of a society in order to get the appropriate responses to the queries a foreign agent asks is specially important in MAS, if and when they are composed of the kind of complex agents that establish artificial cultures. Being autonomous, self-interested[6], and subject to complex social relations, such agents are not likely to contribute freely to the foreign agent's adaptation unless they can have some sort of reward for their help (given those characteristics of such agents, a reward that the overall society is improved, but which is not an individual one, may not be sufficient). Nevertheless, some agents may exist that are *benevolent* (Conte and Castelfranchi 1995; Castelfranchi 1990) enough to contribute to the anthropological work for pure "pleasure" (as happens sometimes in real life), so that it may reward a foreign agent to spend some of its efforts in looking explicitly for them. Therefore, we think MAS designers should consider including "informant" tendencies in at least some of the agents, if they intend to facilitate interoperation. This is also an important aspect for a recent thread of research concerned with devising agent-based software development methods (Iglesias, Garijo, and González 1999).

Note that, although we are not interested here in issues of security, an anthropologist agent's work may be undermined if local agents are not persuaded that providing it with information about the local culture is "safe" for their own business, whatever the goals of that society. As a mere illustration, we remark that some of the rules followed by anthropologists when entering the field for participant observation are: go to the field with written documentation about

---

[6] But note that we question the widespread notion that autonomous agents are necessarily selfish agents, in Section 8.3.

him/herself and the project, or even about cultural anthropology itself; be prepared to answer questions about his/her intentions, about the person or group for whom s/he is doing the work, duration of the visit, why s/he intends to learn about that society, and so forth. Other things anthropologists consider when preparing for field work, and which can be helpful here, are: in case of multiple sites, choose the one that seems best for easy access to data; create maps of the region or of the "social scene" in case there is no physical location to be mapped (where the latter means finding out the names of the key players and their relationships).

In general it takes as much as one year for an anthropologist in participant observation to study a community (Bernard 1994). It would be interesting to evaluate, should anthropological approaches to adaptation come to be used in the promising wide range of practical MAS that may be implemented in the future, the time needed for an agent to study a whole computational society through participant observation, and hence to relate time-scales for the human and computational processes.

4.3.2 Informants

Ethnography, a routine activity of anthropologists, depends at base on a few *key informants.* It is therefore indispensable to choose competent informants (i.e., people with whom one can talk easily, who understand the information needed, and are willing to give, or try to get, such information), and ask good questions (i.e., those that informants can answer). Participant observation is also important in guiding the anthropologist in asking sensible questions regarding that culture (Bernard 1994). We name the computational equivalent of an anthropologist's informant an *Informant Agent* (IA), i.e., an agent that provides the anthropologist agent with information about its specific artificial culture.

If an anthropologist agent has "informant agents" who help it create anthropological descriptions of a society, the immigrant agents, in their turn, may have help from local agents as well. We mention some ideas for this extra support for the adaptation of an immigrant agent:

- An immigrant agent may find *compatriot agents* in the target society, i.e., agents that come from the same society of origin and have migrated previously; or, alternatively, agents that have migrated from some third society, but have also had some experience of what it is to be an outsider in that target society[7]. Those agents (if there are any of them) are likely to help the new agent in its process of adaptation, as they may have experienced similar difficulties.

---

[7] Lave and Wenger say that apprentices learn substantially from other apprentices; see Section 4.4.

- Looking for *sympathetic* (or *altruistic*) *agents* is a good tactic for the immigrant agent (i.e., looking for agents that show themselves interested in helping). In our point of view, this should be one more feature in a general scheme of searching for suitable informants, as confirmed in the anthropological counterpart.

- Some societies may have what we call a *novice master*[8] *agent*, which can be handy in the adaptation period. To the extent that features of a society can be engineered, we suggest that designers should incorporate the "novice master" characteristic in the MAS that they specify.

  Some of the properties that we included in the anthropologist and informant agents used in Chapters 5 and 6 are related to those of a "novice master" agent. Some societies, however, may (need to) provide a specific novice master (i.e., a specialised agent).

4.3.3 Language

Bernard (1994) points out that knowledge of the local language improves the rapport and allows a more reliable check on the accuracy of elicited data. In their notion of learning as performing (i.e., participating) in a community, Lave and Wenger (1991) emphasize that language (also seen as a way of acting in the world) is of decisive importance "..., since language use entails multiple participatory skills, and is one of the most basic modes of access to interaction in social life." This leads to the idea that an immigrant agent must acquire at least some acceptable level of knowledge of the local language (the initial motivation for the work briefly outlined in Chapter 3, and in other aspects for the work in Chapter 5).

In certain circumstances, it is also possible to deal with a very simplified language which could stand for some sort of gesture language that is universally understandable. In the cricket exercise (in Chapter 6; see Section 6.3.3) one can consider gestures made by players to celebrate a positive achievement or a failure as an example. This also suggests that a somewhat neglected classical AI work can be very useful in the future of MAS interoperability: Schank's Conceptual Dependency Theory ((1975); Schank and Abelson (1977)).

4.3.4 Field Notes

Field notes are what differentiate field work from mere experience (Bernard 1994). Thus, dealing with *field notes* in an appropriate manner is also essential to the task of a social anthropologist. Those that are of the most important kind in the context of the present work are the *descriptive*

---

[8] This term is used in monasteries to designate the monk who is in charge of helping new monks to become adapted to a monastery and its routines.

notes or *ethnographic notes*, which are the basic means for capturing details of behaviour and the environment. There is a suggested format for coding them (Bernard 1994), which is:

$$\#\langle \textit{fieldnote\_number}\rangle\ \langle \textit{date}\rangle\ \langle \textit{place}\rangle\ \langle \textit{name\_of\_informant}\rangle\ \langle \textit{topical\_codes}\rangle\ \langle \textit{content}\rangle$$

where $\langle \textit{place}\rangle$ and $\langle \textit{name\_of\_informant}\rangle$ may be encrypted for security reasons and $\langle \textit{topical\_codes}\rangle$ are (either personal or standard) codes for the topics covered in the $\langle \textit{content}\rangle$ part of the note. For MAS, the elaboration of a good scheme (which, unlike the case in anthropology, will not be natural language) for expressing $\langle \textit{content}\rangle$ is a promising focus for future research, and a good test of the value of KIF (Genesereth and Ketchpel 1994) and similar proposals.

It would be interesting to have, in the future, a computational equivalent of the Human Relations Area Files[9] (HRAF) (Bernard 1994), which would record public information of all known (already studied by anthropologist agents) cultural information from all the existing societies of agents accessible in international computer networks.

4.3.5 Further Techniques

Several other methods in (Bernard 1994) are relevant to the conception of anthropologist agents. We mention just a few others below—refer to that book for the full concepts. Anthropologist agents could have in their knowledge bases information about techniques used by anthropologists for choosing informants, e.g. the *consensus model*. They should be "aware" of the *response effect* and *deference effect*; know how to run *focus groups*; and have "*probing techniques*" (for interview with complex (truly autonomous) informant agents).

Further, the distinction between *direct, reactive observation* and *unobtrusive observation* also applies here. The technique of *continuous monitoring* is a helpful tool for the first type. The second type, where a subculture is being studied without its members being aware of that fact, means that the anthropologist needs to be a member of the general culture, so as not to be perceived as a foreigner (in the MAS case, the anthropologist agent would have to be a local agent as well). Also, the statistical techniques presented by Bernard, which are used in anthropology and social sciences in general, e.g. *sample taking*, can be useful for an anthropologist agent.

In summary, (Bernard 1994) provides a vast spectrum of anthropological methods which are relevant for the computational counterpart activity: that of an anthropologist agent studying target societies. It is not our intention to explain a larger range of them here. The sample shown is sufficient, though, to demonstrate our point on the usefulness of anthropological field work techniques for the problem we consider in this thesis. It makes a strong suggestion that MAS

---

[9] The HRAF comprises primary ethnographic materials (copies of pages of articles and books) from 350 cultures coded with the Outline of Cultural Material (OCM); see (Bernard 1994).

designers and particularly researchers who will deal with agent-based software development methods (the non-existence of which many, including (Wooldridge 1997) and (Iglesias, Garijo, and González 1999), claim to be a major hindrance preventing the wide spread of agent technology) should study in detail anthropological methods among all the other humanistic disciplines of interest to MAS.

4.4 Situated Learning in Anthropological Multi-Agent Systems

Before we turn to Lave and Wenger's elaboration (1991) on the concept of *situated learning*, we discuss a previous work by Lave (1988) which provides a basis for those more recent ideas. Lave suggests, which is of strong significance for the thesis we are constructing here (the necessary inclusion of anthropology as a discipline in DAI), that there is an indivisible character in the study of psychology (of cognition in particular) and anthropology. She argues that both have made wrong assumptions in order to allow their studies to be carried out "independently." Cognitive anthropology has made the well criticised assumption of cultural uniformity (i.e., neglecting the subjective nature of sociocultural phenomena, or ignoring intracultural variation), while cognitivists try to isolate the individual from the settings of their everyday[10] practice, believing that only by experimentation inside laboratories can they understand subjective processes (which is well argued to be a mistake, in that book).

Further, she is highly critical of functionalist propositions that abstract general knowledge verbally transmitted is the source of cognitive skills available for transfer across situations (time and settings). She calls for a conception of *knowledge-in-practice*, constituted in the settings of "everyday" social practice, as the most powerful source of knowledge for people in the "lived-in" world (Lave 1988).

In accordance with the emphasis on "practice," (Lave 1988) proposes cognition as a societal phenomenon, yet not removing totally its subjective character. She criticises common conceptions of social practice for their individualistic and rationalistic biases, i.e., their tendency to emphasise utilitarian interest as the motivation for human action. (It is worth mentioning that this is related to our work on moral sentiments, presented in Chapter 7; interestingly, we only encountered this particular reference in the anthropological material that was the basis for this thesis after our work on moral sentiments was started.) She insists that present theories of cognition, with their emphasis on the rational aspects of cognition (i.e., to see cognition as

---

[10] "Everyday" is a term used in that book to refer to what people do in ordinary cycles of activity. This is a gross simplification of this important notion for that author, see (Lave 1988).

general *problem solving* and rationally motivated) are culturally and historically determined (i.e., a manifestation of contemporary western culture). In her own words:

> If the analytic concept of the individual is reduced to a self-contained, disembodied technology of cognition, knowledge is reduced to scientific "discoveries," and society to a set of actors whose lives are structured only by self-interested motives, then both the analyses and conclusions that follow must surely involve deep impoverishment and distortion of their object. ... a more appropriate unit of analysis is the whole person in action, acting with the settings of that activity. This shifts the boundaries of activity well outside the skull and beyond the hypothetical economic actor, to persons engaged with the world for a variety of "reasons;" it also requires a different version of the everyday world.
>
> It is within this framework that the idea of cognition as stretched across mind, body, activity and setting begins to make sense. But we have arrived at the limits of the sociological theories of practice, for they do not specify a novel theory of cognition itself. Instead, "cognition" seems to represent one limit of the field of their inquiry. ...

Notice that this *decentralised* view of cognition is also already a major rationale for the work on DAI (Gasser 1991). Also note that she maintains that the whole process is beyond sociology; this argument supports our point about the need also for an *anthropological stance* in MAS.

Her analysis of everyday activity leads to new conceptions of knowledge, situation, and problem solving. Knowledge takes on the character of a process of knowing, the active engagement of consciousness in a reciprocal relation with the world (simultaneously knowing and changing the world). For her, the notion of situation cannot be separated from activity, therefore it must be transformed into a concept of dialectically constituted, *situated activity*. This dialectical constitution of the relations among activity, settings and processes of dilemma resolution implies that it is not possible to separate the means of problem-solving activity from its ends. She goes on to say:

> Further, if goals are not exogenous to the constitution of problems, then a problem is not structured as an end in itself or by a goal set elsewhere and presented to problem solvers by problem givers. A problem is a dilemma with which the problem solver is emotionally engaged; conflict is the source of dilemmas. Processes for resolving dilemmas are correspondingly deprived of their assumed universalistic, normative, decontextualized nature. As studies of math in practice have demonstrated, problems generated in conflict can be resolved or abandoned as well as solved, and often have no unique or stable resolution. Since quantitative relations, embodying value directly, bear direct relations with aspects of dilemmas that aren't quantitative, most dilemmas which involve relations among quantities are not well-formed arithmetic problems. In short, both theoretical critique and empirical evidence recommend that we recognize the cultural character and historical continuity of the contemporary study of cognition, and act accordingly to broaden the search for alternative conceptualizations that might encompass a richer, less stylized, investigation of the world as is.

> Reordering relations between means and ends leaves in question that status and meaning of "rational action." ...

Also of interest to this thesis, inasmuch as we argue towards anthropological MAS, is the point made by Lave on the *cultural* character of activity and, therefore (in her view), of cognition as well. Again we quote her own words (1988): "By arguing that activity, including cognition, is socially organized and quintessentially social in its very existence, its formation, and its ongoing character, we have committed the enterprise to theorizing about the social production of action as well as its cultural character." She also argues against transmission and internalization as the primary mechanisms by which culture and individual come together. This is a very complex problem in the social sciences, which is related to problems in MAS as we see them in our anthropological approach. It also resembles the micro-macro link problem thoroughly discussed (also in the realm of DAI) by Conte and Castelfranchi (1995; we refer further to this work in Section 8.3).

Lave also argues, at a theoretical level, against the dualism of mind and body. After she wrote that book, there have been major advances in the understanding of this point through neurological studies, in particular in the well-known book "Descartes's Error: Emotion, Reason and the Human Brain" by Damasio (1995), which, interestingly, is one source of evidence for Ridley's points on the innateness of moral sentiments that we present in Section 7.2. Lave's views on that issue of dualism at the time (1988) already went like this:

> The claim that the person is socially constituted conflicts with the conventional view in its most fundamental form, with the venerable division of mind from body. For to view the mind as easily and appropriately excised from its social milieu for purposes of study denies the fundamental priority of relatedness among person and setting and activity. The strategy adopted here is to replace dichotomous divisions, especially between the mind and the body, with ones that cross-cut them and reflect what appear to be more fundamental categories of everyday experience. ...

On the same grounds, that learning is an integral part of generative social practice in the lived-in world, Lave and Wenger (1991) go on to create the concept of *Legitimate Peripheral Participation* (LPP). The reason for the first word in that expression[11] is that "The form that the legitimacy of participation takes is a defining characteristic of ways of belonging [to a community of practice[12]], and is therefore not only a crucial condition for learning, but a constitutive element of its content." The second word, in its turn, refer to the idea "that there are multiple, varied, more- or less-engaged and inclusive ways of being located in the fields of participation defined

---

[11] Note that that expression is not intended to be considered in parts but as a concept to be regarded as a whole. We are simplifying matters so as to make it straightforward the presentation of that composite expression (see (Lave and Wenger 1991) for an extensive discussion of the term).

[12] The concept of *community of practice* is also important in that work. A detailed discussion on it can be found there.

by a community." And, of course, the term *participation* emphasises the fact that it is all about actions located in the social world. Further, they say that "*Changing* locations and perspectives are a part of actor's learning trajectories, developing identities and forms of membership." (Lave and Wenger 1991; their italics). In brief, *legitimate peripheral participation* is Lave and Wenger's concept related to situated learning, which emphasises that *learning is the accomplishment of an individual who engages progressively in the social practices that are characteristic of a community.*

The idea of learning in practice immediately brings to mind the notion of apprenticeship. One should be careful, though, with the analogy. Lave and Wenger are highly critical of the usual (mediaeval) view of apprenticeship, related to master-apprentice relations: this is not a general model of apprenticeship in their view. They also point out that in the relation of masters to apprentices, the issue of *conferring legitimacy* is more important than providing teaching, as usually thought.

One of their claims which is quite significant for this thesis (in regard to the theme of cultural variety of MAS) is that of the cultural specificity of learning, which was made clear by the theory of LPP: "In short, the form in which such legitimate access is secured for apprentices depends on the characteristics of the division of labor in the social milieu in which the community of practice is located." This entails that migrant agents must be prepared to adjust their *learning* activities to the organisation of each particular target society, and the society itself be prepared to guide that learning process. Also, the general form of learning as LPP may serve as a basis for the continuity of activity of a migrant agent across societies (even though an explicit account of it is not presented, so mapping to MAS is not a trivial task).

Also, they noticed in their case-studies of apprenticeship (reported in (Lave and Wenger 1991)) that senior members of a community of practice often avoid giving information specific to advanced stages of a profession to new apprentices (which raises the possibility of local agents withholding information justifiably from foreign agents), until they are "ready" for a next step through increasing participation in the community.

They also point out a typical behaviour in apprenticeship which is relevant here too: apprentices learn mostly in relation to other apprentices (recall our earlier suggestion of migrant agents searching for other migrant agents). Where the circulation of knowledge among peers is possible, it spreads exceedingly rapidly and effectively. This suggests that engaging in practice, rather than being its object (as in conventional school education), is likely to be a *condition* for the effectiveness of learning.

Further, they claim that "legitimate peripherality" provides "newcomers" (apprentices) with more than just opportunities for observation, but most importantly it involves *participation* as a way of learning, of absorbing and being absorbed in the "*culture of practice*," which connects directly with the motives in this thesis. "It includes an increasing understanding of how, when, and about what old-timers collaborate, collude, and collide, and what they enjoy, dislike, respect, and admire. In particular, it offers exemplars (which are grounds and motivation for learning activity), including masters, finished products, and more advanced apprentices in the process of becoming full practitioners." (Lave and Wenger 1991)

Again in opposition to traditional schooling, they distinguish a *learning curriculum* from a *teaching curriculum*: the former consists of situated opportunities for engagement in improvised practice (i.e., it is not specified as a set of dictates for proper practice), whereas the latter is constructed for the instruction of newcomers. Significantly, they claim that as a consequence, a learning curriculum is *characteristic of a community*. These issues deserve attention in future conceptions of DAI mechanisms for learning.

They briefly consider the implications of their theory for pedagogical practices. The result is an interesting critique of the very existence of schools (for these remove learning from the community of practice which produces what is to be learnt). In their view, test-taking is a parasitic practice, the goal of which is to increase the exchange value of learning independently of its use value (Lave (1988) also makes a point on the lack of transferability of "abstract" knowledge learned at school). This particular aspect of Lave and Wenger's theory is not particularly related to this thesis, so we do not comment further on it here.

They claim that productive activity and understanding are not separate (or even separable), but dialectically related. This is an argument against the dichotomy between learning experimentally and learning by abstraction. Rather, they claim that these differences reside within the cultural practices[13] in which learning takes place, on issues of *access* and *transparency* (which we mention next) of the cultural environment with respect to the meaning of what is being learnt.

One important issue of legitimate peripherality is related to "*access*" by newcomers to the community of practice and all that membership in that community entails. This is delimited by relations of power and conflicting interests (including economic interests). Therefore, the issues of social power in DAI (Castelfranchi 1990) ought to be considered in a possible DAI conception of the ideas of LPP. Lave and Wenger show, in one of the case studies they present, an example of a community where newcomers are not given access by their senior counterparts to the practice that would lead them to progress in understanding/participating fully in the community. Another important issue mentioned earlier is that of "*transparency*": this refers to the idea that the

---

[13] For example, they argue that participation in school practices is neither more nor less abstract or concrete, experimental or cerebral, than in any other.

knowledge and typical ways of perceiving and manipulating technologies/artifacts used within communities are encoded in them in ways that can be more or less revealing to the newcomers in their process of trying to understand the use and significance of that technology/artifact.

They also raise the interesting question of discourse in practice, opposing the traditional emphasis given to verbal instruction. They suggest that the purpose for newcomers is not to learn *from* talk as a substitute for LPP; it is to learn *to* talk (in ways that are appropriate in the practice of that community) as a key to LPP. The same arguable applies to MAS (at least in the usual situation where interoperation is desirable).

Their conception about motivation for apprentices relates strongly to issues of identity; they say: "Moving toward full participation in practice involves not just a greater commitment of time, intensified effort, more and broader responsibilities within the community, and more difficult and risky tasks, but, more significantly, an increasing sense of identity as a master practitioner." (Lave and Wenger 1991). This gives clues as to some characteristics that designers should consider for the motivation of their migrant agents.

One of the contradictions in the core of learning processes, made clear by the LPP approach, is between continuity and displacement. LPP is both a means of achieving continuity over generations for the community of practice, and an inherent process of displacement, as "full participants" are thereby replaced by "newcomers-become-old-timers." This contradiction is fundamental to the social relations of production and to the social reproduction of labour. It is ever-present in apprenticeship:

> The different ways in which old-timers and newcomers establish and maintain identities conflict and generate competing viewpoints on the practice and its development. Newcomers are caught in a dilemma. On the one hand, they need to engage in the existing practice, which has developed over time: to understand it, to participate in it, and to become full members of the community in which it exists. On the other hand, they have a stake in its development as they begin to establish their own identity in its future. (Lave and Wenger 1991)

Again relations of power are influential here, in regard to the way this contradiction is played out. It is in participation that the conflicts unfold between the old and the new: "Each threatens the fulfillment of the other's destiny, just as they are essential to it."

Lave and Wenger point out (as does Frake (1969), see our discussion in Section 5.2.3) that *naïveté* is beneficial in newcomers' behaviour[14] in the appropriate settings, and further: "Legitimacy of participation is crucial both for this naive involvement to invite reflection on ongoing activity and for the newcomer's occasional contributions to be taken into account." An interest-

---

[14] For this reason, a naïve behaviour that is typical of "newcomers" needs to be modelled in migrant agents, or at least it is usefull to consider this point in regards to believability (we mention this further later on in the thesis).

ing comment they make is that all of us are "legitimately peripheral" in some respect: "everyone can to some degree be considered a 'newcomer' to the future of a changing community."

In summary, they seek relational, historical conceptions in this particular enterprise of transforming *situated learning* in *legitimate peripheral participation in communities of practice*, considering that it "takes place in a *social world*, dialectically constituted in social practices that are in the process of reproduction, transformation and change," having to address the structural character of that social world at the "level at which it is lived." They generate analytical terms and questions fundamental to this analysis, including: "... forms of membership and construction of identities ... the location and organization of mastery in communities; problems of power, access and transparency; developmental cycles of communities of practice; change as part of what it means to be a community of practice; and its basis in the contradictions between continuity and displacement."

We believe that these are good theoretical grounds for the development of new concepts of learning in a DAI context, which Durfee *et al.* (1997) argues to be an important open research question of DAI. However, we have left any general attempt to build such learning mechanism out of the scope of this thesis. We include the discussion of LPP here as it is an important conceptual framework within which to start such research, which is in tune with some of the themes discussed in this thesis (as learning is closely related to adaptation). What is more, it is in line with our anthropological approach to MAS for it recognises the role of anthropology, alongside psychology and sociology, for a proper understanding of cognition; it is, therefore, one more support for our claims regarding the suitability of the use of cultural anthropology to solve problems in DAI.[15]

4.5 Conclusion

We have presented some concepts from Cognitive Anthropology and techniques used by field-workers in Social Anthropology which contribute, in several levels of abstraction, to the conception of agents that can migrate among different societies. We have, therefore, recommended that Anthropology should figure among the DAI disciplines when one considers the problem of interoperability among potentially very different multi-agent systems.

The material in this section serves as a theoretical connection between the different pieces

---

[15] By mentioning the role of anthropology in understanding cognition, we make it clear the importance of this approach to MAS as far as cognitive agents are concerned; but we also maintain that other types of agents do have a culture too (which can be analysed by cognitive migrant agents, as a means towards interoperability). We further point out that an interesting difference between cognitive and other types of agents is that the former may, in principle, be capable of creating and changing their own cultures!

of work (or "contributions," as we refer to them at times) we have done on aspects of the problem of interoperation—they all relate to these anthropological conceptions in some way. The cognitive approach to anthropology has inspired our work on ascription of intensional ontologies and recovery of taxonomical relations presented in Chapter 5. Part of the anthropological material has not been presented in this chapter, but delayed until Section 5.2.3, for that reason. The ideas of LPP are referred to in Chapter 6 as they can be used in future stages of our simulation of the game of cricket, aiming at the generation of novel (DAI-based) learning techniques for which the game of cricket is a suitable testbed. Anthropological findings are also the motivation for our work on advocating the use of emotions in agent architectures, which is the subject in Chapter 7.

Chapter 5

Anthropological Descriptions of MAS: Ascription of Intensional Ontologies of Terms and Retrieval of Taxonomical Relations

The chapter presents an approach to the description of ontologies used in Multi-Agent Systems as a means to allow interoperability of such systems. It is inspired by a pragmatic theory of intensionality worked out in our anthropological approach to agent migration. A formalisation of how an intensional ontology can be ascribed to a society of agents is presented, together with the formalisation of the recovery of taxonomical relations from such ontologies. This process of discovering taxonomies is inspired by ethnographic studies in social anthropology; the anthropological material presented here is complementary to that presented in Chapter 4. The formalisations are developed using a framework for agent theories, based on the Z specification language. Further, the approach is illustrated by the ascription of an ontology and associated taxonomies for an exotic application: the game of cricket. Several issues related to this approach are discussed in Section 8.2.

5.1 Introduction

We have shown in (Bordini, Campbell, and Vieira 1997) how an agent can ascribe ontological descriptions for the terms used in the communication language to a society being observed; this idea was first mentioned in (Bordini, Campbell, and Vieira 1996). For this particular problem, we have proposed the use of a pragmatic theory of intensionality, which is based on the work of Martin (1959), and has been revived and adapted to the MAS context by Vieira and da Rocha Costa (1993). Further, we have presented our approach using Luck and d'Inverno's (1995) formal framework for specification of agent theories. In terms of our concern with interoperability, these ideas are particularly relevant for interoperation of systems that do not employ similar definitions for the terms used to communicate and represent knowledge. Our approach can be seen as a step towards a solution to the problem of ontological mismatch among disparate MAS.

In (Bordini, Campbell, and Vieira 1998), we extended our work on ascription of intensional ontologies to show how an agent can work out the taxonomical relations existing among the terms in the intensional ontology it has ascribed to a society of agents. We have noted that some initial taxonomical relations can be recovered directly from an ascribed intensional ontology. We therefore extended our previous formalisation to include such a process; it too was inspired by the methods used by cultural anthropologists (also from the cognitive approach to anthropology presented in Section 4.2), as we shall discuss later. A taxonomy is clearly important from an agent's-reasoning point of view; this has been a recurrent observation in Artificial Intelligence research since the early days. Furthermore, from experience in anthropology, it is known that a taxonomy can be quite revealing about the traits of a particular culture (as mentioned in Section 4.2). This latest extension of our work is, thus, related to a fundamental aspect of the procedures of an anthropologist studying a particular society. We suggest that the same approach is of value for an "anthropologist agent" studying a MAS.

The next section summarises the main concepts and definitions we have used or created for this work; these are needed for an understanding of our ideas and the formalisations presented in this chapter. In particular, Section 5.2.3 discusses the elaboration of taxonomies in the context of social anthropology, which has been the main influence on the latest extension of this work (related to taxonomies). We then present the formal specifications in Section 5.3, while Section 5.4 shows a case study of those specifications in the context of a game of cricket. Some discussions will be presented in Chapter 8 (Section 8.2). Appendix B has some auxiliary Z-related material.

5.2 Background

This section reviews some important concepts and definitions. They are essential for an understanding of the specifications we shall present in Section 5.3. This section is organised according to whether the definitions are from the underlying pragmatical theory of intensionality (Vieira and da Rocha Costa 1993), whether they are specific to our work on ascription of intensional ontologies, or whether they concern the elaboration of taxonomies by cultural anthropologists.

5.2.1 Subjective Intensionality

This section covers only the main concepts related to subjective intensionality which we shall use next. We have given a larger account of these concepts and some discussion of its advantages in (Bordini, Campbell, and Vieira 1997); for further details it is necessary to refer to (Vieira and da Rocha Costa 1993), or even to their source (Martin 1959). These concepts will be made clearer when we use them in the formalisation presented in Section 5.3.

The *intension* of an expression is what is known about it in order to identify the object/entity to which it refers. We can say that intension is related to notions of mental entities, properties, relations, and concepts, while *extension* is related to objective entities (i.e., objects, structures). Further, we have the concept of *subjective intensions*. These are associated with the intuitive notion of connotation of a term or name; that is, related to the properties that are associated with a term in an individual's mind in such a way that they are normally borne in mind when the individual uses that term at a certain time[1]. Further, *quasi-intensions* are linguistic reductions of the mental entities relative to intensions. Therefore, the terminology *subjective quasi-intensions* emphasises that the theory deals with virtual classes of expressions related to particular users of the language; in other words, it is a linguistic reduction of the cognitive notion of connotation.

In order to define subjective quasi-intensions and related notions, the acceptance relation between agents and expressions is introduced. The definition, originally described in (Martin 1959), follows.

Definition 5.1 (Acceptance Relation). Acceptance is an empirical relation between users and sentences of a language, observed by an experimenter at a certain time who asks questions by means of a set of sentences forming a logical theory. Whenever an agent answers affirmatively to (has a positive attitude towards) one of these sentences we say that the agent *accepts* that sentence (which must belong to the set of sentences given by the experimenter) at that time.

Definition 5.2 (Subjective Quasi-Intension). The notion of *subjective quasi-intension* for an individual constant (term) is defined as the properties a language user associates with the term, as expressed in the sentences that the given user *accepts* at a certain time.

Definition 5.3 (Intersubjective Quasi-Intension). This concept regards groups of language users, rather than individuals, at a certain time. An *intersubjective quasi-intension* is the equivalence class of all the *subjective quasi-intensions* of a certain group of users of the language.

---

[1] Since these are notions intrinsic to the users of the language, they can also be called *pragmatical intensions*.

Intertemporal Quasi-Intensions are relative to a particular language user at all times. *Objective Quasi-Intensions* can also be defined on the basis of acceptance. They are at the same time the intertemporal and intersubjective quasi-intension of expressions, that is, a class whose members are members of the subjective quasi-intensions of all language users at all times. They are said to be an essential property, as they are universally accepted (within a specific community). One last type of quasi-intensions is that of *Societal Quasi-Intension* which relates to a particular group of agents[2]. In Martin's theory, *Co-Intensiveness* is defined as a relation between terms that have the same subjective quasi-intension (or indeed for any of the types of quasi-intensions mentioned above).

In this theory, a proper "understanding" of a concept can be defined as the situation in which the subjective intension of a term relative to an agent is the same as the intersubjective intension of all agents, some expert group or a specialist.

5.2.2 The Process of Ascription of Intensional Ontologies of Terms

5.2.2.1 Intensional Ontologies of Terms

We here take *ontology* to mean very much the same as proposed by Gruber (1993a), i.e. the definition of a set of representational terms[3] (stated as a logical theory). However, it is important to bear in mind that the theory of intensionality presented here deals only with individual terms. The major contribution of this approach to description of ontologies is that its underlying theory allows us to work towards providing agents with mechanisms for dealing with ontologies themselves (i.e., ascribe possible ontological representations to societies in case they are not available; effect changes in ontologies without consequent interoperability problems, etc.).

The following definition expresses our conception of ontology:

Definition 5.4 (Intensional Ontology of Terms). An *Intensional Ontology of Terms (IOT)* is a set of terms where each one is associated with the (minimal) set of predicates (properties) that is necessary and sufficient to distinguish (unequivocally) itself from every other term in the universe of discourse of a communicating society of agents.

In this approach, the definition of a term is a set of predicates that are considered to hold for that term. It is important to appreciate that not all predicates that hold for the term are needed for its ontological description: there is a difference between knowledge representation and commitment to ontological conventions (Gruber 1993a). Therefore, if some notion of order

---

[2] We do not formalise the idea of groups of agents (and therefore the concept of Societal quasi-intension cannot be formalised either) in this thesis, but the extension should not impose any difficulties.

[3] In this particular context, the representational terms are those used in the communication language of a MAS.

for the predicates is available (e.g. a hypernymy relation[4]), this can reflect on the minimal set of predicates: it would include only the most generic ones which are enough to distinguish the term unequivocally.

5.2.2.2 The Process of Ascription

We have seen in Sections 5.2.1 and 5.2.2.1 that, based on quasi-intensions, a definition for an expression can be given by a set of properties that are accepted by a group of agents as being related to the expression. This is the key point for allowing an anthropologist agent to ascribe an ontological description to a community of agents; it can do so by interviewing the group of informant agents (recall the discussion in Section 4.3.2) that it takes from that particular community. Properties that are associated with a term's definition unanimously among the informant agents should be registered in the construction of an ontology for that community. It is important to note that the anthropologist agent itself needs a theory (i.e., a set of attributes for each term) with which to interview the informant agents. The sentences in this theory will be submitted to the informant agents in order to check whether they accept the sentences or not. In general, the set of sentences to be used in an interview should be the result of observations of the use of language in that society, in the fashion of ethnographers.

We have argued in (Bordini and Campbell 1995) that an anthropologist agent can use its past experience with other communities in its current fieldwork with an unfamiliar society. This idea also applies for the particular problem of creating the set of sentences to be used in the interview. The initial theory for a term to be tested with the informant agents can be taken from previous experiences with other communities or from the definition of that term for the anthropologist agent itself. In case of completely unknown terms, for which observations of communicative actions in the society also do not lead to a useful set of hypothetical sentences, a more elaborate interview will have to take place. The anthropologist agent should then ask the informant agents to state all sentences they have associated with that term in their knowledge representation, instead of just asking them to confirm or deny an initial theory. Once the anthropologist agent has an initial theory obtained from at least one informant agent, it can proceed with the interview as described before. It is reasonable to expect that not all informant agents in all societies will be able to participate in this more complex form of interview. However, in most cases the anthropologist agent should be able to figure out a reasonable initial set of sentences by simply monitoring local conversations, and in these cases our approach requires very little

---

[4] An example in the context of the game of cricket, which is used as a case study in Section 5.4, is the predicate "**is a cricket player**" in hypernymy relation to "**is a batsman**," as batsman is a type of player.

from the informant agents (only that they be able to accept or reject sentences of the language they use). This is a great advantage of the approach, as our aim is to restrict or interfere as little as possible with societies of agents while still allowing them all to interoperate.

Finally, for societies with reduced communication languages, it may be worth presenting all possible sentences (instead of wasting time with observations) for the informant agents to verify them. It therefore appears that the creation of the set of sentences to be used in interviews is culture-dependent (how complex the interview can be depends on how elaborate the society is). In any case, it should normally be based on observations of the language use in particular societies. The techniques that anthropologist agents may use for this can also be drawn from the types of observations carried out by social anthropologists. However, they are not in the scope of the present formalisations.

5.2.3 Discovering Taxonomies in Social Anthropology

This section describes the approach to ethnographic study of cognitive systems as seen by social anthropologists of the cognitive school, and the main concepts involved in the elaboration of taxonomies by anthropologists. It relies heavily on the ideas presented in (Frake 1969), which have allowed us to see that our previous approach to ontologies contained the necessary means to augment ascribed ontologies of terms with the description of the taxonomical relations among those terms.

A common practice among ethnographers is simply "getting names for things" by, e.g., pointing or holding up apparent objects, eliciting their native names and matching them with the investigator's own words for the objects. Instead, Frake proposed the redefinition of the task as one of "finding the 'things' that go with the words" (which, interestingly, resembles the task of our proposed anthropologist agent which must elicit the properties of the words it observes in the traffic of messages in a target society).

Cognitive anthropologists have redefined the task in that way because of their understanding that objects[5] must be defined according to the conceptual systems of the people being studied rather than that of the anthropologist/ethnographer. They are interested in finding out what are the "things" that the members of that community find relevant in their environment. This leads to the discernment of how they interpret their world of experience from the way they talk about it, therefore moulding the "analysis of terminological systems in a way which reveals the conceptual principles that generate them." (Frake 1969).

---

[5] Note that Frake uses the word *object* meaning anything regarded as a member of a category, whether perceptible or not.

Frake goes on to say that different peoples see "things" differently, and mentions that in the past anthropologists sometimes believed this to indicate deficient abstractive ability in primitive societies[6]. He quotes an example concerning a Brazilian Indian tribe which allegedly has no word for *parrot* but only words for *kinds of parrots.* These Indians clearly have a mode of classification for the birds they see which "means that individual bird specimens must be matched against the defining attributes of conceptual categories and thereby judged to be equivalent for certain purposes to some other specimens but different from still others. Since no two birds are alike in every discernible feature, any grouping into sets implies a selection of only a limited number of features as significant for contrasting kinds of birds." Further, the features that are significant are learnt culturally (by every individual from his fellows). Therefore, there is no reason why Brazilian Indians should consider the same attributes which, for an English speaker, make equivalent all the individual birds labelled *parrots.* Knowing how those Indians group objects and which attributes they select as dimensions to generate their taxonomies helps in the construction of a sketch map of the world in the view of the tribe (Frake 1969). Accordingly, we find that some treatment of taxonomical relations must occur in computational societies if one is trying to facilitate migration of agents. We are now able to generate taxonomies (as we shall show in the next section) with the explicit purpose of allowing migrant agents to use a target society's language; but we believe that, in the future, much can be learnt by migrant agents about the culture of target societies by analysing formal descriptions of ontologies and of taxonomical relations among the terms that occur there (see further discussion in Section 8.2).

One might wonder whether (or why) a culture's terminological system is really revealing about the cognitive world of its members (even if not exhaustively so). On this point, it suffices to quote Frake: "Culturally significant features must be communicable between persons in one of the standard symbolic systems of the culture. A major share of these features will undoubtedly be codable in a society's most flexible and productive communication device, its language. ... To the extent that cognitive coding tends to be linguistic and tends to be efficient, the study of referential use of standard, readily elicitable linguistic responses—or *terms*—should provide a fruitful beginning point for mapping a cognitive system." In MAS that is particularly true, as the communication language is likely to include all the terms used for cognitive purposes. In an abstract level, our approach to ontology and taxonomies (which, incidentally, concerns only *terms*) is comparable to ethnography of communication (Hymes 1977) and should therefore be a well-based starting point for a treatment of MAS (concerning interoperability) resembling

---

[6] The word "primitive" is used in the cited source; recall the question we mentioned before about the arrogance on referring to such cultures.

cultural anthropology.

Next we introduce the concepts, given in (Frake 1969), related to the method of arrangements used by anthropologists that is of concern here, namely taxonomies:

Segregate A terminologically distinguished array of objects is a *segregate*[7].

Contrast Culturally appropriate responses which are distinctive alternatives in the same kinds of situations (i.e., occur in the same "environment") can be said to *contrast.*

Contrast Set A series of terminologically contrasted segregates forms a *contrast set.*

Attributes and Dimensions of Contrast The criterial *attributes* which generate a contrast set fall along a limited number of *dimensions of contrast*, each with two or more contrasting values or components (see further comments on attributes below). For example, if in a particular culture a contrast set includes a term for the segregate of "woody plants," another for "herbaceous plants" and another for "vines," two *dimensions of contrast* may be used in the classification of objects: woodiness and rigidity. The first segregate includes objects having *attributes* "woody" and "rigid," the second "not woody" and "rigid," and the third "not rigid."

Inclusion The notion of contrast cannot account for all the significant relations among segregates. Some of them include a wider range of objects than others and are subpartitioned by a contrast set. Those with a wider range are said to *include* each of the segregates in their subpartitioning contrast sets.

Taxonomy A system of contrast sets where some segregates in different contrast sets may be related by inclusion is a *taxonomy.*

It is helpful to make some observations about various consequences of the definitions above:

- Segregates are categories, but not all categories known or knowable to an individual are segregates by the definition above. Segregates are categories designated by particular speech units (terms).
- When one makes a decision about the category membership of an object (by giving it a verbal label), one is selecting a term out of a set of alternatives. When one asserts "This is an X" one is implicitly stating that it is not other things, which are not all conceivable things but those alternatives among which a decision was made. These alternatives are what characterise a contrast set.

---

[7] We shall use that technical term below, even though it does not sound natural in ordinary English as a noun.

- Therefore, the cognitive relation of contrast is not equivalent to the relation of class exclusion in formal logic and set theory. Two categories contrast only when the difference between them is significant for defining the use of the terms associated with them. In other words, having three categories that are mutually exclusive as far as membership is concerned does not mean that they form a contrasting set (i.e., distinctive alternatives in a classifying context).

- To find the relevant attributes is an important task, because having a list of known members of a category is not sufficient to decide how to categorise objects properly. One must find out what natives *know* in order to classify an object correctly as belonging to a specific segregate or to distinguish objects that belong to contrasting segregates in their cultural context. (This is another connection with our approach to ontology: we register the properties, i.e. attributes, that are relevant for a term, which allows the classification of objects as belonging to the segregate it denotes). As Frake puts it, "Categorization, in essence, is a device for treating new experience as though it were equivalent to something already familiar." (Therefore, it is closely related to our motives in this thesis.) A definition of a category in a particular culture is not given by a list of the specimens it contains but by a rule for distinguishing newly-encountered specimens of that category from contrasting alternatives. Again the maxim of the approach of cognitive anthropology stands out very clearly when Frake says: "The distinctive 'situations,' or 'eliciting frames,' or 'stimuli,' which evoke and define a set of contrasting responses are cultural data to be discovered, not prescribed, by the ethnographer. ... It is those elements of *our informants'* experience, which *they* heed in selecting appropriate actions and utterances, that this methodology seeks to discover." This is also the basis of the inspiration of our work on anthropological migration of agents.

Anthropologists (or ethnographers) start their work by recording culturally significant noises and movements from what is heard or seen during observation of a particular community. Recording complementary names applied to the same objects (and eliminating referential synonyms) may yield a recorded sequence like[8]:

Object A is named: *something to eat, sandwich, ham sandwich.*
Object B is named: *something to eat, pie, apple pie.*
Object C is named: *something to eat, pie, cherry pie.*
Object D is named: *something to eat, ice-cream.*

---

[8] The example given in (Frake 1969) concerns a conversation at a lunch counter, and has been abridged here.

The diagram of the sub-partitioning of the segregate "something to eat," as revealed by the naming responses to the four objects above, is in Figure 5.1.

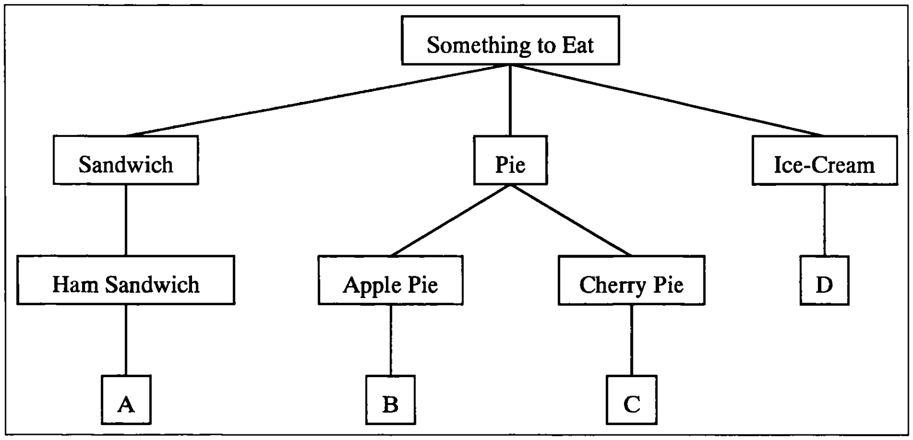


Figure 5.1: Sub-partitioning of the Segregate "Something to Eat" Based on the Naming Responses of Objects A–D (adapted from Frake 1969)

This resembles remarkably our approach to intensional ontology. It therefore allowed us to realise that we already had all the information we needed for the generation of taxonomical relations. Instead of complementary names applied to each object, we have the properties (attributes) that characterise each term in the communication language. Accordingly, by retrieving properties in common and those that differ, we should be able to do exactly the same as ethnographers do, and actually create taxonomies extended with the relevant attributes used to classify objects in one or other segregate. This is what we present in Section 5.3.3.

The ethnographers' work in arriving at diagrams like the one above from their observation and recordings can involve complications, which are not mentioned here. The reader interested in further details of the proceedings of anthropologists should refer to our source (Frake 1969) and more generally (Tyler 1969). One can also find there: the advantages of the use of a taxonomy (e.g., regulation of the amount of information communicated about a particular object in a given situation, hierarchical ordering of categories), with which researchers in the Artificial Intelligence are already familiar; and what are the factors that are likely to determine the degree of elaboration of taxonomies along vertical dimensions of generalization and horizontal dimensions of discrimination in particular cultural contexts. For those who need convincing, the source says about taxonomy's use that "it is a fundamental principle of human thinking."

Another aspect of ethnographic work that is clearly relevant for counterpart computational work on MAS concerns the verification of generated taxonomies. Once a taxonomic partitioning has been worked out, anthropologists have to test it systematically for terminological contrast by means of questions such as "Is that an X?" with an expectation of a negative reply[9]. Using the lunch counter example again, one could point out an apple pie and ask the questions below at the left, to which the answers at the right, reflecting the taxonomy presented above, would be expected:

1. "Is that something to drink?"	1. "No, it is something to eat."
2. "Is that a sandwich?"	2. "No, it is a pie."
3. "Is that a cherry pie?"	3. "No, it is an apple pie."

Frake mentions that "it is easier to do this kind of questioning in a culture where one can assume the role of a naive learner" (as (Lave and Wenger 1991) also pointed out, see Section 4.4). This is exactly how our anthropologist agents should behave, but it does require that societies of agents provide informant agents that are willing to cooperate, as we have remarked in Sections 4.3.1 and 4.3.2 (perhaps emotional agents would be particularly inclined to help naïve learners: see our work on the subject of emotions in Chapter 7).

The similarities also apply at the level of the general motivations of our work. Frake concludes his paper by saying that the real content of culture is how people organise their experience conceptually so that it can be transmitted as knowledge from person to person and from generation to generation. He quotes W.H.Goodenough in saying that culture does not consist of things, people, behaviour, or emotions, but the forms or organisation of these things in the minds of people. The principles of a particular culture reveal how people within it segregate the pertinent from the insignificant, how they code and retrieve information, how they anticipate events, how they define courses of action and make decisions among them. One of the consequences of this approach to ethnographic description (giving central place to the cognitive processes of the actors involved) is that it should result in "descriptions of cultural behaviour, descriptions which, like the linguists' grammar, succinctly state what one must know in order to generate culturally acceptable acts and utterances appropriate to a given socio-ecological context." This is exactly our goal, except that the "socio-ecological context" concerns computational (societies of) agents rather than living organisms. The cognitive approach to anthropology also aims at pointing up critical dimensions for meaningful cross-cultural comparison (as do we in our anthropological approach to MAS) and contributing reliable cultural data to problems of the relations between language, cognition, and behaviour, which is also a key problem for the MAS community.

---

[9] Although not explicitly mentioned in (Frake 1969), it seems to us that, as the ethnographer is interested in testing terminological contrast, asking a question with the expectation of a *negative* reply has the following rationale. Forming a question with a wrong segregate, but one that is at the one particular level of inclusion (i.e., the vertical dimension of generalisation) where contrast is to be tested, will direct the answer from the informant—in its complementary part, that is, after "No"—to state the appropriate segregate *at that particular level of inclusion*, thus giving evidence of the contrasting relation between the inquired-about segregate and the one that occurs in the reply.

This concludes the description of ideas from outside the MAS area that have influenced the extension of our approach to ontologies which accounts for taxonomical relations. We now turn to a formal presentation of our approach.

5.3 Formal Specifications

In this section, an improved formalisation for the process of ascription of intensional ontologies (Bordini, Campbell, and Vieira 1997; 1996) is given, together with a formalisation for retrieval of taxonomical relations. We make reference to the framework for formalisation of agent theories specifically, which was elaborated by Luck and d'Inverno (d'Inverno and Luck 1998; 1996a; Luck and d'Inverno 1995; 1996) based on the Z formal specification language (Spivey 1992; Potter, Sinclair, and Till 1996). A brief account of their framework was given in Section 2.2.1, where further references are also given. We provide, in Appendix B Section B.1, basic notions of the Z notation; however, some familiarity with formal specification methods and mathematical and logical notions is assumed.

5.3.1 The Basic Setting

Before we introduce the formalisation of how an anthropologist agent can ascribe intensional ontologies to societies of agents, the basic setting in which it can occur must be presented (this is much improved in comparison with our previous formalisation given in (Bordini, Campbell, and Vieira 1997)). This section introduces the basic types used in the formalisation, and provide the specifications for informant agents, target societies, and some global functions needed in the rest of the specifications for access to available target societies and to deal with time instants.

We begin by introducing the basic types:

$$[\mathit{Term}, \mathit{Pred}, \mathit{TimeInstant}]$$

and the following abbreviation for the type *Sent*:

$$\mathit{Sent} == (\mathit{Pred} \times \mathit{Term})$$

where *TimeInstant* is taken to be the set of constants representing time instants as it is intuitively understood. *Term* is the set of terms (also called individual constants) of the Communication Language (CL) used by the agents in any Target Society (TS)[10]. *Pred* is the set of predicative constants (predicates) from CL. A sentence (*Sent*) of CL is a pair containing a predicate and a term, meaning that the term has the property (attribute) indicated by the predicate. We

---

[10] Note that *Term* (and *Pred*, mentioned next) are infinite domains; particular target societies will specify the subset of terms (and predicates) they use, as we shall see later.

consider here only sentences of this sort; the consistent acceptance of sentences including the logical connectives by communicating agents within the quasi-intensional approach is given in (Vieira and da Rocha Costa 1993).

We now present the definition of an *InformantAgent*, which is built on the definition of *AutonomousAgent* that is part of Luck and d'Inverno's framework (see, e.g., d'Inverno and Luck 1996a). The only requirement that we impose on the agents that will work as Informant Agents (IA) to the Anthropologist Agents (AA) (the specification of AA is in the next section) is that they make available an acceptance relation *accepts*, in the sense of acceptance we mentioned in Section 5.2.1, Definition 5.1 (note that the relation *accepts* is used in the prefix notation). This should be seen as the interface between IAs and the world, as it is how AAs access the information they need from these agents (for the particular purpose of ontological ascription). The type of the relation makes it clear that each individual IA may or may not accept a certain sentence *s* of its CL, given (by the AA) a set of sentences *S*, at one particular time instant *ti*.

$$
\begin{array}{|l}
\textit{InformantAgent} \\
\hline
\textit{AutonomousAgent} \\
\textit{accepts}\,\_ : \mathbb{P}(\textit{Sent} \times \mathbb{P}\,\textit{Sent} \times \textit{TimeInstant}) \\
\hline
\forall\, s : \textit{Sent};\ S : \mathbb{P}\,\textit{Sent};\ ti : \textit{TimeInstant} \bullet \\
\qquad \textit{accepts}(s, S, ti) \Rightarrow s \in S \\
\hline
\end{array}
$$

The single explicit constraint in the predicate part of the schema above says that IAs only manifest their acceptance of sentences that have been presented to/inquired of them by the experimenter (an AA, in our case); this is how Martin (1959) conceived it in his theory of subjective intensionality. Clearly, this is not sufficient to specify whether an agent accepts a sentence or not. However, the definition of *accepts* is purposely left loose. A complete definition of that relation would need to refer to particular informant agents' mental states and their architectures (and this is of course not desirable in a project aiming at interoperability). For example, if the informant agent works internally as a theorem prover (Fisher 1995), accepting a sentence means simply trying to prove it, and accepting it if it is a theorem in the present set of beliefs of the agent. If the IA is a database system with an agent "wrapping" (Genesereth and Ketchpel 1994), all that is necessary is to check whether the information affirmed in a particular sentence is consistent with the information in the database or not. However agents work, it should always be quite straightforward for designers of agent systems to add this relation as an interface to some particular agents so that they can work as IAs. This is the only requirement that we impose to allow interoperation of agents as far as ontological ascription is concerned. It appears to be quite a reasonable one (especially when contrasted with the degree of constraint implied by the alternative approach of having everything standardised).

Having defined the state space of informant agents, we can now show what designers of societies of agents need to add to their systems so that ascription of intensional ontologies can occur (i.e., the specification of *TargetSociety* below). Before that, we introduce two more basic types. These are the set of constants used as identifiers to TSs (*TSocId*) and to IAs (*InfAgId*). The anthropologist agents and the migrant agents (presented later in Section 5.3.4) should be able to refer to all existing societies of agents (the former analyse them and the latter may need to migrate to them), and for each TS its set of IAs must be identified as well (by the AAs only; remember that AAs can only ascribe IOTs by relying on the IAs of each society). That is why we introduce the following basic types for the constants used for identification (of TSs and IAs).

$$[\mathit{TSocId}, \mathit{InfAgId}]$$

A *TargetSociety* is based on the schema *MASystem* defining what a MAS is (d'Inverno 1998; Section 4.2.2) to which we add all the necessary features for a society of agents to be a *target society* (so that agents can, in our approach, migrate to them). It has a partial injection *iag* from informant agents' identifiers to the actual informant agents in the society. This is used to access the IAs in that particular society (which is why a partial injection is used: two identifiers cannot correspond to the same IA, and only the IAs in that society are in the domain of *iag*). In order to make certain predicates to be introduced later easier to specify, we add a variable *ias* which is constrained to contain all members of the current domain of *iag* (in other words, it contains the set of all identifiers of the IAs that are available in that TS). It is necessarily a non-empty set ($\mathbb{P}_1$) because, as we have said, informant agents are fundamental in this context. Further, a TS has three non-empty sets related to the CL used in it. First, *clterms* is the specific set of terms used in that particular CL. Second, *clpreds* is the set of predicates (or predicative constants) of that CL. As one can see in the predicate part of the schema below, these sets are defined by checking all the terms and predicates that happen to exist in the acceptance relations of all IAs (if there are others at all, they are not at any rate of interest to our work here). Finally, *clsents* is the set of all possible sentences created from the particular terms and predicates of that CL.

$$
\begin{array}{|l}
\multicolumn{1}{l}{\textit{TargetSociety}} \\
\hline
\textit{MASystem} \\
\textit{ias} : \mathbb{P}_1\ \textit{InfAgId} \\
\textit{iag} : \textit{InfAgId} \text{ ⤔ } \textit{InformantAgent} \\
\textit{clterms} : \mathbb{P}_1\ \textit{Term} \\
\textit{clpreds} : \mathbb{P}_1\ \textit{Pred} \\
\textit{clsents} : \mathbb{P}_1 (\textit{Pred} \times \textit{Term}) \\
\hline
\textit{ias} = \operatorname{dom} \textit{iag} \\
\forall\, t : \textit{Term} \bullet (\ t \in \textit{clterms} \Leftrightarrow \\
\qquad \exists\, \textit{ia} : \textit{InfAgId};\ p : \textit{Pred};\ S : \mathbb{P}\ \textit{Sent};\ \textit{ti} : \textit{TimeInstant} \mid \\
\qquad\qquad \textit{ia} \in \textit{ias} \bullet \textit{iag}(\textit{ia}).\textit{accepts}((p, t), S, \textit{ti})\ ) \\
\forall\, p : \textit{Pred} \bullet (\ p \in \textit{clpreds} \Leftrightarrow \\
\qquad \exists\, \textit{ia} : \textit{InfAgId};\ t : \textit{Term};\ S : \mathbb{P}\ \textit{Sent};\ \textit{ti} : \textit{TimeInstant} \mid \\
\qquad\qquad \textit{ia} \in \textit{ias} \bullet \textit{iag}(\textit{ia}).\textit{accepts}((p, t), S, \textit{ti})\ ) \\
\textit{clsents} = (\textit{clpreds} \times \textit{clterms}) \\
\hline
\end{array}
$$

We now introduce in the axiomatic description below some global variables and functions, which will be needed throughout the rest of the specification. The bijection *tsoc* gives a mapping between target society identifiers and actual TSs (there should be a one-to-one correspondence between them, thus a bijection). This is to represent the idea that all existing societies of agents should have a unique identification, and that there is always a way to access the actual TS through their identifiers[11]. Again, for convenience, we add a variable *tsocs* which contains the set of identifiers for all existing TSs (i.e., all members of the current domain of *tsoc*).

Because some of the concepts to be formalised are dependent on time, we need some definitions for handling it. An injective sequence over time instants *the_time* must be available. It is supposed to be the clock of the "system[12]:" it defines the order in which each constant of type *TimeInstant* occurs. Being an injective sequence, it is assured that a constant denoting a time instant occurs no more than once over time, and we must add a predicate saying that all possible time instants are present in the range of the sequence *the_time*, thus giving a complete order for their occurrence. We then have a binary relation (in infix notation) *before_eq* stating whether a time instant t_1 either occurs before or is the same as a time instant t_2. It is defined by checking whether the natural number associated with t_1 in the sequence *the_time* (we do this by using the inverse relation denoted by "$\sim$" superscript) is less than or equal to the one associated with t_2 (it is auxiliary in the definition of *most_recent*, given next). It will make our next definitions easier if we provide a global function *most_recent* which, given a non-empty set of time instants, returns the one that is most recent (i.e., the one with the largest number associated with it in the domain

---

[11] Given the actual infrastructure of international network services, this is not an unrealistic assumption.

[12] Actually, a "world" time, i.e., one that is meaningful across societies.

of *the_time*, which is the last to occur). This is easily defined in terms of the relation *before_eq*, by means of a μ-expression which gives the one *ti* in the provided set of time instants for which it is true that each time instant in the provided set either occurs before *ti* or is *ti* itself.

$$
\begin{array}{|l}
tsocs : \mathbb{P}\ TSocId \\
tsoc : TSocId \rightarrowtail TargetSociety \\
the\_time : \mathrm{iseq}\, TimeInstant \\
\_\ before\_eq\ \_ : TimeInstant \leftrightarrow TimeInstant \\
most\_recent : \mathbb{P}_1\ TimeInstant \rightarrow TimeInstant \\
\hline
tsocs = \mathrm{dom}\, tsoc \\
\mathrm{ran}\, the\_time = TimeInstant \\
\forall t_1, t_2 : TimeInstant \bullet \\
\qquad t_1\ before\_eq\ t_2 \Leftrightarrow the\_time^{\sim}(t_1) \leq the\_time^{\sim}(t_2) \\
\forall tis : \mathbb{P}_1\ TimeInstant \bullet \\
\qquad most\_recent(tis) = (\mu\, ti : TimeInstant \mid ti \in tis \land \\
\qquad\qquad (\forall t : TimeInstant \mid t \in tis \bullet t\ before\_eq\ ti))
\end{array}
$$

Given these basic definitions, we are now ready to see how an anthropologist agent can ascribe an intensional ontology to a target society.

5.3.2 Ascription of Intensional Ontologies of Terms

First we define abbreviations for some types which will be used later. Referring back to Section 5.2.2.1 makes it easy to understand that the signature *IntensionalOntologyOfTerms* is a partial function from terms to non-empty sets of predicates. It is a partial function because it is possible that the AA will not be able to find definitions for all terms used in the TS (and TS itself only uses a subset of them), but if there is an entry for a term in the Intensional Ontology of Terms (IOT), then there must be a non-empty set of predicates which defines it. Referring to Section 5.2.1 leads to the definition of the type *SubjectiveQuasiIntension*: the subjective quasi-intension of a term, for a particular IA, who is from a TS , given a set of sentences (informed by an AA), at a specific time, is a set of predicates which are the properties that the agent accepts as being related to that term. Note that this can be an empty set of predicates if it happens that the IA does not know the particular term in question. The type *IntersubjectiveQuasiIntension* is similar, except that it does not depend on a specific IA (recall that these are relative to the whole group of IAs from a particular TS).

$$
\begin{array}{l}
IntensionalOntologyOfTerms == Term \nrightarrow \mathbb{P}_1\ Pred \\
SubjectiveQuasiIntension == \\
\qquad (Term \times InfAgId \times TSocId \times \mathbb{P}\ Sent \times TimeInstant) \rightarrow \mathbb{P}\ Pred \\
IntersubjectiveQuasiIntension == \\
\qquad (Term \times TSocId \times \mathbb{P}\ Sent \times TimeInstant) \rightarrow \mathbb{P}\ Pred
\end{array}
$$

We now give axiomatic definitions for the functions *subjective_quasi_intension* and *intersubjective_quasi_intension*, which will be used later (when defining the ascription of IOTs). For all terms t used in the CL of that TS, all ia that are informant agents of a target society ts, all sets of sentences S (which are necessarily from that TS's particular CL), and all time instants ti, the *subjective_quasi_intension* of t, for an ia from ts, given S, at time ti, is the set of predicates that ia accepts as being associated with term t, for the set of sentences S, at ti. The *intersubjective_quasi_intension* of t in the TS ts, given S, at ti, is the set of predicates accepted by all informant agents from ts: it is the intersection of the subjective quasi-intensions of all IAs from that TS for that term t (again given S and ti). Note that we need to make use of the *tsoc* function and of the TS's *iag* function to map from identifiers to actual TSs or IAs. We also give below the definition of *co_intensive* which is a function that, given an IOT *iot* as argument, returns the set of all pairs of different terms in the domain of *iot* which have the same set of predicates associated with them in the given *iot*. (Note that in this case we refer to co-intensiveness in terms of intersubjective quasi-intension, which is the type of quasi-intension used in the ascription of ontologies as we shall see later, so as to be able to deal with synonyms.)

$$
\begin{array}{|l}
\mathit{subjective\_quasi\_intension} : \mathit{SubjectiveQuasiIntension} \\
\mathit{intersubjective\_quasi\_intension} : \mathit{IntersubjectiveQuasiIntension} \\
\mathit{co\_intensive}\ \_ : \mathit{IntensionalOntologyOfTerms} \rightarrow \mathbb{P}(\mathit{Term} \times \mathit{Term}) \\
\hline
\forall\, t : \mathit{Term};\ ia : \mathit{InfAgId};\ ts : \mathit{TSocId};\ S : \mathbb{P}\,\mathit{Sent}; \\
\quad ti : \mathit{TimeInstant} \mid t \in \mathit{tsoc}(ts).\mathit{clterms} \land \\
\qquad ia \in \mathit{tsoc}(ts).\mathit{ias} \land S \subseteq \mathit{tsoc}(ts).\mathit{clsents} \bullet \\
\qquad\quad \mathit{subjective\_quasi\_intension}(t, ia, ts, S, ti) = \{p : \mathit{Pred} \mid \\
\qquad\qquad \mathit{tsoc}(ts).\mathit{iag}(ia).\mathit{accepts}((p, t), S, ti)\} \\
\forall\, t : \mathit{Term};\ ts : \mathit{TSocId};\ S : \mathbb{P}\,\mathit{Sent};\ ti : \mathit{TimeInstant} \mid \\
\quad t \in \mathit{tsoc}(ts).\mathit{clterms} \land S \subseteq \mathit{tsoc}(ts).\mathit{clsents} \bullet \\
\qquad \mathit{intersubjective\_quasi\_intension}(t, ts, S, ti) = \\
\qquad\quad \bigcap\{ia : \mathit{InfAgId} \mid ia \in \mathit{tsoc}(ts).\mathit{ias} \bullet \\
\qquad\qquad \mathit{subjective\_quasi\_intension}(t, ia, ts, S, ti)\} \\
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms} \bullet \\
\quad \mathit{co\_intensive}(\mathit{iot}) = \{t_1, t_2 : \mathit{Term} \mid t_1 \neq t_2 \land \\
\qquad t_1 \in \operatorname{dom} \mathit{iot} \land t_2 \in \operatorname{dom} \mathit{iot} \land \mathit{iot}(t_1) = \mathit{iot}(t_2)\}
\end{array}
$$

The definition of *AnthropologistAgent* is given below. It is based, as for the definition of IA, upon the fact that it is an *AutonomousAgent* (provided in the framework) with some additional particular features.

We first say that an AA is able to generate the questions that are needed to interview the IAs (function *generate_sentences*). It is evident that this function is not properly defined in the predicate part of the schema below. The process of generating the necessary questions (i.e., the set of sentences that are submitted for IAs to accept or reject) is discussed in Section 5.3.2, but no formalisation is as yet available for this. However, it is known that the generation of sets of sentences is dependent on the target society and the particular time instant when the interview will take place, hence the signature of *generate_sentences* as given below. Note that this function can return an empty set in situations where the AA does not have much experience with a particular TS at a particular time.

Next, there is the function *history_of_intensional_ontologies*. This is the most important part of AAs because it keeps track of all IOTs an AA has ascribed. Because a TS's IOT may vary over time (we shall comment further on this later), the function *history_of_intensional_ontologies* maps a pair stating a TS and a time instant to the IOT that was ascribed to that TS at that time.

The two items mentioned above are the important aspects of AAs, but we have included a few more variables in the schema in order to make the access to the information from the AA easier in later specifications. The set *known_societies* records all TSs that have been analysed by a particular AA so far; it is the set of all *TSocId* that appear as the first members of the pairs belonging to the domain of *history_of_intensional_ontologies*. There is also *all_versions* which is a function that, given a target society identifier *ts*, provides all the time instants at which IOTs were ascribed to *ts*, provided, of course, that *ts* is in the set of *known_societies*. Further, *current_ontology* maps *ts* (which is as before) to the time interval *ti* that is the *most_recent* of the time intervals associated with *all_versions* of IOTs existing for that *ts* in the *history_of_intensional_ontologies*. Finally, we provide a relation which may be useful for migrant agents' reference: *current_synonyms* gives, for each target society *ts* in *known_societies*, a synonymy relation; it is the (reflexive and transitive) relation over terms created with the help of the function *co_intensive* (in the axiomatic definition above) applied to *ts*'s *current_ontology*.

$$
\begin{array}{|l}
\hline
\mathit{AnthropologistAgent} \\
\hline
\mathit{AutonomousAgent} \\
\mathit{generate\_sentences} : (\mathit{TSocId} \times \mathit{TimeInstant}) \rightarrow \mathbb{P}\, \mathit{Sent} \\
\mathit{history\_of\_intensional\_ontologies} : \\
\qquad (\mathit{TSocId} \times \mathit{TimeInstant}) \nrightarrow \\
\qquad\qquad \mathit{IntensionalOntologyOfTerms} \\
\mathit{known\_societies} : \mathbb{P}\, \mathit{TSocId} \\
\mathit{all\_versions} : \mathit{TSocId} \nrightarrow \mathbb{P}\, \mathit{TimeInstant} \\
\mathit{current\_ontology} : \mathit{TSocId} \nrightarrow \mathit{IntensionalOntologyOfTerms} \\
\hline
\mathit{known\_societies} = \{ s : (\mathit{TSocId} \times \mathit{TimeInstant}) \mid \\
\qquad s \in \operatorname{dom} \mathit{history\_of\_intensional\_ontologies} \bullet \mathit{first}\ s \} \\
\forall\, \mathit{ts} : \mathit{TSocId} \mid \mathit{ts} \in \mathit{known\_societies} \bullet \mathit{all\_versions}(\mathit{ts}) = \\
\qquad \{ \mathit{ti} : \mathit{TimeInstant} \mid (\mathit{ts}, \mathit{ti}) \in \\
\qquad\qquad \operatorname{dom} \mathit{history\_of\_intensional\_ontologies} \} \\
\forall\, \mathit{ts} : \mathit{TSocId};\ \mathit{ti} : \mathit{TimeInstant} \mid \mathit{ts} \in \mathit{known\_societies} \land \\
\qquad \mathit{ti} = \mathit{most\_recent}(\mathit{all\_versions}(\mathit{ts})) \bullet \mathit{current\_ontology}(\mathit{ts}) = \\
\qquad\qquad \mathit{history\_of\_intensional\_ontologies}(\mathit{ts}, \mathit{ti}) \\
\mathit{current\_synonyms} = \{ \mathit{ts} : \mathit{TSocId} \mid \mathit{ts} \in \mathit{known\_societies} \bullet \\
\qquad \mathit{ts} \mapsto \{ t_1, t_2 : \mathit{Term} \mid (t_1, t_2) \in \\
\qquad\qquad \mathit{co\_intensive}(\mathit{current\_ontology}(\mathit{ts})) \} \} \\
\hline
\end{array}
$$

Having defined the state space of *AnthropologistAgent*, to be precise with the Z method we now need to say what are the initial values for the variables in it. The only relevant variable is $\mathit{history\_of\_intensional\_ontologies}'$ (given that the invariants in the schema above relate all other variables to this), and its initial value is evidently the empty set.

$$
\begin{array}{|l}
\hline
\mathit{InitialAnthropologistAgent} \\
\hline
\mathit{AnthropologistAgent}' \\
\hline
\mathit{history\_of\_intensional\_ontologies}' = \varnothing \\
\hline
\end{array}
$$

We can now specify the operation *AscribeIntensionalOntologyOfTerms*, which alters the state of an AA (Δ*AnthropologistAgent*). This operation is given two inputs: $\mathit{ts}?$ is the target society for which an IOT should be ascribed, and $\mathit{ti}?$ gives the time when the ascription is to take place. The operation consist of asserting that $\mathit{history\_of\_intensional\_ontologies}'$ should be overridden from its previous definition to map the pair $(\mathit{ts}?, \mathit{ti}?)$ to the IOT which maps each of the terms t of that TS to its (current) intersubjective quasi-intension, provided this is not an empty set[13]. The set of sentences S that must be provided to the function *intersubjective_quasi_intension* as a parameter

---

[13] Note that in the present formalisation we do not constrain the ontological description of a term to have a minimal set of properties as suggested in Definition 5.4. See further discussion about this in Section 8.2.

(alongside *ts*?, *ti*? and, of course, the term *t*) is produced by the function *generate_sentences* for that particular *ts*? at *ti*?.

$$
\begin{array}{|l}
\hline
\mathit{AscribeIntensionalOntologyOfTerms} \\
\Delta \mathit{AnthropologistAgent} \\
ts? : \mathit{TSocId} \\
ti? : \mathit{TimeInstant} \\
\hline
\mathit{history\_of\_intensional\_ontologies}' = \\
\quad \mathit{history\_of\_intensional\_ontologies} \oplus \\
\quad (\mathbf{let}\ S == \mathit{generate\_sentences}(ts?, ti?) \bullet \\
\qquad \{(ts?, ti?) \mapsto \{t : \mathit{Term} \mid t \in \mathit{tsoc}(ts?).\mathit{clterms} \land \\
\qquad \mathit{intersubjective\_quasi\_intension}(t, ts?, S, ti?) \neq \varnothing \bullet \\
\qquad\quad t \mapsto \mathit{intersubjective\_quasi\_intension}(t, ts?, S, ti?)\}\}) \\
\hline
\end{array}
$$

In brief, the non-empty intersubjective quasi-intension of a term is its definition, in our approach. When the intersubjective quasi-intension is an empty set, the AA cannot ascribe a definition to that term. Recall that by the type of the IOTs (i.e., a partial function) we express the fact that there may not be definitions for all existing terms (not even for all those used in that TS).

Because agents only accept sentences that are in the set of sentences they were given by an AA (stated in *InformantAgent*), and the intersubjective quasi-intension of a term is based on accepted sentences (stated in the axiomatic descriptions), and an ascribed ontology only contains those terms whose intersubjective quasi-intensions are non-empty (in the schema above), we can derive the theorem below which concerns the state space of *AnthropologistAgent* (but only now are we able to introduce it). It says that if there is a term *t* in an ascribed IOT, it is guaranteed that there was at least one sentence concerning that term in the set of sentences generated by the *AnthropologistAgent*. (A corollary would be that if the set of generated sentences is empty, the ascribed IOT is an empty set too.)

$$
\begin{array}{l}
\mathit{AnthropologistAgent}; \\
t : \mathit{Term};\ ts : \mathit{TSocId};\ ti : \mathit{TimeInstant} \mid \\
\quad t \in \mathit{tsoc}(ts).\mathit{clterms} \land \\
\quad (ts, ti) \in \operatorname{dom} \mathit{history\_of\_intensional\_ontologies} \vdash \\
\qquad t \in \operatorname{dom} \mathit{history\_of\_intensional\_ontologies}(ts, ti) \Rightarrow \\
\qquad\quad t \in \{s : \mathit{Sent} \mid s \in \mathit{generate\_sentences}(ts, ti) \bullet \mathit{second}\ s\}
\end{array}
$$

We emphasise that, given that we use the notion of intersubjective quasi-intension, which is time-specific, for the definitions of the terms in the ontology (see *AscribeIntensionalOntologyOfTerms*), these definitions may not be valid *ad infinitum*. Thus, the anthropologist agent may need to review the ontology it has ascribed to a particular society from time to time, as autonomous evolution within societies takes place or the AA alters its set of IAs, or the AA's set of sentences to be given to the IAs is changed, etc. That is why we refer to this type of ontology as *evolutionary*: we intend agents to be able to improve them (and their ascribed versions) with time. Since in our definitions we state that AAs keep track of the whole history of ontologies they have ascribed to each of the TSs, this allows one to analyse how that TS has evolved as far as ontology is concerned. Some agents may be able to analyse the historical evolution of ontologies provided by an AA: one could find it interesting in the future to consider "*historian agents*," or "*linguist agents*" interested in "*agent archaeology*," who might make use of that information.

However, based on the concept of objective quasi-intentions (see Section 5.2.1), some subset of the ontology may form an immutable part of it, composed of the terms universally accepted in that community. In order to deal with this point, we start by providing abbreviations for the appropriate types (as we did for subjective and intersubjective quasi-intension). One should note that *IntertemporalQuasiIntension* is similar to *SubjectiveQuasiIntension* except that it does not depend upon *TimeInstant*. Likewise, *ObjectiveQuasiIntension* resembles *IntersubjectiveQuasiIntension* except for the dependence on time; alternatively, one can see *ObjectiveQuasiIntension* as related to *IntertemporalQuasiIntension* except that the former does not concern particular informant agents.

$$
\begin{aligned}
&\mathit{IntertemporalQuasiIntension} == \\
&\qquad (\mathit{Term} \times \mathit{InfAgId} \times \mathit{TSocId} \times \mathbb{P}\,\mathit{Sent}) \rightarrow \mathbb{P}\,\mathit{Pred} \\
&\mathit{ObjectiveQuasiIntension} == \\
&\qquad (\mathit{Term} \times \mathit{TSocId} \times \mathbb{P}\,\mathit{Sent}) \rightarrow \mathbb{P}\,\mathit{Pred}
\end{aligned}
$$

The axiomatic description below states that the function *intertemporal_quasi_intension*, given a term t, informant agent ia, target society ts, and set of sentences S, yields a set of predicates which ia accepts as being associated with term t at all times, given the set of sentences S. The *objective_quasi_intension* of t in society ts, given S, is the set of predicates accepted by all IAs from ts, for that term, at all times: it is the intersection of the intertemporal quasi-intensions of all IAs in that TS for that term t (again, given S).

$$\begin{array}{l}
\mathit{intertemporal\_quasi\_intension} : \mathit{IntertemporalQuasiIntension} \\
\mathit{objective\_quasi\_intension} : \mathit{ObjectiveQuasiIntension} \\
\hline
\forall\, t : \mathit{Term};\; ia : \mathit{InfAgId};\; ts : \mathit{TSocId};\; S : \mathbb{P}\,\mathit{Sent} \mid \\
\quad t \in \mathit{tsoc}(ts).\mathit{clterms} \land ia \in \mathit{tsoc}(ts).\mathit{ias} \land \\
\quad S \subseteq \mathit{tsoc}(ts).\mathit{clsents} \bullet \\
\qquad \mathit{intertemporal\_quasi\_intension}(t, ia, ts, S) = \\
\qquad\quad \{p : \mathit{Pred} \mid (\forall\, ti : \mathit{TimeInstant} \bullet \\
\qquad\qquad \mathit{tsoc}(ts).\mathit{iag}(ia).\mathit{accepts}((p, t), S, ti))\} \\
\forall\, t : \mathit{Term};\; ts : \mathit{TSocId};\; S : \mathbb{P}\,\mathit{Sent} \mid \\
\quad t \in \mathit{tsoc}(ts).\mathit{clterms} \land S \subseteq \mathit{tsoc}(ts).\mathit{clsents} \bullet \\
\qquad \mathit{objective\_quasi\_intension}(t, ts, S) = \\
\qquad\quad \bigcap\{ia : \mathit{InfAgId} \mid ia \in \mathit{tsoc}(ts).\mathit{ias} \bullet \\
\qquad\qquad \mathit{intertemporal\_quasi\_intension}(t, ia, ts, S)\}
\end{array}$$

In order to say that AAs may also provide immutable intensional ontologies, based on the concepts specified above, we introduce the schema *ExperiencedAnthropologistAgent* which is built on the schema *AnthropologistAgent* and includes a function *immutable_intensional_ontology* which maps TSs to IOTs (it does not depend on time as before, as these IOTs are the ones that are not supposed to change). As in the case of *history_of_intensional_ontologies*, it only maps terms that have a non-empty set of predicates to define them, except that in this instance the set of predicates is given by *objective_quasi_intension* instead of *intersubjective_quasi_intension.* Note that the TS identified by *ts* must necessarily be in the set of *known_societies* of that AA, and the set of sentences S to be verified by informants is defined here as the union of all sentences that the AA generates for that society at all times (the larger this set is the better, as the chances of finding which are the immutable terms in that society are increased).

ExperiencedAnthropologistAgent

$$\begin{array}{l}
\mathit{AnthropologistAgent} \\
\mathit{immutable\_intensional\_ontology} : \mathit{TSocId} \mathrel{+\!\!\!\!\!\rightarrow} \\
\quad \mathit{IntensionalOntologyOfTerms} \\
\hline
\forall\, ts : \mathit{TSocId} \mid ts \in \mathit{known\_societies} \bullet \\
\quad \mathit{immutable\_intensional\_ontology}(ts) = \\
\qquad (\mathbf{let}\; S == \bigcup\{ti : \mathit{TimeInstant} \bullet \mathit{generate\_sentences}(ts, ti)\} \bullet \\
\qquad\quad \{t : \mathit{Term} \mid \mathit{objective\_quasi\_intension}(t, ts, S) \neq \varnothing \bullet \\
\qquad\qquad t \mapsto \mathit{objective\_quasi\_intension}(t, ts, S)\})
\end{array}$$

Migrant agents may well find it useful to know which subset of the intensional ontology is immutable (e.g., if they join a certain society, leave it, and turn to join it in the future, they should try to restrict the terms they use initially to those less likely to have changed in the meantime, to avoid a possible "embarrassment" resulting from the use of an obsolete meaning for a term). Note that some societies may never keep immutable terms, or it may take a long time to arrive at a sound conclusion that there is a immutable subset of an intensional ontology. There is further discussion on this point in Section 8.2.

This completes our specification of the process of ascription of intensional ontologies of terms. We continue by considering how anthropologist agents can retrieve taxonomical relations from the ontologies they have ascribed to target societies using this approach.

5.3.3 Retrieval of Taxonomical Relations

First of all, we describe the type *Taxonomy* using a free type definition. We start by defining *Segregate* (refer to Section 5.2.3 for the concepts of segregate, contrast set and taxonomy) by means of the type constructor *segregate*, which takes a non-empty set of terms and a non-empty set of predicates of the agents' CL. It means that one or more terms (in case of synonyms) are used to refer to a particular segregate in that society and what characterises it is that set of predicates (i.e., the set of properties common to all members of that segregate in the view of the IAs). In this respect, we are presenting here taxonomies that are *augmented* with the characteristic properties of each segregate. Therefore, a taxonomy not only gives a logical structure for the terms used to communicate in that society, but it also records the peculiarities of each segregate in the taxonomy (the *attributes* in the terminology of Section 5.2.3), which allows the identification of the objects that belong to them. Evidently, there are circumstances when such information will be very useful for migrant agents.

A taxonomy is, structurally, a tree. We define the type *Taxonomy* by means of a type constructor *contrastset* which takes a *Segregate* and a set of taxonomies, each member of this set being a subtree representing one of the segregates in the contrast set of *Segregate*, i.e., these segregates are included by inclusion to the segregate in question. The leaves of the tree are segregates that have an empty set of *Taxonomy* for their contrasting sets (i.e., there are no contrast sets for them).

$$\begin{aligned} &\mathit{Segregate} ::= \mathit{segregate}\langle\langle \mathbb{P}_1\ \mathit{Term} \times \mathbb{P}_1\ \mathit{Pred} \rangle\rangle \\ &\mathit{Taxonomy} ::= \mathit{contrastset}\langle\langle \mathit{Segregate} \times \mathbb{P}\ \mathit{Taxonomy} \rangle\rangle \end{aligned}$$

The axiomatic description, given as Algorithm 5.1, presents four functions which are responsible for the retrieval of the taxonomical relations from IOTs.

First, *segregate_size* takes an intensional ontology of terms *iot* and a predicate *p* as parameters. It provides the size of the segregate related to that particular predicate in *iot*. In order to do that, it suffices to count how many terms have the predicate *p* associated with them in *iot*.

Second, *find_segregate* is always given an *iot* whose terms all belong to a single segregate; that is, they all have at least one predicate in common. It then produces that *segregate* and a new IOT

$$
\begin{array}{|l}
\mathit{segregate\_size} : (\mathit{IntensionalOntologyOfTerms} \times \mathit{Pred}) \rightarrow \mathbb{N} \\
\mathit{find\_segregate} : \mathit{IntensionalOntologyOfTerms} \rightarrow \\
\quad (\mathit{Segregate} \times \mathit{IntensionalOntologyOfTerms}) \\
\mathit{find\_contrast\_set} : \mathit{IntensionalOntologyOfTerms} \rightarrow \\
\quad \mathbb{P}\ \mathit{IntensionalOntologyOfTerms} \\
\mathit{find\_taxonomical\_relations} : \\
\quad \mathit{IntensionalOntologyOfTerms} \rightarrow \mathit{Taxonomy} \\
\hline
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms};\ p : \mathit{Pred} \bullet \\
\quad \mathit{segregate\_size}(\mathit{iot}, p) = \\
\qquad \#\{t : \mathit{Term} \mid t \in \operatorname{dom} \mathit{iot} \land p \in \mathit{iot}(t)\} \\
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms} \bullet \mathit{find\_segregate}(\mathit{iot}) = \\
\quad (\mathbf{let}\ P == \{p : \mathit{Pred} \mid (\forall\, t : \mathit{Term} \mid \\
\qquad t \in \operatorname{dom} \mathit{iot} \bullet p \in \mathit{iot}(t))\} \bullet \\
\qquad\quad (\mathbf{let}\ T == \{t : \mathit{Term} \mid t \in \operatorname{dom} \mathit{iot} \land \mathit{iot}(t) = P\} \bullet \\
\qquad\qquad (\mathbf{let}\ \mathit{siot} == \{t : \mathit{Term} \mid t \in \operatorname{dom} \mathit{iot} \land t \notin T \bullet \\
\qquad\qquad\quad t \mapsto \mathit{iot}(t) \setminus P\} \bullet \\
\qquad\qquad\qquad (\mathit{segregate}(T, P), \mathit{siot})))) \\
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms} \bullet \\
\quad (\mathit{iot} = \varnothing \Rightarrow \mathit{find\_contrast\_set}(\mathit{iot}) = \varnothing) \land \\
\quad (\mathit{iot} \neq \varnothing \Rightarrow (\exists\, \mathit{sp} : \mathit{Pred} \mid \\
\qquad \mathit{sp} \in \bigcup(\operatorname{ran} \mathit{iot}) \land \\
\qquad \mathit{sp} \in \{p : \mathit{Pred} \mid p \in \bigcup(\operatorname{ran} \mathit{iot}) \land \\
\qquad\quad (\forall\, q : \mathit{Pred} \mid q \in \bigcup(\operatorname{ran} \mathit{iot}) \bullet \\
\qquad\qquad \mathit{segregate\_size}(\mathit{iot}, p) \geq \mathit{segregate\_size}(\mathit{iot}, q))\} \bullet \\
\qquad\qquad\quad \mathit{find\_contrast\_set}(\mathit{iot}) = \\
\qquad\qquad\qquad (\mathbf{let}\ \mathit{csiot} == \{t : \mathit{Term} \mid t \in \operatorname{dom} \mathit{iot} \land \\
\qquad\qquad\qquad\quad \mathit{sp} \in \mathit{iot}(t) \bullet t \mapsto \mathit{iot}(t)\} \bullet \\
\qquad\qquad\quad \{\mathit{csiot}\} \cup \mathit{find\_contrast\_set}(\mathit{iot} \setminus \mathit{csiot})))) \\
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms} \bullet \\
\quad \mathit{find\_taxonomical\_relations}(\mathit{iot}) = \\
\qquad \mathit{contrastset}(\mathit{first}\ \mathit{find\_segregate}(\mathit{iot}), \\
\qquad\quad \{\mathit{csiot} : \mathit{IntensionalOntologyOfTerms} \mid \mathit{csiot} \in \\
\qquad\qquad \mathit{find\_contrast\_set}(\mathit{second}\ \mathit{find\_segregate}(\mathit{iot})) \bullet \\
\qquad\qquad\quad \mathit{find\_taxonomical\_relations}(\mathit{csiot})\})
\end{array}
$$

Algorithm 5.1: Algorithm for the Retrieval of Taxonomical Relations in an Ascribed Intensional Ontology of Terms *siot* in the following way. Let P be the set of all predicates common to all terms in *iot*. Then T is the set of all terms having exactly P as their definition in *iot* (note that if they have the same set of predicates, they are synonyms). These are the terms that represent the whole segregate (and P is what characterises it), and all other items in *iot* belong to the segregates that form the contrast set for this particular segregate. We define *siot* as an intensional ontology which is the same as *iot* but does not include any of the terms that are in T (as these are already used in the taxonomy, to define the present segregate). Besides, we remove all the predicates (P) that are common to the definition of all the remaining terms (the ones in *siot*); again this information is already in the segregate of which the terms in *siot* form the contrast set. The function *find_segregate* then returns the *segregate* formed by T and P, and also returns *siot*, to be used by the function described next in this process of retrieving taxonomical relations. (Once the common predicates are removed from the definitions of the remaining terms (*siot*), the remaining predicates can be used to split *siot* into contrasting segregates, in the next step).

Third, *find_contrast_set* is a recursive function which splits *iot* into a set of IOTs so that each of these IOTs resulting from the split is guaranteed to have at least one predicate in common (that is, all the terms in it belong to a single segregate and can then be used by the function described above). The end of the recursion occurs when the current *iot* is empty, in which case an empty IOT is returned. Otherwise, the predicate *sp* which yields the largest segregate[14] in the current *iot* is selected and an IOT *csiot* is created by gathering all terms which belong to that segregate (i.e., have *sp* in their definition in *iot*). Thus, the terms in *csiot* belong to one of the contrasting segregates that is present in *iot* and the remaining terms in it ($iot \setminus csiot$) are given recursively to *find_contrast_set* for the remaining IOTs (representing the other contrasting segregates) to be added to the final set of IOTs which is returned by this function eventually. We emphasise here that the criteria used to select the segregates for the contrast set (i.e., based on the predicate that yield the largest segregate first) may in some cases create taxonomies that do not have the exact structure used by the IAs. This is why we have mentioned that the taxonomy created by the process introduced here is just tentative, and the AA should check it by interviewing the IAs again in order to determine whether they accept that structure (see further discussion in Section 8.2).

Finally, *find_taxonomical_relations* is also a recursive function which takes an IOT and returns a *Taxonomy*. Given an *iot*, it returns a *contrastset* whose *Segregate* is given by the *first* coordinate of the pair returned by *find_segregate* applied to that *iot*. The set of taxonomies associated with that

---

[14] Note that a μ-expression is not allowed here because several predicates yielding the same segregate size may exist. We therefore state only that *sp* belongs to the set of predicates yielding the same largest segregate size (which is a somewhat ambiguous, but suitable for the purpose here).

Segregate in *contrastset* is created by applying *find_taxonomical_relations* recursively to a set of IOTs. These are returned by *find_contrast_set* when applied to the IOT that is the *second* coordinate of the pair returned by *find_segregate* applied to that *iot*. In other words, one creates a segregate, eliminates from *iot* the terms already used in the creation of the segregate, eliminates from the remaining terms the common predicates (that already belong to the created segregate), splits the resulting ontology into IOTs representing contrasting segregates, and proceeds in the same way for each of them. At some point, *find_segregate* returns, as second coordinate, an empty IOT which will lead to an empty set of taxonomies (denoting a leaf of the tree), thus ending the recursion.

One more definition will be necessary for the specification of the extended capabilities of an Anthropologist Agent (*TaxonomistAnthropologistAgent* given later). The function *generate_taxonomical_relations* maps an IOT to functions from predicates to taxonomies. The idea is that for each *iot* there is a function which, given a predicate, returns the taxonomy that can be retrieved in a subset of *iot* which only contains the terms that have that particular predicate in their definitions. We need this because when the whole IOT of a TS is considered, several of the terms may not be related to each other (in the example to be given in the next section, the ontology includes terms for cricket players bearing no relation to the terms for field divisions in the same ontology). Therefore, it is necessary to give a general subject for which a taxonomy is required before actually proceeding with the algorithm for retrieval of taxonomical relations (presented in Algorithm 5.1) in order to retrieve a particular taxonomy related to that subject (note that the algorithm assumes to start from an IOT where all terms are related to a single subject).

The first constraint in the axiomatic description below states that for a predicate p to be in the domain of the function yielded by the application of *generate_taxonomical_relations* to *iot*, there must exist at least one term having p in its definition in *iot*. Further, when one applies *generate_taxonomical_relations* to *iot* and then applies the resulting function to a predicate p, an IOT *piot*, containing only the terms in *iot* which have p in their definitions, is passed as a parameter to the function *find_taxonomical_relations* defined earlier so that a taxonomy for that particular subject (the predicate p) is generated.

$$
\begin{array}{|l}
\mathit{generate\_taxonomical\_relations} : \\
\qquad \mathit{IntensionalOntologyOfTerms} \rightarrow (\mathit{Pred} \mathrel{+\!\!\!\!\rightarrow} \mathit{Taxonomy}) \\
\hline
\forall\, \mathit{iot} : \mathit{IntensionalOntologyOfTerms};\ p : \mathit{Pred} \bullet \\
\qquad p \in \operatorname{dom} \mathit{generate\_taxonomical\_relations}(\mathit{iot}) \Rightarrow \\
\qquad\qquad (\exists\, t : \mathit{Term} \bullet p \in \mathit{iot}(t)) \\
\forall\, \mathit{iot}, \mathit{piot} : \mathit{IntensionalOntologyOfTerms};\ p : \mathit{Pred} \mid \\
\qquad (\forall\, t : \mathit{Term} \mid t \in \operatorname{dom} \mathit{iot} \bullet \\
\qquad\qquad (t \in \operatorname{dom} \mathit{piot} \Leftrightarrow p \in \mathit{iot}(t)) \land \mathit{piot}(t) = \mathit{iot}(t)) \bullet \\
\qquad\qquad\qquad \mathit{generate\_taxonomical\_relations}\ \mathit{iot}\ p = \\
\qquad\qquad\qquad\qquad \mathit{find\_taxonomical\_relations}(\mathit{piot})
\end{array}
$$

We can now improve the capabilities of an AA to include the generation of taxonomies as well. The schema *TaxonomistAnthropologistAgent* is built upon *ExperiencedAnthropologistAgent*. The function *generate_subjects* resembles the function *generate_sentences*, but it is intended to generate a set of predicates (for a given TS at a certain time *ti*) which are the subjects for which the AA believes that taxonomies exist (or could exist) in that particular TS at *ti*. Similarly to the process of finding the sentences to be used in the interview, the AA should be able to generate these subjects by means of its observation of the use of the CL of that TS. As in the case of *generate_sentences*, this process is not yet formalised, so *generate_subjects* is also loosely defined. Also as before, *generate_subjects* returns an empty set should the AA not be experienced enough in the study of a TS to be able to decide what are the subjects for which it is worth generating taxonomies in that society.

Apart from the ability to generate subjects, a *TaxonomistAnthropologistAgent* makes available a function *taxonomical_relations* which, given a target society identifier, leads to a set of taxonomies used in that society. The target society represented by *ts*, for which taxonomies are required, clearly must be in the set of *known_societies* (otherwise the AA could not have any ontological information about it, let alone generate taxonomical relations). We take *ti* to be the *most_recent* of *all_versions* of IOT existing for that *ts*. The *taxonomical_relations* for *ts* will then be the set of all taxonomies recovered from the *current_ontology* of *ts*. Each of these taxonomies is associated with each of the predicates (subjects) that are generated for that *ts* at *ti* (the time at which the *current_ontology* was generated).

$$
\begin{array}{|l}
\multicolumn{1}{l}{\textit{TaxonomistAnthropologistAgent}} \\
\hline
\textit{ExperiencedAnthropologistAgent} \\
\textit{generate\_subjects} : (\textit{TSocId} \times \textit{TimeInstant}) \rightarrow \mathbb{P}\,\textit{Pred} \\
\textit{taxonomical\_relations} : \textit{TSocId} \nrightarrow \mathbb{P}\,\textit{Taxonomy} \\
\hline
\forall\, ts : \textit{TSocId};\ ti : \textit{TimeInstant} \mid \\
\qquad ts \in \textit{known\_societies} \land \\
\qquad ti = \textit{most\_recent}(\textit{all\_versions}(ts)) \bullet \\
\qquad\qquad \textit{taxonomical\_relations}(ts) = \\
\qquad\qquad\qquad \{p : \textit{Pred} \mid p \in \textit{generate\_subjects}(ts, ti) \bullet \\
\qquad\qquad\qquad \textit{generate\_taxonomical\_relations}\ \textit{current\_ontology}(ts)\ p\} \\
\hline
\end{array}
$$

Similarly to the theorem derived from *AnthropologistAgent* before, we can now derive a useful theorem concerning the schema above. It says that, given that *ts* belongs to the set of *known_societies*, *ti* is the time at which the most recent IOT has been ascribed to *ts*, and *SP* is the set of all the sets of predicates that are at the root of each of the current taxonomies for *ts* ($\textit{taxonomical\_relations}(ts)$), the following applies. For each of the sets of predicates in *SP*, there exist at least one predicate *p* such that: (i) *p* is one of the subjects for *ts* at *ti*, generated by an AA ($\textit{generate\_subjects}(ts, ti)$); and (ii) at least one term in the $\textit{current\_ontology}(ts)$ has *p* associated with it. This follows from the definitions of *taxonomical_relations* (which generates a taxonomy for each of those subjects, so that each is in the set of predicates at the root of a taxonomy), and the definition of *generate_taxonomical_relations* (which selects a subset of the IOT concerning a particular subject).

$$
\begin{array}{l}
\textit{TaxonomistAnthropologistAgent}; \\
ts : \textit{TSocId};\ ti : \textit{TimeInstant};\ SP : \mathbb{P}\,\mathbb{P}_1\,\textit{Pred} \mid \\
\qquad ts \in \textit{known\_societies} \land \\
\qquad ti = \textit{most\_recent}(\textit{all\_versions}(ts)) \\
\qquad SP = \{T : \mathbb{P}_1\,\textit{Term};\ P : \mathbb{P}_1\,\textit{Pred} \mid \\
\qquad\qquad \textit{segregate}(T, P) \in \textit{taxonomical\_relations}(ts) \bullet P\} \vdash \\
\qquad\qquad\qquad \forall\, sp : \mathbb{P}_1\,\textit{Pred} \mid sp \in SP \bullet \\
\qquad\qquad\qquad\qquad \exists\, p : \textit{Pred} \mid p \in sp \bullet \\
\qquad\qquad\qquad\qquad\qquad p \in \textit{generate}_s\textit{ubjects}(ts, ti) \land \\
\qquad\qquad\qquad\qquad\qquad p \in \bigcup(\operatorname{ran} \textit{current\_ontology}(ts))
\end{array}
$$

A corollary, similar to the one for the previous theorem, would be that if the AA's observation does not lead to indications of what are the main subjects relevant to that TS (i.e., $\textit{generate\_subjects}(ts, ti)$ returns an empty set), no taxonomical relation can be retrieved for it; the same applies if the predicate is one of the subjects but is not present in the current IOT. It is interesting to note that the theorems are similar, the one for *AnthropologistAgent* being concerned with propositions about terms, whilst the one for *TaxonomistAnthropologistAgent* is concerned with propositions about predicates.

This ends the formalisation of the retrieval of taxonomical relations. It is additionally worth formalising a few general ideas about migration of agents in this context, which we examine next.

5.3.4 Migration of Agents

In order to round off the ideas formalised so far, we present some specifications on migration of agents (showing what is the information that an AA has collected which can be useful for a migrant agent (MA) in its process of migration). Below is the schema *MigrantAgent*; as before for IA and AA, it is based on *AutonomousAgents* with a few extra features. Its variable *ts* identifies the TS that the MA intends to join. Further, two IOT variables are needed, one for the current *intensional_ontology* from *ts* and another for the *immutable_ontology*, in case there is any. Also, there is the relation on terms for the *synonyms* in that *ts*'s CL, and the set of *taxonomies* used in that *ts*.

MigrantAgent

$AutonomousAgent$

$ts : TSocId$

$intensional\_ontology, immutable\_ontology :$
$\quad IntensionalOntologyOfTerms$

$synonyms : Term \leftrightarrow Term$

$taxonomies : \mathbb{P}\ Taxonomy$

As for *AnthropologistAgent*, we need to define the initial state for the schema above. We introduce $\perp$ to mean an identifier to an undefined TS. The variable ts' begins with this value, and all other variables stand initially for empty sets.

$\perp : TSocId$

InitialMigrantAgent

$MigrantAgent'$

$ts' = \perp$

$intensional\_ontology' = immutable\_ontology' = \varnothing$

$synonyms' = \varnothing$

$taxonomies' = \varnothing$

Finally, we define the operation *Migrate* which changes an MA ($\Delta MigrantAgent$) and also needs the mediation of an AA whose state is not thereby altered ($\Xi\, TaxonomistAnthropologistAgent$). This operation takes as input a TS identifier $ts?$ which must belong to the set of *known_societies* of the relevant AA[15]. The value of the input *ts*? is used to set *ts*′; *intensional_ontology*′ is given the *current_ontology* for that *ts*? (i.e., the last IOT the AA has ascribed to *ts*?); *immutable_ontology*′ gets its value from the function provided in the *ExperiencedAnthropologistAgent* schema; *synonyms*′ receives the *current_synonyms* (recall, from the *current_ontology*) from *ts*?; and *taxonomies*′ has the *taxonomical_relations* for *ts*?.

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{Migrate}} \\
\Delta \mathit{MigrantAgent} \\
\Xi \mathit{TaxonomistAnthropologistAgent} \\
\mathit{ts?} : \mathit{TSocId} \\
\hline
\mathit{ts?} \in \mathit{known\_societies} \\
\mathit{ts}' = \mathit{ts?} \\
\mathit{intensional\_ontology}' = \mathit{current\_ontology}(\mathit{ts?}) \\
\mathit{immutable\_ontology}' = \mathit{immutable\_intensional\_ontology}(\mathit{ts?}) \\
\mathit{synonyms}' = \mathit{current\_synonyms}(\mathit{ts?}) \\
\mathit{taxonomies}' = \mathit{taxonomical\_relations}(\mathit{ts?}) \\
\hline
\end{array}
$$

We have found, as d'Inverno (1998) claims and demonstrates for applications of a size similar to this one, that Z is an excellent basis for clear specification of agents with special properties. It has allowed an effective progression from the initial qualitative ideas on agency with an anthropological flavour to the precise form given above, and to theorems that they satisfy. Furthermore, it should allow us to formalise the missing parts incrementally (some are mentioned in Section 8.2). Additionally, building on definitions from Luck and d'Inverno's framework (i.e., *AutonomousAgent* and *MASystem*), makes it possible for us to integrate our approach with other agent theories specified in the same framework (e.g., Sichman *et al.* 1994, see d'Inverno and Luck 1996a; others were mentioned in Section 2.2.1), besides the obvious advantage of exempting us from specifying those basic concepts. Also, as d'Inverno and Luck indicate (1996a), the framework can be used directly in the implementation of simulations of the agent theories that have been formalised. As a matter of fact, we have type-checked these specifications using ZTC (Jia 1995) and animated a simplified[16] version of them using PIZA (Hewitt 1997). The executable version of the specifications given in this section can be found in Appendix B, Section B.2, and the results for the cricket example given in the next section, can be found in Section B.3 of that appendix.

---

[15] But note that this is anthropologically mediated migration; we have mentioned that migration can occur without anthropologist agents, despite the increased problems of adaptation in such cases.

[16] The simplifications concern mainly some of the global definitions for the basic setting (e.g. access to TSs), which are not directly implementable in the Z tools used. All main operations are seen to work in the animated version as intended in the specified theory.

Any difficulties remaining from the reading of the formalisation presented should be resolved by the illustration of their use given in the next section.

5.4 Case Study: an Ontology from the Game of Cricket and Two of its Taxonomies

In order to illustrate the meaning of the specifications presented in the previous section, we shall make use of some terms used in the game of cricket. We have made the exotic choice of cricket because we already use that context in the simulation in Chapter 6, and because the information content of this example is not trivial. (Routinely, it baffles foreign human observers.)

5.4.1 Two Cricket Taxonomies

A few cricket terms are presented in the form of two taxonomies, some of which will be used later in the example of the use of the specifications. Figure 5.2 includes some of the terms used to refer to the players. Figure 5.3 holds terms related to the field where the game is played. In the figures, "..." means that a whole subtree of terms is omitted from the figure (for the sake of space and clarity). In Figure 5.3, the symbol "*" means that there are alternative and non-exclusive ways of partitioning a cricket field, which we do not mention there.

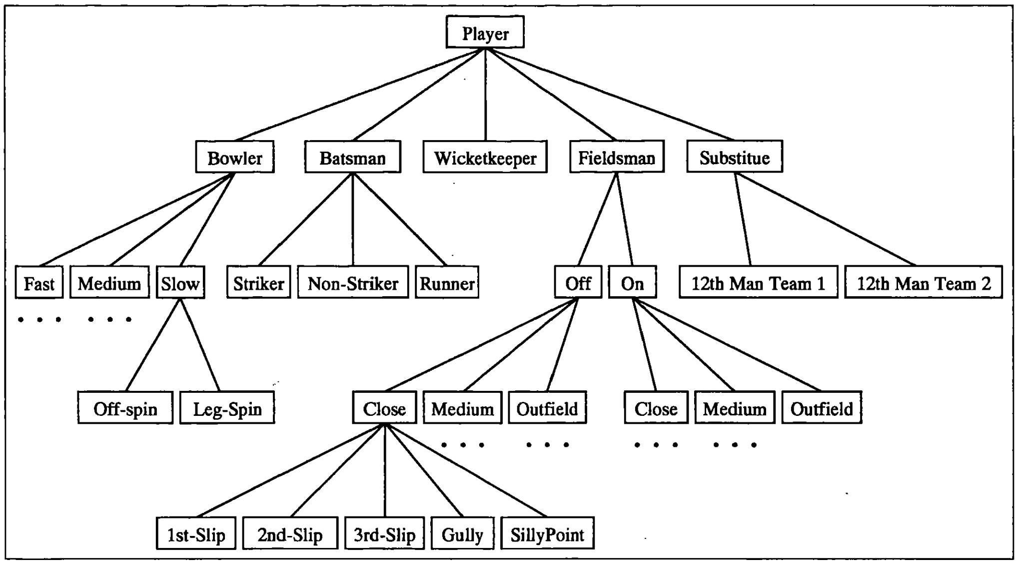


Figure 5.2: Taxonomy of Terms Referring to the Cricket Players

In the next section we shall use a subset of these terms to show how an intensional ontology for them can be ascribed to a *cricket society of agents*. Later, we show how the taxonomical relations among them can be retrieved from the intensional ontology.

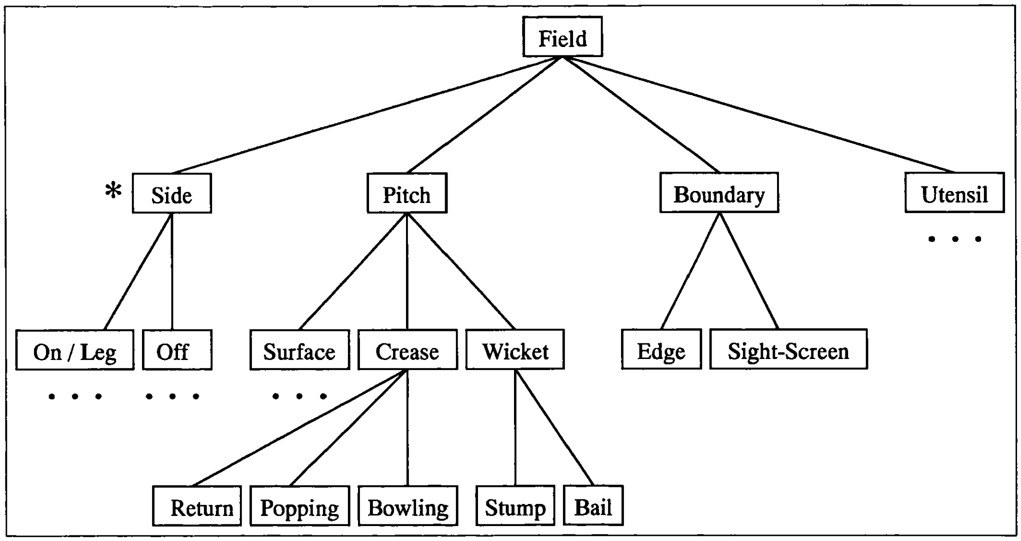


Figure 5.3: Taxonomy of Terms Related to the Cricket Field

5.4.2 Ascribing an Intensional Ontology to the Cricket Society

The ascription of intensional ontologies relies essentially on the acceptance relations of informant agents. In order to specify their acceptance relations, we first need to introduce the particular set of terms we use, and give acronyms for the appropriate predicates. We do so in the form of the following Z predicates:

$$
\begin{aligned}
&tsoc(CSoc).ias = \{IAg_1, IAg_2\}\\
&tsoc(CSoc).clterms = \{player, bowler, fastblr, slowblr, batsman, wicketkpr,\\
&\quad field, side, on, leg, off, pitch\}\\
&tsoc(CSoc).clpreds = \{HB, TB, TBF, IS, TBS, IA, CB, WG,\\
&\quad WP, FP, LRB, RRB, CF\}
\end{aligned}
$$

where *CSoc* is a *TSocId* referring to the cricket society; we say that the cricket society has two informant agents identified by IAg_1 and IAg_2 (only two for the sake of simplicity); and we have used the following acronyms: Human Being (*HB*), Throws the Ball (*TB*), Throws the Ball Fast (*TBF*), Throws the Ball Slowly (*TBS*), Is Strong (*IS*), Is Accurate (*IA*), Carries a Bat (*CB*), Wears Gloves (*WG*), location Where the game is Played (*WP*), Field Partition (*FP*), to the Left of a Right-handed Batsman (*LRB*), to the Right of a Right-handed Batsman (*RRB*) and Centre of the Field (*CF*). We have also used *fastblr* for "fast-bowler," *slowblr* for "slow-bowler," and *wicketkpr* for "wicketkeeper."

We now define the IAs' acceptance relations. Again for simplicity we assume that all possible sentences ($tsoc(CSoc).clsents$) are presented by the AA to the IAs for them to verify each of them, and we only consider one time instant ti_1. In most practical cases, to produce all possible sentences would be intractable, but given the small number of terms and predicates of the

CL used here as an example, and that the point to be made clear here does not concern the interview (as it is not formalised), this simplification is quite reasonable.

$$
\begin{array}{l}
\{s : \mathit{Sent} \mid \mathit{tsoc}(\mathit{CSoc}).\mathit{iag}(\mathit{IAg}_1).\mathit{accepts}(s, \mathit{tsoc}(\mathit{CSoc}).\mathit{clsents}, \mathit{ti}_1)\} = \{ \\
\quad (\mathit{HB}, \mathit{player}), (\mathit{HB}, \mathit{bowler}), (\mathit{TB}, \mathit{bowler}), \\
\quad (\mathit{HB}, \mathit{fastblr}), (\mathit{TB}, \mathit{fastblr}), (\mathit{TBF}, \mathit{fastblr}), (\mathit{IS}, \mathit{fastblr}), \\
\quad (\mathit{HB}, \mathit{slowblr}), (\mathit{TB}, \mathit{slowblr}), (\mathit{TBS}, \mathit{slowblr}), (\mathit{IA}, \mathit{slowblr}), \\
\quad (\mathit{HB}, \mathit{batsman}), (\mathit{CB}, \mathit{batsman}), \\
\quad (\mathit{HB}, \mathit{wicketkpr}), (\mathit{WG}, \mathit{wicketkpr}), \\
\quad (\mathit{WP}, \mathit{field}), (\mathit{WP}, \mathit{side}), (\mathit{FP}, \mathit{side}), \\
\quad (\mathit{WP}, \mathit{on}), (\mathit{FP}, \mathit{on}), (\mathit{LRB}, \mathit{on}), \\
\quad (\mathit{WP}, \mathit{leg}), (\mathit{FP}, \mathit{leg}), (\mathit{LRB}, \mathit{leg}), \\
\quad (\mathit{WP}, \mathit{off}), (\mathit{FP}, \mathit{off}), (\mathit{RRB}, \mathit{off}), \\
\quad (\mathit{WP}, \mathit{pitch}), (\mathit{CF}, \mathit{pitch})\} \\
\{s : \mathit{Sent} \mid \mathit{tsoc}(\mathit{CSoc}).\mathit{iag}(\mathit{IAg}_2).\mathit{accepts}(s, \mathit{tsoc}(\mathit{CSoc}).\mathit{clsents}, \mathit{ti}_1)\} = \{ \\
\quad (\mathit{HB}, \mathit{player}), (\mathit{HB}, \mathit{bowler}), (\mathit{TB}, \mathit{bowler}), \\
\quad (\mathit{HB}, \mathit{fastblr}), (\mathit{TB}, \mathit{fastblr}), (\mathit{TBF}, \mathit{fastblr}), \\
\quad (\mathit{HB}, \mathit{slowblr}), (\mathit{TB}, \mathit{slowblr}), (\mathit{TBS}, \mathit{slowblr}), (\mathit{IA}, \mathit{slowblr}), \\
\quad (\mathit{HB}, \mathit{batsman}), (\mathit{CB}, \mathit{batsman}), \\
\quad (\mathit{HB}, \mathit{wicketkpr}), (\mathit{WG}, \mathit{wicketkpr}), \\
\quad (\mathit{WP}, \mathit{field}), (\mathit{WP}, \mathit{side}), (\mathit{FP}, \mathit{side}), \\
\quad (\mathit{WP}, \mathit{on}), (\mathit{FP}, \mathit{on}), (\mathit{LRB}, \mathit{on}), \\
\quad (\mathit{WP}, \mathit{leg}), (\mathit{FP}, \mathit{leg}), (\mathit{LRB}, \mathit{leg}), \\
\quad (\mathit{WP}, \mathit{off}), (\mathit{FP}, \mathit{off}), (\mathit{RRB}, \mathit{off}), \\
\quad (\mathit{WP}, \mathit{pitch}), (\mathit{CF}, \mathit{pitch})\}
\end{array}
$$

Note that IAg_2 does not agree with IAg_1 with respect to whether a *fast-bowler* is necessarily strong (*IS*) or not. Therefore, the AA will not consider this predicate as part of the intensional definition for *fastblr*, as this is likely to be the consequence of individual observations or experiences of some of the IAs in particular. (Interestingly, to do so would be closer to knowledge representation rather than ontological commitment.)

If we execute the operation *AscribeIntensionalOntologyOfTerms* in this context, that is:

ExampleAscription

$$
\begin{array}{|l}
\mathit{AscribeIntensionalOntologyOfTerms} \\
\hline
\mathit{ts}? = \mathit{CSoc} \\
\mathit{ti}? = \mathit{ti}_1 \\
\hline
\end{array}
$$

we get the following IOT:

$$
\begin{array}{l}
\mathit{ExampleAscription} \mid \\
\quad \mathit{generate\_sentences}(ts?, ti?) = \mathit{tsoc}(\mathit{CSoc}).\mathit{clsents} \\
\qquad \vdash \mathit{current\_ontology}'(ts?) = \{ \\
\qquad\qquad \begin{array}{lll}
\mathit{player} & \mapsto & \{HB\}, \\
\mathit{bowler} & \mapsto & \{HB, TB\}, \\
\mathit{fastblr} & \mapsto & \{HB, TB, TBF\}, \\
\mathit{slowblr} & \mapsto & \{HB, TB, TBS, IA\}, \\
\mathit{batsman} & \mapsto & \{HB, CB\}, \\
\mathit{wicketkpr} & \mapsto & \{HB, WG\}, \\
\mathit{field} & \mapsto & \{WP\}, \\
\mathit{side} & \mapsto & \{WP, FP\}, \\
\mathit{on} & \mapsto & \{WP, FP, LRB\}, \\
\mathit{leg} & \mapsto & \{WP, FP, LRB\}, \\
\mathit{off} & \mapsto & \{WP, FP, RRB\}, \\
\mathit{pitch} & \mapsto & \{WP, CF\}\}
\end{array}
\end{array}
$$

Note that the AA's synonymy relation for *CSoc* in this context (that is, the value of *current_synonyms*) would be $\{(on, leg), (leg, on)\}$ (i.e., these are co-intensive terms, or synonyms).

We do not give an example regarding immutable ontologies of terms, since we are only representing a single time instant for the sake of clarity and space.

5.4.3 Recovering the Taxonomical Relations from the Cricket Ontology

What remains to be illustrated now is how the taxonomical relations are recovered from the intensional ontology presented in the previous section. We use Tx_1 and Tx_2 as global variables over *Taxonomy* and we show what their contents are in Figure 5.4 and Figure 5.5 respectively so as to make their visualisation clearer.

$$
\mid \; Tx_1, Tx_2 : \mathit{Taxonomy}
$$

$$
\begin{array}{l}
\mathit{ExampleAscription}; \\
\mathit{TaxonomistAnthropologistAgent} \mid \\
\quad \mathit{generate\_subjects}(ts?, ti?) = \{HB, WP\} \vdash \\
\qquad \mathit{taxonomical\_relations}(ts?) = \{Tx_1, Tx_2\}
\end{array}
$$

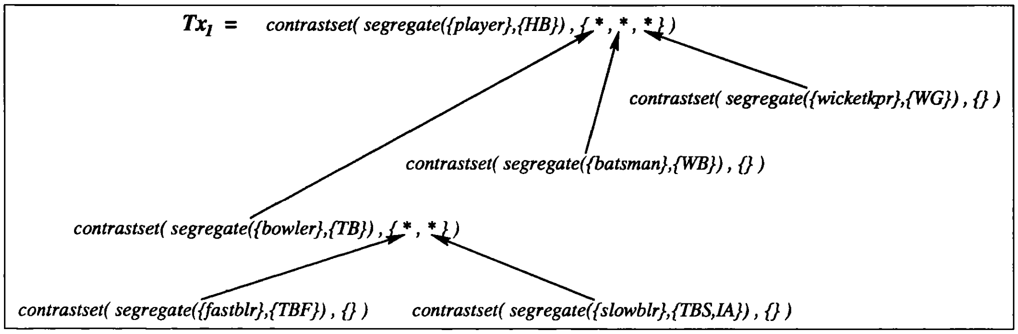


Figure 5.4: Tx_1—Recovered Taxonomy of Terms Referring to the Cricket Players

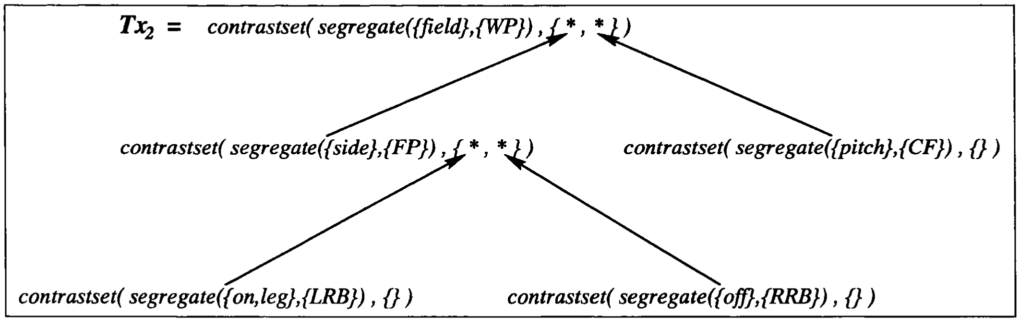


Figure 5.5: Tx_2—Recovered Taxonomy of Terms Related to the Cricket Field

We now show some steps of the generation of the taxonomy Tx_1 above so that the algorithm used may become clearer. We shall use *csiot* to refer to the IOT ascribed to the cricket society in the previous section. When using *generate_taxonomical_relations*(*csiot*, *HB*) we would get the following new IOT *hbiot*:

$$
\begin{array}{lll}
\multicolumn{3}{l}{hbiot = \{} \\
\quad player & \mapsto & \{HB\}, \\
\quad bowler & \mapsto & \{HB, TB\}, \\
\quad fastblr & \mapsto & \{HB, TB, TBF\}, \\
\quad slowblr & \mapsto & \{HB, TB, TBS, IA\}, \\
\quad batsman & \mapsto & \{HB, CB\}, \\
\quad wicketkpr & \mapsto & \{HB, WG\}\}
\end{array}
$$

which would be passed on to *find_taxonomical_relation*. This would call *find_segregate*(*hbiot*) which would return a pair, its first coordinate being *segregate*(*player*, *HB*), because *HB* is the predicate that all terms have in common. The second coordinate returned by *find_segregate* would be *rhbiot* (i.e., the remaining terms from *hbiot*):

$$
\begin{array}{lll}
\multicolumn{3}{l}{rhbiot = \{} \\
\quad bowler & \mapsto & \{TB\}, \\
\quad fastblr & \mapsto & \{TB, TBF\}, \\
\quad slowblr & \mapsto & \{TB, TBS, IA\}, \\
\quad batsman & \mapsto & \{CB\}, \\
\quad wicketkpr & \mapsto & \{WG\}\}
\end{array}
$$

Next, *find_taxonomical_relations* calls *find_contrast_set*(*rhbiot*), which returns a set with three IOTs, split according to the largest *segregate_size*; *TB* is chosen first because it yields a segregate with three terms, and the same is done until there are no more terms in the IOT to be split. This gives the set of IOTs *cshbiot* (which will become a contrast set after these IOTs are also converted to taxonomies) as below:

$$
\begin{array}{l}
\mathit{cshbiot} = \{ \\
\quad \{ \\
\qquad \begin{array}{lll}
\mathit{bowler} & \mapsto & \{\mathit{TB}\}, \\
\mathit{fastblr} & \mapsto & \{\mathit{TB}, \mathit{TBF}\}, \\
\mathit{slowblr} & \mapsto & \{\mathit{TB}, \mathit{TBS}, \mathit{IA}\}\ \},
\end{array} \\
\quad \{ \\
\qquad \begin{array}{lll}
\mathit{batsman} & \mapsto & \{\mathit{CB}\}\ \},
\end{array} \\
\quad \{ \\
\qquad \begin{array}{lll}
\mathit{wicketkpr} & \mapsto & \{\mathit{WG}\}\ \}\ \}
\end{array}
\end{array}
$$

Now *find_taxonomical_relations* is called recursively to each of the three IOTs in *cshbiot* and this leads, by the same process, to a set of taxonomies *txhbiot*. The final result of calling *find_taxonomical_relation*(*hbiot*) is *contrastset*(*segregate*(*player*, *HB*), *txhbiot*), which is the same as Tx_1 in Figures 5.4. A similar process occurs then with *generate_taxonomical_relations*(*csiot*, *WP*), for the taxonomy for field terms, yielding Tx_2, shown in Figure 5.5. The taxonomies are free of mistakes about cricket and therefore more reliable than the reasoning of some foreign observers of cricket.

This concludes our case study. The discussion related to this work can be found in Section 8.2.

5.5 Conclusion

We have presented a new way of specifying ontologies used in societies of agents based on a theory of intensionality, which connects with our anthropological approach to migration of agents. It allows intensional ontologies to be ascribed to societies of agents, hence its importance for this thesis, insofar as it aims at allowing interactions among agents of disparate communities.

Further, inspired by works on ethnography (which supports our point about the usefulness of anthropology as DAI discipline), we have presented a means for an anthropologist agent to recover taxonomical relations[17] from the intensional ontologies it has ascribed to societies of agents, and we have formalised it along with the ascription process. This expanded capability of an anthropologist agent is also a necessary step towards the generation of cultural descriptions of societies of agents. Migrant agents should be able to adapt much more effectively to a target society if they are given access not only to intensional ontologies, but also to taxonomies for (subsets of) the terms that are used in the target's (foreign) culture.

We have also presented an example of our approach using an ontology and taxonomies

---

[17] Recall also that the segregates in the taxonomies in our approach are extended with *atributes* (the predicates), allowing the classification of new objects into those categories.

from a non-trivial ball game in order to illustrate the process given in the formal specifications; we have demonstrated that the process is effective even for the exotic culture of cricket.

The work described in this chapter is an indispensable part in the process of an anthropologist agent creating cultural descriptions of societies of agents, if migrant agents are to be able to use foreign communication languages (which are not merely fixed languages that can be programmed into the design of all relevant agents) properly. In turn, the possbility of ascribing cultural description to MAS is fundamental for the viability of our approach to interoperability.

Chapter 6

The Cricket-Playing Society

This chapter presents a simulation of a simplified version of the game of cricket in which one of the players is a "foreign agent," thus unfamiliar with the game. It therefore needs to learn how to play by observing the other players. This player is relevant to the idea of *migrant agent*, in particular the need to join in the activities of a target society in order to have access to information about it, as discussed by anthropologists; it consequently suggests general knowledge which may be of use for general migrant agents, e.g. (and perhaps surprisingly), simple geometry. It is also intended as a testbed for future work on the idea of learning as legitimate peripheral participation and ontology ascriptions. The simulation is based on very simple geometry and kinematics, and has a simple graphical interface. A Symbolic Game Summary is also generated; it can be used for future studies on an anthropologist agent trying to create a model of the game supporting the behaviour of the players. The learning of the foreign agents is, currently, accomplished using the ID3 algorithm, running on independent threads within the simulator.

6.1 Introduction

In this chapter, we consider a problem that is realistic in English society, and related to the general ideas of the project considered in this thesis, i.e., migration of agents. Cricket, the national English team sport in summer, is played by teams of 11 "agents." Informal games between villages, where a team from one village visits another, occasionally lead to difficulties for the visitors because not all of their 11 players arrive (e.g. as in (Macdonell 1933; chapter 7)). If a team therefore consists of 10 players, and it observes an able-bodied individual arriving as a spectator, this spectator may find himself strongly urged to join the team, which may already be taking part in a game. If the new person happens to be a Brazilian graduate student with no idea of what cricket is, and with limited colloquial English, we have an interesting instance of a migrant agent or "anthropological field work" problem.

Our simulation consists of a representation of the state of play (position of players in the field, their strength and speed, etc.) and a means of generating, displaying and recording what happens between deliveries of the ball. A game consists of a succession of such cycles (presented in detail in Section 6.3). The goal of the *foreign agent* (as a migrant agent or an anthropologist agent is called in this context) is to join in the game, infer regularities and expected behaviour of players, especially players occupying the field positions to which it is assigned, and to come to behave in a way that resembles the behaviour of an ideal *local agent*—i.e., agents that, unlike the nature of the foreign agent, were conceived for that society and are experts in the role—in the same circumstances (detailed in Section 6.4). In the future, the foreign agent should be enabled to develop a model of the game that supports and generates this behaviour, using the Symbolic Game Summary (see Section 6.3.3). It is also intended as a testbed for future work on the conception of new DAI learning algorithms (e.g., based on the idea of legitimate peripheral participation, which was introduced in Section 4.4). The aim of this experimental work is, at this stage, is to investigate traditional AI methods and general knowledge that can be helpful for an anthropologist agent to engage in "participant observation" (defined in Section 4.3), as we shall emphasise again later in the chapter.

The next section comments on the historical and cultural importance of cricket and presents our motivation in choosing cricket for this part of our work. Section 6.3 gives all the details of the functioning of the simulator and Section 6.4 concentrates on the work that has been done so far in relation to the foreign agent. The final section reviews the chapter and discusses this work in the context of the thesis. We postpone any mention of our future intentions connecting cricket and the idea of legitimate peripheral participation (presented in Chapter 4) until Section 9.2.

6.2 The Game of Cricket

Cricket can have some remarkable complications: according to some foreign human observers, the complexity is arbitrary. Brodribb (1952) devotes a whole book to complex, unusual occurrences in first-class cricket and comments on historical changes in the game resulting from them. Although cricket is such a complex game, it is very flexible, in the sense that we can simulate a wide variety of realistic versions of cricket, and therefore tune the simulations to a point where, on the one hand, the game is complex enough to provide a good test of the agent's adaptive behaviour, and on the other hand it is not so simple that it becomes, to borrow an idiom that is quite well known in England, "not cricket." (This expression actually means to play within the laws of cricket but outside the conventions and spirit of the game; we shall mention this idea below.) This flexibility is one of the reasons, alongside its realistic aspect, for the choice of cricket. Further, the game of cricket, being a non-trivial ball game, has not received any attention[1] from the Artificial Intelligence community, which has remarkable accomplishments in regard to other games (e.g., the famous case of the game of chess).

The choice of the game of cricket is also related to its cultural relevance, which is important for the future research we envision using the simulations of cricket as presented in the course of this chapter (we comment on these future directions in Chapter 9.2). This has to do with an anthropologist agent learning general cultural attributes of the *cricket society* from the game itself. There is plenty of support for the fact that cricket has indeed a profound cultural basis. Take, as an example, the bodyline story (Sissons and Stoddart 1984). The English national team, also called Marylebone Cricket Club (MCC) for the non-Test games, won the "Ashes" in the 1932–33 tour to Australia by means of a rather dubious new manner of fast bowling: many thought it *not cricket*, to use the expression we introduced earlier. This new bowling "technique" consisted of the English fast bowlers aiming on or outside the batsmen's leg stump, and considerably short of a good length. Therefore, deliveries came to Australian batsmen fast, high and in line with their bodies, which is obviously dangerous for the batsman, who has to protect himself rather than play properly. In addition, the fieldsmen were arranged in such a way that if the batsman was at all able to play a shot while defending his body, there was a strong chance that he would be dismissed caught.

The repercussions of this conduct were remarkably far-reaching, even to the point of cricket risking the attachment of Australia to the British Empire (or so Sissons and Stoddart claim), instead of having its usual role as the main cultural bridge between the countries of the Empire. At least there is evidence that the strained relations between the "mother country" and the dominion at the time were aggravated by the episode. The social, economical and political tensions of the time had strained relations between specific circles; cricket, however, was "*general currency*" (Sissons and Stoddart 1984). A large number of episodes are mentioned in that source showing further the (overall) cultural importance of cricket, which we do not mention here. The point they make, quite convincingly, is that cricket was the cultural link of the Empire and an instrument of diffusion of the English culture throughout it. Reaffirmingly, Brodribb agrees with Sissons and Stoddart on the role of cricket in the postwar Commonwealth: "Though it [cricket] is English in spirit, it has now become one of the chief links of the Commonwealth." (Brodribb

---

[1] However, our choice of cricket as an area of application is not totally peculiar. In (Wyeth 1997), "robo-cricket" is used to teach mechatronics for first year undergraduate students. That paper presents a completely different use of the game, but worth mentioning here.

1952).

A recent episode of the MCC is also quite revealing about the cultural traditions that cricket involves. The members of the MCC were to vote whether women should now be allowed to join the club, which the members rejected by not recording a high enough percentage of positive votes. A television reporter enquired, at the end of the meeting which made the decision, from an MCC member: "Is this a sensible decision in 1998?" (with emphasis on the year, as we are supposed to be politically correct these days). To which the MCC member wittily (although he probably had no humorous intentions) replied: "This is a sensible decision in any year." And this is why, when we refer to human players of cricket in the text, we unashamedly use pronouns with masculine gender.

This discussion on the importance of cricket in the English culture has no particular implication for the state of the research as presented here. Nevertheless, its arguments are a strong motivation behind the use of cricket. It could be effective for a long-term project aiming at studying: the discovery of cultural aspects of societies of agents from their activities; cultural adaptation and cultural domination (spread) from one society towards others; and numerous other issues relating cricket to the interdisciplinary project with which we are concerned.

The reader who is not familiar with the game should be warned that its description of the simplified version which we present next is far from enough information for acquaintanceship with the real game. Anyone watching a real cricket game is bound to became baffled with several observed behaviours if only the material below is considered. Although it does include some of the general aspects of the game that are omitted below, this quotation in (Brodribb 1952) from some devotee from far overseas (or even Mark Twain—the attribution is not clear) is likely to make things even worse for a observer in that condition, but it is included for its delightfulness:

> Cricket is really quite simple. You have two sides—one out in the field, one in. Each man in the side that's in goes out in turn to have his innings, and when he's out he comes in (or out), and the next man goes in until he's out. Then, when all are out, that ends the innings and the side that's out in the field comes in, and the side that's been in, goes out and tries to get those coming in, out. Sometimes, of course, you get men who are still in and not out. When both sides have been in and out including not outs, that ends the game.

6.3 The Simulation of the Game

This section describes in some detail the functioning of the simulation of the game of cricket that we have implemented. The description of the foreign agent that is part of this simulation is given apart in Section 6.4.

This section is organised as follows. Section 6.3.1 gives the simplified version of cricket that is actually simulated; it also provides details on the main structure of the simulator as well as a briefing on the user interface. Next, in Section 6.3.2, the behaviour of each type of player (namely, bowler, batsman, wicketkeeper, and fieldsman), alongside the "umpiring," is detailed. A symbolic representation of the episodes of the game, which we call Symbolic Game Summary, is discussed in Section 6.3.3 (which also mentions the contents of the text subwindow of the simulator).

The Implementation

The system has been programmed in C and C++. The graphical user interface is built using the XView toolkit[2] (Van Raalte 1993) with some added code using the Xlib (Nye 1993) directly. Altogether, there are some 8,850 lines of code (n.b., this includes the code for graphical interface, and personal libraries for general data structures and the ID3 learning algorithm—the core of the simulation is about 4,900 lines long).

Although the implementation does not use proper agent-oriented programming (given that there was not much support for it easily available at the beginning of the project), it does make use of general agent notions like "persistent goals;" we have notice that being aware of such notions during programming is quite useful.

6.3.1 The Functioning of the Simulator

This section has three parts: Section 6.3.1.1 explains what are the basic rules of the game of cricket as we simulate it; the general structure of the simulator is discussed later in Section 6.3.1.2; then, in Section 6.3.1.3, the instructions on how to use the controls in the graphical interface are summarised.

6.3.1.1 The Simplified Form of Cricket

In the simulation of the game that we have implemented, a number of simplifications were made to the game of cricket. We explain exactly how the game proceeds in this simulation with the help of Figure 6.1 below. It shows the graphical interface to the simulation, which allows one to follow the simulated game in an effortless way.

The *players* are displayed in that image as coloured circles with a black contour. The *ball* is a

---

[2] The initial XView code was set up from changes in a previous interface generated automatically with the DEVGUIDE tool for creating graphical user interfaces, which was not available for this project.

```
Bowling (1.36)
Runs Scored: 0

Bowling (4.42)

Bowling (5.78)
Runs Scored: 1

Bowling (9.81)
Run Out
Runs Scored: 0

Bowling (13.56)
Runs Scored: 0

Bowling (16.26)
Runs Scored: 3

Bowling (28.59)

Bowling (30.14)

Bowling (31.53)
Out Caught
Runs Scored: 0

Bowling (33.23)

Bowling (34.70)

Bowling (36.13)
Runs Scored: 1
```

Figure 6.1: The Graphical Interface to the Cricket Simulator

slightly smaller circle with a white contour; it is coloured either red (of a shade similar to the real cricket ball) when the ball is "in the air" or in a purplish colour when it is "on the ground." This is a simple scheme to show whether the ball is (still) in the air (which is important for the game: see the umpiring of Out Caught in Section 6.3.2), given that the graphical interface is two-dimensional. The *creases* are the white marks in the centre of the green "*field*" delimited by white lines which are the *boundaries*; together, the two opposite creases and the space between them form the *pitch*. The *wickets* are represented by two small rectangles of a wooden colour, each one in one of the creases. We refer to the wicket at the end of the pitch where the batsman/striker initially is (the one to the right-hand side in the display of the simulator) as W1 and the opposite wicket as W2.

The *bowler* is displayed in a dark orange colour. Before the simulation starts, it is holding the ball and its position is within the creases that appear in the left of the pitch. The *batsman* is drawn in medium blue, and its initial position is within the right creases (note that, unlike the real game, there is only one batsman here). All players but the batsman belong to the same "team"—the batsman is thus the "opponent" (one can accordingly refer to the "fielding team" and the "batting team"). The brown circle is the *wicketkeeper*, whose position is behind W1 (to its right in the interface). All the pale circles represent the *fieldsmen*. Finally, the *foreign agent* is the yellow circle (this changes to a pinkish colour under certain circumstances, as explained in Section 6.4).

A cycle of the game starts with the *bowling* (i.e., the bowler "throwing" the ball) and, if the ball is hit by the batsman, it ends either if the wicket is *put down*, also described as "*broken*" (i.e, the ball has reached the wicket; see Section 6.3.2.5) or when the ball is in the possession of the wicketkeeper or, less frequently, of the bowler. If the batsman chooses not to hit the ball (the reason for this is explained later), a new cycle starts immediately. In case the ball is actually hit by the batsman (who, in the real game, carries a *bat*: a wooden club with a handle and one flat surface), the ball is then chased by the fieldsmen (including the wicketkeeper, and the bowler in unusual situations), and the batsman keeps running from one end of the pitch to the other, until it calculates that it is not going to be able to reach the opposite end before the ball is returned to either wicket (again the reasons for this will be made clear later).

There is a special scoring if the ball reaches the boundaries of the field before it is retrieved by a fieldsman (see Section 6.3.2.5). If the ball is retrieved by a fieldsman, it then decides whether to throw the ball directly to the nearest wicket or to the wicketkeeper. The reason is that to "break the wicket" with the ball (in the simulation this means to make the ball reach one of the wickets) while the batsman is outside the creases (i.e., running between the two ends of the pitch) is one of the goals of the fielding team. The fieldsman who has retrieved the ball may also throw it to the wicketkeeper because it may be in better conditions to break the wicket due to its strategic position. If the fieldsmen manage to do it, it is said that the batsman is *out*, and they *take a wicket*. There are other ways of taking wickets (see again Section 6.3.2.5). The batsman, in its turn, aims at *scoring runs* (that is, every time the batsman reaches opposite creases it scores a run).

Before the next cycle of the game starts, the simulation is stopped for approximately two seconds so that the user has a little time to see what has happened at the end of the current cycle before the simulation carries on (unless the ball is *not played*, i.e., not hit by the batsman). The time of the simulation (in seconds) and the scoring (runs and wickets) are shown in the top part of the graphical interface, within the control panel.

The current default *fielding positions* are (the usual ones for a fast bowling/right-handed batsman): two in slip positions[3], gully, deep fine leg, mid-off, leg slip, silly mid-off, silly mid-on. The foreign agent is presently in the mid-on position (it is not an important position, so a "foreigner" is likely to be placed there[4]); further studies with the foreign agent is other positions remain to be performed. Those are the positions used unless a file named **`Cricket.cfg`** is available to be read by the simulator at the beginning of the simulation. (This file must have the same format as the file **`Config.hh`** which is part of the system and provides the default positions during its compilation.)

6.3.1.2 The General Structure of the Simulator

The constants used in the system are all expressed in kilograms, metres or seconds (therefore, forces in newtons). The time interval *ti* between steps of the simulation is 0.01 seconds, but, of course, the state of the game is not displayed on the screen for each of these steps of simulation (Section 6.3.1.3 below explains how to control the frequency of display updates). The time step of the simulation was set experimentally: larger values caused the missing of some events (e.g., it was not possible to check, within reasonable thresholds, whether the ball had reached a certain position, as in one instant it was too far behind and in the next too far ahead). The field is 155 m long and 111 m wide. The measures for the creases and the wickets (including the distance between them) were all taken from (Smith and the Association of Cricket Umpires 1993).

---

[3] Previously, there were three fieldsmen in slip positions. One of them was removed when we introduced the foreign agent elsewhere in the simulation.

[4] In a real-life situation, a more likely position for a foreigner is mid-off. However, placing the foreign agent in the mid-off position would leave no local agent in a symmetrical position to the foreign agent, thus possibly leaving no examples in the game's past episodes for some of the situations in which the foreign agent may find itself.

The simulator encompasses all the players plus the ball (which are the "movable" objects) and a representation of the field (all these are the "displayable" objects; that is, the ones that appear in the graphical interface). At every step of the simulation, the players are given the chance to "reason;" that is, each player in turn decides if any action is required of it at that time (the reasoning of each type of player is detailed in Section 6.3.2). After that, the new positions of all the movable objects are recalculated. This is done with simple kinematics (Halliday, Resnick, and Walker 1993). If a player p_n, at position X_{ip_n}, is running with velocity $\mathbf{v}_{\mathbf{i}p_n}$, its new position X_{p_n} is:

$$X_{p_n} = X_{ip_n} + ti\,\mathbf{v}_{\mathbf{i}p_n}$$

The calculation of the new position of the ball is slightly more complicated. As in the case of the players, the representation of the ball includes its current velocity. However, when a player throws it or the batsman hits it with the bat, the force applied to the ball needs to be converted to a velocity (the mass of the ball considered for this is 0.160 kg). At every step, the new velocity of the ball is calculated considering both gravity and attrition (basically frictional retardation). The acceleration due to gravity $\mathbf{g}$ is represented as the vector $(0, 0, -9.8)$. When the ball is in the air (i.e., the third component of the coordinate giving the present ball position is greater than zero), attrition is 0.0001, otherwise it is 0.0025 (these are constants multiplied by the current velocity of the ball, and were set experimentally). The same equation as used for the movement of the players (shown above) is then used here for the ball.

We comment now on some characteristics that are common to all the players (the properties of particular types of players are left for when each one is presented in Section 6.3.2). When the simulator is started, certain attributes of each player are assigned random values (within certain thresholds). All players have a *maximum speed* which gives the magnitude of the velocity the player employs when running. For the sake of simplicity, whenever a player decides to run, it runs with its maximum speed (which is constant for the same player during the same execution of the simulator). However, the fact that not all players have the same running ability is expressed in the simulation by each player having its own maximum speed. The same applies to the *maximum strength* of each player; that is, players always use their maximum strength when throwing the ball, but the force applied to the ball varies among the players. Further, players have different *heights*, also randomly attributed. The maximum speed of a player is a value between 6 and 7 m/s, the maximum strength between 2 and 2.5 N (i.e., kg.m/s^2), and their heights range between 1.5 and 2 m; all these ranges are inclusive and with a precision of $\frac{1}{10}$ (so that, e.g., a player's maximum speed is one of 11 possible values).

6.3.1.3 The Use of the Graphical Interface

We explain here the use of the buttons and the slide gadget in the control panel in the top part of the interface. Refer back to Figure 6.1 for a visualisation of what is described here. If the executable `Cricket` is called with a parameter, which must be the number of seconds of simulation, the graphical interface is not used. Regardless of whether the interface (interactive mode) is used or not, when the program finishes the results will be in the appropriate files (we comment on each of the three files generated for every run of the simulator in the appropriate sections).

The leftmost large button, initially with a triangular shape drawn inside, starts the simulation. Once the simulation has started, this button has a rectangular shape in it, which then serves to stop the simulation (rather like the symbols in a tape recorder). The next large button (to the right of the sliding gadget), with two rectangular shapes (again using the same convention of tape recorders), allows a temporary pause in the simulation. The pause button is used to stop the simulation when one intends it to be resumed later exactly from the same point where it was stopped (the button changes colour; in that state, pressing it resumes the simulation). Stopping the simulation with the start/stop button freezes the animation of the simulation (so that the state it was in at the time can be seen in the screen), as happens with pausing, but restarting it after that has a rather different effect than pausing. It is as if a new game was started—the time, wicket and runs scored, which are displayed in the top of the interface, are all reset. However, this is not the same as exiting the simulator and starting a new simulation, because the game started after a stop is as another game but with the "same players" (i.e., the characteristics of the players that are randomly set are not changed between games in a single run of the simulator). The rightmost large button, which has the drawing of an open exit door, terminates the execution of the simulator.

The slide gadget can be used to control the speed of the animation as it is seen by the user. It actually changes the number of frames shown per second of simulation (so the slower the simulation seems, the more precise the movements appear on the screen). It can be adjusted to 10 different speeds (specified as the interval in seconds of simulation between the updates of the screen): from $\frac{1}{30}$ to $\frac{10}{30}$ of a second. In the slowest setting ($\frac{1}{30}$ s), the display is updated 30 times per second of simulation. In the fastest setting ($\frac{10}{30}$ s), only three frames are shown for each second of simulation time (we commented in the previous section about the time interval between steps of simulation).

The small button on the very top of the control panel of the simulator window, to the left of the area where the current simulation time is displayed, opens a subwindow where some text is printed in order to make clear some aspects of the simulated game which are not always easily captured in the graphical animation (specially in a fast mode). Whenever a game is stopped, the statistics of that game are printed in that window too (further details are given in a specific item under Section 6.3.3). If the text subwindow is already open, pressing that button only makes the window "come forward" in the window system (i.e., to be displayed in the foreground). To close it, the user should use the normal window controls: the text subwindow behaves like a normal, independent window in the system. It is actually implemented as a (read-only) text editor of the XView toolkit, therefore allowing the saving of the text that is displayed on it, if the user wishes to do so (e.g., to keep the history, or parts of it, and the final scores of a particularly interesting game), even though the complete contents of the text window are also saved in one of the simulation files, as we explain later. The user needs to click with the right button of the mouse inside the text panel in order to have access to the text editor controls.

6.3.2 The Reasoning of the Players

We now consider each of the types of players in turn, presenting the details concerning the issues related particularly to each class of player: the bowler, the batsman, the fieldsmen, and finally the wicketkeeper. Then the issues relating to the "umpiring" of the game are presented (i.e., the decisions about scoring and whether the next cycle is due).

The algorithms described here are included in Appendix C as C++ programs but of very high level. The reasoning part of the object interface for each type of player has been programmed as a series of calls to high-level functions, with suggestive names, which makes that particular piece of code quite readable, almost like a pseudocode program. The reader can refer to that Appendix for details; we provide here only textual descriptions. The implementation of each of the high-level functions presented there are not included in the Appendix nor discussed at great length in the thesis. When a certain algorithm used in the implementation of one of those high-level functions is particularly important, it is presented in this section, as in the case of how the batsman calculates how long it will take before the ball returns to the wicket, which is needed to decide if another run may be practicable; the other algorithm given here refers to the choice of the target where to throw the ball after a fieldsman has retrieved the ball. (Note that this type of information is not in the Appendix; only the general description of the behaviour of each type of player is given there.)

6.3.2.1 The Bowler

The main action of the *bowler* is, obviously enough, the bowling (which starts a new cycle of the game). Because of the the special way in which real bowlers project the ball (*to "bowl"* is to throw the ball towards the wicket with a straight arm making a relatively small angle with the vertical above one's head), the force applied to the ball by a bowler is considered to be 15% greater than the maximum strength originally attributed to the bowler as a "player" (more precisely, all types of players in the game inherit general attributes from a base class **Player**). However, the force the bowler applies to each bowled ball has a random reduction of not more than 10% in the value of the final bowler's strength; this denotes the inconstancy of consecutive bowlings in real life. For this same purpose, the coordinate for the target (W1) also suffers some random variation of a maximum of 0.3 m in all directions (i.e., in the three components of the coordinate, either positive or negative).

After bowling, the bowler seldom has any further actions in a cycle of the game. However, the bowler considers chasing the ball if it is passing very close (less than 7 m) to its position and the batsman is out of its ground (see description of the batsman below); that is, there is a possibility of achieving a Run Out (see Section 6.3.2.5 below). If a fieldsman "grabs" the ball and the batsman is not out of its ground (i.e., not attempting runnings anymore), it throws the ball to the bowler (unless the wicketkeeper is nearer to that fieldsman), so another possible action for the bowler is to receive the ball in these situations. Once the ball is "safely" in the possession of one of these two special players, a new cycle is started.

6.3.2.2 The Batsman

The goals of the *batsman* are basically two:

protect the wicket: the batsman hits the ball with the bat if it calculates that the bowled ball is in a collision course with the wicket; and

score runs: if the batsman has hit the ball that had been thrown by the bowler, it may start "running" as explained before.

However, the batsman must be careful before running to avoid being out as explained earlier (in the real game, this batsman would be dismissed and his team would have to send another player in to bat; we do not discuss further such complications of the real game). In this version of the game, when the batsman is "sent out" or "dismissed" (for whatever reason[5]) we

---

[5] The other forms of scoring (for both the batsman and the fielding team) are presented below in the discussion on umpiring, Section 6.3.2.5.

say that the fielding team has taken a wicket, but otherwise it is as if the same batsman continued playing. This is to say that the characteristics of the batsman as a player are not changed (as we stated earlier, this only happens when the program is run again).

A (human) batsman is considered "out of his ground" unless any part of his body (or the "bat" that he holds) is inside the creases at either end of the pitch (more precisely, behind the line of the popping crease; see (Smith and the Association of Cricket Umpires 1993; Oslear and Mosey 1993)). To determine whether the batsman is *out of its ground* in the simulation, we check whether the batsman's position, considering also the length of the bat, is inside the creases (the length of the bat is **0.965** m). If the batsman is out of its ground and the wicket is broken, the batsman is out—this is better explained in the Section 6.3.2.5, which is about umpiring. (Note that in this simplification, as we have only one batsman, both wickets are "safe" if the batsman is inside the creases at either end of the pitch).

The batsman has, besides its characteristics as a player, one more parameter which is attributed to it randomly when the simulator is started. This is its *accuracy*: an integer in the range of **5** to **10** (inclusive). This value is used to determine the amount of variations (i.e., errors) used in the batsman's calculations[6] that are needed when it has to decide whether to run, which we present later. As in the case of the bowler, the batsman also applies a force to the ball which is higher than for the normal player: this is because, in the case of a human batsman in the real game, he hits the ball with the bat, so striking it with much more strength than a simple throw. Thus, for the "computational" batsman, we multiply its randomly-assigned strength, i.e., the one it inherits from the class common to all players, by **3** (a value that is also set experimentally).

In what follows, when we need to refer to the different parts of any vector $\mathbf{x}$, we represent its magnitude or length by x and its direction or orientation by the unit vector $\mathbf{X}$. The notation $|\mathbf{x}|$ is the magnitude of $\mathbf{x}$.

When a bowled ball is about to be hit, that is, the batsman reckons that the bowler was accurate enough and the ball is in a trajectory passing close enough to the wicket, the batsman prepares to hit it by moving its bat. In the simulation, a random position for the bat is set at the time of collision with the ball (we do not consider any further tactics of batting that real batsmen use). The bat is represented by a random unit vector $\mathbf{N}$ normal to the flat surface. This indicates the direction of motion of the bat. The magnitude of the vector giving the velocity of the bat is the batsman's strength, subject to a **20%** random variation (similar to what happens

---

[6] Note that the accuracy determines only how well the batsman predicts other player's moves in the game: a high accuracy batsman does not necessarily mean a high run-scoring batsman, as this also depends on its maximum speed and strength. It also depends on the opportunities that the batsman has for running, which are determined by particular traits of the other players and, for short simulations, by some other random variables used at several moments of the generation of simulation cycles.

to the strength of a bowler). When the ball collides with the bat, its new velocity is calculated as explained below. The real situation is complicated because of facts beyond the reach of simple physics, e.g., the details of conversion of kinetic energy into other forms of energy when the ball hits the bat, and the reality that the bat is not a free body but is attached to the batsman, whose input to the physics at the moment of hitting the ball is basically inaccessible to us. We have therefore simplified the treatment, checking that the results from the cricket simulator are nevertheless kinematically convincing and realistic.

Let m_1 be the mass and $\mathbf{v_1}$ the initial velocity of the ball (the velocity of the bowled ball at the time of the collision with the bat). Let m_2 be the mass and $\mathbf{v_2}$ the initial velocity of the bat (determined as explained above). The final velocity of the ball is represented by $\mathbf{v}_{1f}$. Then $\mathbf{V}_{1f}$ is as follows (where "^" is used to take the corresponding unit vector):

$$\mathbf{V}_{1f} = \widehat{\mathbf{i} + 2\,\mathbf{N}}$$

where

$$\mathbf{i} = \frac{\mathbf{v_1}}{|\,\mathbf{v_1} \cdot \mathbf{N}\,|}$$

These equations are based on the ones used, e.g., in the well-known rendering technique of *raytracing* in computer graphics (Foley *et al.* 1990); they serve our purpose adequately, as noted above, even though a real cricket ball is not a ray.

Now, as for the magnitude of $\mathbf{v}_{1f}$, we first need to calculate the velocity $\mathbf{w}$ of the centre-of-mass system (with respect to the stationary coordinate system of the collision of the ball and bat):

$$\mathbf{w} = \frac{m_1\,\mathbf{v_1} + m_2\,\mathbf{v_2}}{m_1 + m_2}$$

To an observer in the system, the (magnitude of the) apparent velocity of the ball before it is hit by the bat is $\mathrm{u}_1 = |\,\mathbf{v_1} - \mathbf{w}\,|$ and after being hit is $\mathrm{u}_{1f} = Cr_C\,\mathrm{u}_1$, where Cr_C is the coefficient of restitution of a cricket ball, which we consider[7] to be **0.6**. Any velocity in the centre-of-mass system can be converted to the stationary system by the inverse process, i.e., by adding $\mathbf{w}$. Thus the velocity of the ball after being hit ($\mathbf{v}_{1f}$) is:

$$\mathbf{v}_{1f} = \mathbf{w} + \mathbf{u}_{1f}$$

Note that taking both parts of each of these vectors, the only missing values are v_{1f} and $\mathbf{U}_{1f}$. Because the square of a unit vector is 1, and

---

[7] We have not been able to find a precise value in the literature. However, we are aware of some informal experiments with a real cricket ball that have been performed and have led to that number (after some rounding).

$$\mathbf{U}_{1f} = \frac{\mathrm{v}_{1f}\,\mathbf{V}_{1f} - \mathrm{w}\,\mathbf{W}}{\mathrm{u}_{1f}}$$

we know that

$$\mathrm{u}_{1f}{}^2 = \mathrm{v}_{1f}{}^2 + \mathrm{w}^2 - 2\,\mathrm{v}_{1f}\,\mathrm{w}\,(\mathbf{V}_{1f} \cdot \mathbf{W})$$

We therefore take the quadratic equation where $a = 1$, $b = -2\,\mathrm{w}\,(\mathbf{V}_{1f} \cdot \mathbf{W})$, and $c = \mathrm{w}^2 - \mathrm{u}_{1f}{}^2$. We then use one of the roots of that quadratic equation as v_{1f} (that is, as the magnitude of $\mathbf{v}_{1f}$, which is the final velocity of the ball after collision, for which, remember, we already know the direction). If there is only one positive root, that one is chosen. Otherwise, we choose the greater one (in absolute value) in an average of 30% of the relevant instances (randomly) and the lesser one at other times.

As we described earlier, after batting, the batsman needs to decide whether to "run". We turn to this decision-making of the batsman now. In summary, if a player is chasing the ball, the batsman needs to calculate the time it will take for the player to catch the ball and throw it to the nearest wicket (which would be, in perfect conditions, the fastest way of breaking the wicket); if the ball is on its way from a fieldsman to the wicketkeeper, than a similar calculation must be made considering the time for the wicketkeeper to receive the ball and return it to W1; if the ball is already travelling in the direction of a wicket, the batsman needs to consider how long it will take to reach the target. After a calculation of this time is complete (with random errors added for realism), the decision of the batsman is simple: all it has to do is check whether this time is greater than the time the batsman takes to run from one wicket to the other. That calculation of the time for the ball to reach the wicket explained informally above is clearly presented in Algorithm 6.1.

In the algorithm, **ChasingBall**(f) returns in f the fieldsman that is chasing the ball, and **RetrievePos**(f, b) is the position where fieldsman f will retrieve the ball (whose relevant information—in this case position and direction—is given by b). Because of the behaviour of the fieldsmen, first it is necessary to consider the time it will take for the player to reach the nearest point to its current position that is on the current trajectory of the ball (how we calculate this is presented in the next section, when we discuss the fieldsmen), unless, of course, the fieldsman is already in line with the trajectory of the ball. Then it is considered how long the player will be able to run towards the ball from that point on the ball's trajectory until they meet (except, of course, if the fieldsman will miss the ball). After this position r that is suitable for retrieval of the ball is discovered, calculating how long for the retrieval to occur (**TimeToRetrieve**(r)) is trivial.

```
if ChasingBall($f$) then
    $r$ = RetrievePos($f$, $b$);
    $t$ = TimeToRetrieve($r$) + TimeToReturn($r$, $w$);
    (where $f$ is the fieldsman chasing the ball,
     $b$ is the relevant information about the ball,
     and $w$ is the nearest wicket)
if BallToWicketKeeper() then
    $t$ = TimeToReceive($wk$, $b$) + TimeToReturn($wk$, $w$);
    (where $wk$ is the position of the wicketkeeper)
if BallToWicket() then
    $t$ = TimeToReach($w$, $b$);
$t$ = AddVar($t$);
```

Algorithm 6.1: Batsman's Calculation of the Time for the Ball to Reach the Wicket

That calculated time plus the time for the player to throw the ball from r to the nearest wicket w (TimeToReturn(r, w)) gives t, the time before the wicket can be broken.

In case the ball is already travelling in the direction of the wicket, t is the remaining time before the ball reaches that wicket (TimeToReach(w, b)). If the player has thrown the ball to the wicketkeeper (to make it more likely that the wicket will be broken), then t is the time it will take for the wicketkeeper (who is in position wk) to receive that ball (TimeToReceive(wk, b)) plus the time for the ball to be thrown back to wicket w and reach it (TimeToReturn(wk, w)).

The calculated time t then needs to have some variation added to it (in the Algorithm expressed as $t =$ AddVar(t);). Notice that in order to calculate t the batsman needs to use data (e.g., the position and velocity of the ball, the strength of the player that is throwing the ball) which it cannot know precisely, if the simulation is to be realistic; hence the added variation. This variation (either positive or negative) is of a random percentage not greater than $2 + 2(10 - A_b)$ where A_b is the accuracy of the batsman (recall that its range is the interval from 5 to 10).

6.3.2.3 The Fieldsmen

The main function of the *fieldsmen* is to chase the ball after it has been batted. They "coordinate" their action by checking which of the fielders is in the best position to chase the ball. In the usual jargon of MAS, the players here are not "autonomous:" they do not refuse to cooperate neither do they negotiate to decide who will do the chasing. Simply, the fieldsman who can retrieve a particular ball most efficiently (i.e., in the least time) is assumed to be the one responsible for doing the job, and indeed does it. When a ball is batted, every fieldsman, during their turn to reason, calculates whether there is another fieldsman (for those who are in the direction of the batted ball, of course) who would reach the ball in less time than itself. If there is, then this fieldsman just "rests." Very slow balls are always dealt with by the wicketkeeper (which has a different reasoning from the fieldsman, see Section 6.3.2.4 below). The behaviour of the fieldsman who cannot find anyone better than itself to retrieve that ball is explained next.

The fieldsman that is to chase the ball first moves straight to the point in the trajectory of the ball that is nearest to its initial position (we explain how this point is calculated below). This is to make sure the fieldsman will be there to intercept the ball when it passes if that is at all possible (simply running repeatedly towards the current position of the ball from the start made the fieldsmen unnecessarily miss the ball at times). Once in that point, the fieldsman starts running towards the ball if the ball is still in its direction. If the fieldsman missed the ball then all fieldsmen, when the next opportunity to reason arrives, consider which one of them is now the best fieldsman to back up for the one who missed the ball. This decision is simply taken by considering the fieldsman that is closest to the ball at that moment, rather than calculating the time it will take to reach the ball as in the previous situation (which is quite reasonable most of the time in this context of backing up). Also, if it happens that at some point another fieldsman can do better than the present backing-up fieldsman (i.e., the former is now closer to the ball), the back-up decision procedure of the fieldsmen is started again. This same backing up decision process also happens when a fielded ball will not reach its target. If the fielding was not enough, that is, the fieldsmen realise that the ball is currently moving too slowly (i.e., less than 6 m/s) so it will not arrive at its target (recall the effect of attrition, and the limited strength of each fieldsman), the fieldsmen perform the backing-up decision-making again so that another fieldsman is chosen to run to the ball, retrieve it, and throw it towards its target again.

In order to calculate the *closest retrieval point* C, which is the position towards which, as explained above, the fieldsman initially runs, we use the following formulæ, based on basic geometry (Thomas and Finney 1993). In the formulæ, B is the position of the ball, $\mathbf{v}$ is the vector giving the direction of the ball's trajectory, F the position of the fielder, and $\mathbf{bf}$ is the vector from B to F. We first consider the distance a from F to the line given by B and $\mathbf{v}$:

$$a = \frac{|\,\mathbf{bf} \times \mathbf{v}\,|}{|\,\mathbf{v}\,|}$$

and we can then consider the triangle given by B, F and C (the point we are trying to calculate), of which a is the length of one side, and the hypotenuse h is simply the distance from B to F. The distance d from B to C is therefore $d = \sqrt{h^2 - a^2}$. Now we can determine the coordinates of C as follows (below, p_x means the x coordinate of p):

$$C = (B_x + t\,\mathbf{v}_x,\ B_y + t\,\mathbf{v}_y,\ B_z + t\,\mathbf{v}_z)$$

where

$$t = \sqrt{\frac{d^2}{\mathbf{v}_x^2 + \mathbf{v}_y^2 + \mathbf{v}_z^2}}$$

If a fieldsman is able to retrieve the ball (i.e., it arrives at C before the ball is past that point), it must then decide what to do with it. Depending on the conditions, a fieldsman throws the ball to one of the wickets (in an attempt to break the wicket whilst the batsman is out of its ground), or to the wicketkeeper, or even to the bowler (in unusual cases). This is concisely expressed in Algorithm 6.2. (This algorithm is part of the fieldsmen's reasoning included in Appendix C.)

```
if BatsmanOutOfGround() then
    if IsNearer($w$, $wk$) $\wedge$ $\neg$TooFarFromPitch()
        $bt = w$;
        (where $wk$ is the wicketkeeper's position
         and $w$ is that of the nearest wicket)
    else
        $bt = wk$;
else
    $bt$ = Nearest($wk$, $bl$);
    (where $bl$ is the bowler's position)
ThrowBall($bt$)
```

Algorithm 6.2: Fieldsman's Choice of the Best Target for Throwing the Ball

In the algorithm, when it is used in the reasoning of a fieldsman f, **IsNearer**(p_1, p_2) is true if and only if point p_1 is closer to the location of the fieldsman f than p_2, and **Nearest**(p_1, p_2) returns p_1 if the distance from p_1 to f is less than or equal to that of p_2 to f, otherwise it returns p_2. The predicate **BatsmanOutOfGround()** speaks for itself and **TooFarFromPitch()** is true if and only if fieldsman f is positioned at more than half the distance between the pitch and the boundaries of the field at the moment the ball was reached. In this situation, there is no point in a fieldsman trying to break a wicket, as it is too far from it to have good chances of hitting it, so the best target is the wicketkeeper. If fieldsman f has stopped the ball at a point near the wicket and there is a chance of scoring a wicket (the batsman is out of its ground), the best target will be either the nearest wicket or the wicketkeeper, whichever is closer to f (this is a gross simplification of the real game, of course). If there is no chance of scoring, then fieldsman f throws the ball either to the wicketkeeper or the bowler, whichever is nearer to f, so that the next cycle can start. After the variable for the best target bt is set in that way, all the fielder has to do is throw the ball to that position (ThrowBall(bt)).

6.3.2.4 The Wicketkeeper

The *wicketkeeper* is a fieldsman with several special attributions. Its initial position is behind the wicket (at **3.5** m from it), and it runs towards the wicket after the batsman has hit the ball so that it is immediately behind the wicket (at **0.5** m from it). This is a strategic position, for it allows the wicketkeeper to break the wicket easily after receiving the ball from a fieldsman that has retrieved it. The wicketkeeper here is also special in the sense that the cycle only finishes when the ball is its possession (or of the bowler), unless the wicket is put down (the wicket being put down does not mean that a wicket is taken by the fielding team; see further details below, in Section 6.3.2.5).

Because it is important that the wicketkeeper stays in that strategic position, the wicketkeeper is not considered when the fieldsmen decide who is the best one to back up. Also, the wicketkeeper is only considered to be the one initially in charge of chasing a ball if it does not have to move too far (more than **2** m) to be in the closest retrieval point; however, balls that are too slow (less than **2** m/s after the collision with the bat) are always dealt with by the wicketkeeper. If the wicketkeeper finds itself outside an area of **5** m from its initial position when pursuing the ball, it stops, returns to its standing position, and let the fieldsmen do the chasing. The wicketkeeper only runs towards the ball in order to receive it if the ball is already closer than **7** m from the wicketkeeper's position and the batsman is running (i.e., there is still a chance to score); this too is to assure that the wicketkeeper will not be too far from the strategic position behind the wicket.

One last thing to be mentioned about the wicketkeeper is its decision to carry the ball and run to break the wicket whilst holding the ball, rather than throwing the ball for that purpose. This happens only if the batsman has left its ground and is at least **0.5** m away from the creases and the wicketkeeper is not near enough the wicket (i.e., more than **1.5** m away from the wicket) for a throw to be accurate enough to break the wicket.

6.3.2.5 The Umpiring

The *umpiring* function is called at every simulation step after all players have had the chance to "reason" and all movable objects (players and the ball) have had their positions updated. The possible events of the game that must be determined here are listed below, and we describe the conditions under which each one is to happen:

A Run is Scored: if after batting the batsman has run from one end of the pitch to the other (more accurately, if its bat has reached behind a popping crease) and the ball is still being played (i.e., it is not *dead*, see below), the batsman *scores a run.*

The Ball is Dead: the ball is said to be *dead* when: (i) a bowled ball is not hit by the batsman (we shall refer to this case as a ball *not played*), (ii) a played ball is already in the possession of the wicketkeeper or the bowler, (iii) the batsman is "out" (see below the various forms of being sent out), (iv) the ball crosses a boundary (see Four Runs below), (v) the wicket is put down[8]; a new cycle of the game starts thereafter.

Four Runs: if the ball reaches the boundaries of the field, i.e., no fieldsman has managed to retrieve it, the batsman's scoring during this cycle is at least *four* points (of course it can be more than four runs scored[9] if the batsman had already scored more than that before the ball was dead).

The Batsman is Out: the batsman can be *out* (the fielding team *takes a wicket*) in the following ways[10]:

Caught: if the after being hit by the batsman the ball is caught by a fieldsman without having touched the ground (this is simple in the simulation: suffices it to check whether the third coordinate of the position of the ball is greater than zero[11]) the batsman is out *Caught.*

Bowled: if the ball hits the wicket coming directly from the bowler (i.e., the batsman did not manage to protect the wicket, which is its main function) the batsman is said to be out *Bowled.* In this simulation, the ball is always given a new direction after reaching the batsman's position (unless that is a ball not played, as explained above). We therefore consider Bowled Out if after being given this new direction (which represents the batting, but remember that there is a random positioning of the bat) the ball hits the wicket.

Run Out: if the wicket is put down (i.e., the ball hits it) and the batsman is out of its ground (as explained earlier), the batsman is out *Run Out.*

---

[8] In our simplified version of cricket, the ball is dead if the wicket is put down, no matter if a batsman is out or not. For example, if a fieldsman attempts to dismiss the batsman by Run Out but the batsman is in its ground when the ball reaches the wicket, that ball is considered dead (even though the fielding team does not succeed in dismissing the batsman).

[9] Law 19 (Boundaries) says "The runs completed at the instant the ball reaches the boundary shall count if they exceed the boundary allowance." (Smith and the Association of Cricket Umpires 1993).

[10] Note that all these definitions have been simplified for the computational version we are discussing here. To understand precisely what these terms mean in real cricket, one should refer to another source, e.g., (Oslear and Mosey 1993; Smith and the Association of Cricket Umpires 1993).

[11] Note that, as in this very simplified model of a ball, it does not bounce. Therefore, if at some point after being hit the third coordinate of the ball's position is greater than zero, that means that it *has always been so* since the time it was hit, which is a defining condition for dismissing the batsman "caught" in the real game.

Stumped: if the wicket is put down by the wicketkeeper and the batsman is out of its ground (as in the Run Out case), and the wicketkeeper broke the wicket without the intervention of any other fieldsman, the batsman is out *Stumped* rather than Run Out.

Leg Before Wicket (LBW): we have created a simplified mechanism to allow this category of a batsman being sent out to be included in the simulation (given that it is natural only for "non-computational batsmen"). We have simply determined that at random times, when the bowled ball reaches the batsman, it would be considered that the batsman protected the wicket in an analogous way to a real batsman protecting the wicket by intentionally using his legs rather than the bat, which in real cricket is only allowed under conditions (often disputed, and historically quite often changed in the laws) when the ball would otherwise (probably) not have hit the wicket. On average, in 1% of the times at which the ball is batted (chosen randomly) the batsman is considered out *Leg Before Wicket.*

After all the above "umpiring" events apart from the scoring of runs, a new cycle of the game is started.

6.3.3 The Symbolic Game Summary

The Symbolic Game Summary (SGS) is a symbolic representation of the episodes of the game, in the way that would be *observed* by a *foreign agent* (FA). Below we include parts of the SGS generated by a run of the simulation and we comment on the main aspects of the symbolic representation of the game as they appear in those sample events.

The SGS is actually kept in a data structure, but every time the simulator is run, a file `CricketSGS<time>.txt` is created (where `<time>` is the number with the internal representation of the date/time in UNIX, so that a different file is created every time the simulation is run). This file contains a textual representation of the SGS data structure, which is basically a list of *episodes.* An episode of the game is the set of *actions* (i.e., events) that happen simultaneously. We have also introduced a representation of expressions that the players may show (e.g., by means of gestures in the real game, which are arguably universally understandable). In two cases the foreign agent communicates a "sentence" to the captain, but the intention here was to make it clear, in the reading of the SGS, what was going on at that time. Although this ability to express feelings and utter sentences (which may include actions that the receiver agent is expected to take) is so far of use only for the reading of the SGS, it is quite important for future extensions.

We now proceed with an abridged SGS (set in a different type) of one arbitrary simulation, with added comments. These are in a different typeface (*slanted*), and we use "..." to mark the abridgements. Before the episodes of the game(s) are presented in the SGS file, there is a "header" containing symbolic representations of the objects and players of the game:

```
Objects used in the game:

Object 1(ball): leathery, round, small.
Object 2(bat): wooden, flat, large.
Object 3(wicket): wooden, fixed, medium.
Object 4(boundaries): white, fixed, long line.
Object 5(crease): white, fixed, short line.

Players present in the game:
...
```

These are the objects that the FA perceives when joining the cricket society; obviously, the FA is not aware of the names of the objects used in the game. Therefore, they are numbered for their reference in the sequence of episodes that form the games in the SGSs. Clearly, the FA must be able to distinguish the objects, e.g. by means of properties of each one (the properties above are just examples that we inserted in our symbolic structure of the game). This resembles the intensional representations used in Chapter 5. We have provided the real name of the objects in brackets to make it clear for the reader what objects are being referenced.

Similarly to the objects, the FA (or the anthropologist agent who generates the SGS) numbers all the players of the game (including itself). The players are numbered in the following order: bowler, batsman, wicketkeeper, first slip, second slip, gully, deep fine leg, mid-off, leg slip, silly mid-off, silly mid-on, and the FA itself is player number 12. Player number 9, in position leg-slip, is the captain of the fielding team. Below, we comment on a few cycles (or parts of them) of the game so as to show what sorts of episodes (composed of a sequence of simultaneous actions) and what different types of actions can be represented in the SGS[12]:

```
Episode at time 4.37:
Throw - player: 1, threw object: 1, object velocity: V( 16, -1.1, -0.82 ).

Episode at time 5.67:
Throw - player: 1, threw object: 1, object velocity: V( 16, -1.1, -0.77 ).
...
```

---

[12] Note that we have used the word "collect" to refer to the action, perceived by the FA, of a player taking possession of the ball, in whatever circumstances (that is, both for retrieving a ball and for a catch). This corresponds exactly to Schank's *GRASP* primitive act of Conceptual Dependency (1975).

The bowling at 4.37 (i.e., player 1 performing action "Throw" with object 1) resulted in a ball not played (the batsman considered that the wicket was not in danger of being hit by that ball): notice that there is another bowling act (denoting a new cycle) immediately after that one.

```
...

Episode at time 24.24:
Throw - player: 1, threw object: 1, object velocity: V( 15, -0.9, -0.62 ).

Episode at time 25.48:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 9.8, 2.3, -2.1 ).
Run - player: 3, at: P( 30, 0, 1.8 ), velocity: V( -6.1, 0, 0 ).
Run - player: 4, at: P( 40, 5, 2 ), velocity: V( -1.5, 6.2, 0 ).

Episode at time 25.49:
Run - player: 2, at: P( 9.1, 0, 1.7 ), velocity: V( -7, 0, 0 ).

Episode at time 27.69:
Collect - player: 4, at: P( 28, 4.4, 2 ), collected object: 1.
Throw - player: 4, threw object: 1, object velocity: V( -13, -5.1, -0.23 ).

Episode at time 27.84:
Run - player: 3, at: P( 16, 0, 1.8 ), velocity: V( 5.7, 2.3, 0 ).

Episode at time 27.9:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 28.5:
Collect - player: 3, at: P( 17, 0.16, 1.8 ), collected object: 1.
```

The batsman scored one run and was inside the creases (it had stopped running) before the ball returned to the wicketkeeper (that is, the fielding team was not successful in dismissing the batsman); an ordinary cycle of the game.

```
...

Episode at time 56.17:
Throw - player: 1, threw object: 1, object velocity: V( 15, -0.81, -0.41 ).

Episode at time 57.44:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 10, -1.8, 2 ).
Run - player: 3, at: P( 30, 0, 1.8 ), velocity: V( -6.1, 0, 0 ).
Run - player: 9, at: P( 25, -10, 1.9 ), velocity: V( 1.2, 6.8, 0 ).

Episode at time 57.45:
Run - player: 2, at: P( 9.1, 0, 1.7 ), velocity: V( -7, 0, 0 ).

Episode at time 58.85:
Collect - player: 9, at: P( 23, -2.5, 1.9 ), collected object: 1.
Throw - player: 9, threw object: 1, object velocity: V( -7.2, 10, -0.4 ).

Episode at time 59:
Collect - player: 3, at: P( 22, -0.7, 1.8 ), collected object: 1.
Throw - player: 3, threw object: 1, object velocity: V( -16, 0.92, -1.4 ).

Episode at time 59.76:
```

```
Meet - object: 1, met object: 3.
Express - player: 3, is: cheerful.
Express - player: 2, is: disappointed.
```

This complete cycle ended with a Run Out, and no runs scored by the batsman.

Before we continue with more episodes, we make some comments on the other ways of dismissing the batsman, which will not be demonstrated in the SGS for the sake of space. A Stumped dismissal is similar to the one above except that the wicketkeeper would be the only fieldsman involved; an Out Caught ends with a "Collect" action performed by a fieldsman immediately after the batting—the cheerful/disappointed gestures confirm that for a reader of the SGS file; Out Leg Before Wicket ends soon after a bowling act, the batsman being disappointed; in an Out Bowled the ball hits the wicket immediately after the batting (recall this peculiarity of the simulated game).

```
...

Episode at time 475.3:
Throw - player: 1, threw object: 1, object velocity: V( 15, -0.96, -0.66 ).

Episode at time 476.58:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( -0.43, 12, 7 ).
Run - player: 3, at: P( 30, 0, 1.8 ), velocity: V( -6.1, 0, 0 ).
Run - player: 10, at: P( 5, 4, 1.8 ), velocity: V( 6.6, 0.24, 0 ).

Episode at time 476.59:
Run - player: 2, at: P( 9.1, 0, 1.7 ), velocity: V( -7, 0, 0 ).

Episode at time 476.94:
Run - player: 10, at: P( 7.3, 4.1, 1.8 ), velocity: V( 6.5, 0.84, 0 ).

Episode at time 479:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 479.01:
Run - player: 2, at: P( -7.9, 0, 1.7 ), velocity: V( 7, 0, 0 ).

Episode at time 481.26:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 481.27:
Run - player: 2, at: P( 7.9, 0, 1.7 ), velocity: V( -7, 0, 0 ).

Episode at time 483.52:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 483.53:
Run - player: 2, at: P( -7.9, 0, 1.7 ), velocity: V( 7, 0, 0 ).
```

```
Episode at time 483.91:
Meet - object: 1, met object: 4.
```

This cycle resulted in Four Runs (the ball reached a boundary); the batsman had already run successfully for 3 runs.

...

We now concentrate on the actions represented in the SGS that are specific to the behaviours of the FA. There are four types of mistakes that the FA can make:

1. Not to know whether it should chase the ball (a *Failure* node in the decision tree generated by the learning algorithm, see below in Section 6.4); if it actually was the FA's responsibility to chase the ball, another fieldsman will back the FA up.
2. Not to know where to throw the ball (again due to *Failure* in the decision tree); in this case the FA holds the ball and another fieldsman must come to the FA's position to get the ball and carry on with the game (the fieldsmen enter the "backing-up" mode).
3. Not to chase the ball when the fieldsmen have calculated that the FA should have done so (recall that, in this simulation, the fieldsmen have a perfect calculation of who is the best fieldsman to chase the ball) and therefore no fieldsman is chasing the ball; when the fieldsmen realise that this is happening they enter the backing-up mode.
4. Start chasing the ball when it did not need to do so; if the FA realises that another fieldsman is chasing the ball at the same time as itself, it immediately stops running[13]. A FA would reason: "surely the local agent knows better who is supposed to chase the ball."

These misbehaviours of the FA seldom occur, except the one in which the FA tries to chase a ball for which another fieldsman is responsible (item 4). However, it is important to consider what is recorded in the SGS in these cases. They are illustrated below (in the same order as above):

```
Episode at time 523.82:
Throw - player: 1, threw object: 1, object velocity: V( 16, -0.74, -0.67 ).

Episode at time 524.99:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 0.38, -0.13, 0.049 ).
Run - player: 3, at: P( 30, 0, 1.8 ), velocity: V( -6.1, 0, 0 ).
Say - player: 12, says to player: 9, the message: "I don't know if
      I should chase the ball!".
Express - player: 9, is: angry.
Express - player: 12, is: sorry.
```

---

[13] The initial movement of the FA towards a ball that it is not supposed to chase is hardly noticeable in the interface, but the change of colour, as we see in Section 6.4, allows the user to realise that such is the case.

```
Episode at time 525.2:
Run - player: 11, at: P( 5, -4, 1.7 ), velocity: V( 5, 4.8, 0 ).

Episode at time 526.21:
Collect - player: 11, at: P( 9.5, -0.2, 1.7 ), collected object: 1.
Throw - player: 11, threw object: 1, object velocity: V( -14, 0.89, -0.15 ).

Episode at time 527.61:
Collect - player: 1, at: P( -9.1, 1, 1.5 ), collected object: 1.
```

The FA does not know if it should chase the ball; in this case, the fieldsmen expected it to do so (no other fieldsman ran for the ball); therefore, the captain is angry and the FA expresses sorriness for the mistake[14] *(i.e., not running); later on, player 11 backs up for the FA (it gets the ball and throws it to the bowler; the batsman had not left its ground, so there is no scoring).*

```
...

Episode at time 570.95:
Throw - player: 1, threw object: 1, object velocity: V( 16, -0.89, -0.71 ).

Episode at time 572.09:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 3.2, -12, 7.4 ).
Run - player: 3, at: P( 30, 0, 1.9 ), velocity: V( -6.2, 0, 0 ).
Run - player: 12, at: P( 15, -15, 1.8 ), velocity: V( -2.5, 6.2, 0 ).

Episode at time 572.9:
Run - player: 2, at: P( 9.1, 0, 1.5 ), velocity: V( -6.9, 0, 0 ).

Episode at time 575.35:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 575.36:
Run - player: 2, at: P( -7.9, 0, 1.5 ), velocity: V( 6.9, 0, 0 ).

Episode at time 577.64:
Reach - player: 2, reached object: 5.
Express - player: 2, is: cheerful.

Episode at time 577.65:
Run - player: 2, at: P( 7.9, 0, 1.5 ), velocity: V( -6.9, 0, 0 ).

Episode at time 579.9:
Collect - player: 12, at: P( 24, -55, 1.8 ), collected object: 1.
Say - player: 12, says to player: 9, the message: "I don't know what
      to do with the ball now!".
Express - player: 9, is: angry.
Express - player: 12, is: sorry.

Episode at time 579.91:
Run - player: 9, at: P( 25, -10, 2 ), velocity: V( -0.11, -6.9, 0 ).
...
```

---

[14] One may wonder how the FA knows that it should be "sorry." Given that there was a failure in its reasoning and as a result there were not very friendly gestures from the "captain" (player number 9), it would not be difficult for the FA to deduce that it has made a mistake.

The FA has the ball but does not know what to do with it; of course the "captain" is angry, and the FA is sorry for the mistake; in this case the captain itself (player number 9) is the best player to back up, so it runs towards the FA to take the ball from it (the rest of the cycle is omitted here as it continues as an ordinary one).

```
...

Episode at time 634.46:
Throw - player: 1, threw object: 1, object velocity: V( 16, -0.85, -0.68 ).

Episode at time 635.65:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 2.9, -0.82, -1.2 ).
Run - player: 3, at: P( 30, 0, 1.8 ), velocity: V( -6.1, 0, 0 ).

Episode at time 638.43:
Express - player: 11, is: angry.
Express - player: 12, is: sorry.
Run - player: 11, at: P( 5, -4, 1.7 ), velocity: V( 6.8, 1.4, 0 ).
...
```

The FA did not run for a ball as it was supposed to do, fieldsman number 11 backs up for the FA (and complains about it).

```
...

Episode at time 1927.22:
Throw - player: 1, threw object: 1, object velocity: V( 16, -0.84, -1 ).

Episode at time 1928.38:
Hit - player: 2, with object: 2, hit object: 1, object velocity:
      V( 11, -5.2, 3.8 ).
Run - player: 3, at: P( 30, 0, 1.9 ), velocity: V( -6.2, 0, 0 ).
Run - player: 9, at: P( 25, -10, 2 ), velocity: V( 3, 6.2, 0 ).
Run - player: 12, at: P( 15, -15, 1.8 ), velocity: V( -2.5, 6.2, 0 ).

Episode at time 1928.39:
Express - player: 12, is: sorry.

Episode at time 1929.44:
Collect - player: 9, at: P( 21, -5.7, 2 ), collected object: 1.
Express - player: 9, is: cheerful.
Express - player: 2, is: disappointed.
...
```

The FA thought it was its ball but it was not; fieldsman number 9 was responsible for it (and achieved an Out Caught, hence the respective cheerful/disappointed gestures for itself and the batsman).

```
...

Episode at time 9999.15:
Throw - player: 1, threw object: 1, object velocity: V( 16, -0.93, -0.87 ).
```

The simulation was set to finish at time 10,000 s *(this simulation was not run in the interactive mode).*

The Text Window

The SGS is intended as a textual representation of the whole game, particularly for future experiments on an anthropological basis (see Sections 6.5 and 9.2), rather than for an understanding of some details of the game, which can sometimes be missed in the animation. For this purpose, the text subwindow of the simulator is provided. The complete contents of the text window are written to a file named `CricketTXT<time>.txt` (where `<time>` is as before). An abridged version of the contents of a text window (showing only all the possible types of messages displayed) can be seen below (using the same conventions as before for comments and abridgements):

```
Configuration for the Simulation(s)

Player  Height  Speed   Strength
------  ------  -----   --------
BM      1.8     7.0     7.5
BR      2.0     6.4     2.8
WK      1.6     6.0     2.4
S1      1.5     6.0     2.0
S3      1.8     6.6     2.4
GU      1.7     6.6     2.5
FL      1.6     6.1     2.0
MF      1.8     6.3     2.5
LS      1.7     6.0     2.1
SF      2.0     6.3     2.5
SN      1.7     6.7     2.2
FA      2.0     6.2     2.4

Batsman's accuracy: 9


Bowling (0.00)
Runs Scored: 0
```

An ordinary cycle; the ball was actually batted but the batsman scored no runs.

```
Bowling (2.79)

Bowling (3.96)
Runs Scored: 1
```

The bowling at 2.79 resulted in a ball not being played (i.e., not being hit by the batsman); notice that another cycle (bowling) followed without there being any reference to the number of runs—for played balls an explicit number of runs is displayed, even if that is 0 (for the cases when no runs are scored).

```
...

Bowling (30.93)
Four Runs
Runs Scored: 4
```

The "Four Runs" expression indicates that the ball reached a boundary; the total number of runs scored is also displayed (as it can, in principle, be more than 4). Below are examples of the five ways of dismissing a batsman.

```
...

Bowling (190.61)
Out Caught

...
```

```
Bowling (997.48)
Bowled Out

...

Bowling (1897.14)
Run Out
Runs Scored: 0

...

Bowling (3062.80)
Out Leg Before Wicket

...

Bowling (9390.52)
Out Stumped

...
```

Finally, the statistics of the game are displayed at the end of a simulation (this was a non-interactive simulation set to stop at time 20,000 s*).*

```
Bowling (19996.29)


Statistics:

Balls Bowled: 5294
 Not Played:  1687 (31.87%)
Runs:         3995
Wickets:      276
 Out Caught:  150 (54.35%)
 Bowled Out:  67  (24.28%)
 Run Out:     27  (9.78%)
 Out LBW:     29  (10.51%)
 Out Stumped: 3   (1.09%)
```

At the end of a simulation (e.g., when the stop button is pressed, or for non-interactive simulations when the allowed time, passed as a parameter to the system, has expired), some statistics for that "match" are displayed. It states the number of balls that were bowled (and how many of them were not played, alongside the percentage for that), the number of runs scored by the batsmen and the number of wickets taken; it also shows the number of wickets taken through each way of dismissing a batsman (and respective percentages). The final scoring of such simulations can vary significantly (which is expected given the number of random variables in the simulation); the one above is just an arbitrary example.

6.4 The Foreign Agent

This section describes the initial work that has been done in regard to the *foreign agent* (FA). The FA is a player in the cricket society (i.e., our simplified simulation of the game of cricket) that does not know how to play the game. However, we assume that this player is capable of very simple tasks common to all ball games such as running towards a ball and catching it if it is near enough (this is the type of knowledge that such an FA could have brought from its society of origin). Further, for such an agent to be able to play the game properly, it must be endowed with a (machine) learning algorithm. In this implementation, we have used the very simple ID3 algorithm[15] (Quinlan 1986; 1993). The FA uses the learning algorithm on the observations of the behaviour of the other agents in similar roles in the game (in this case the fieldsmen) to determine eventually how to make the decisions that are "culturally-dependent" (that is, solving the local problems that are specific to the game of cricket, with which the FA is therefore not familiar).

That the FA needs very little knowledge of the game (besides the learning algorithm) and needs to be capable of only a few simple actions can be seen in Algorithm 6.3 below (which is also presented in Appendix C in similar form but in C++ syntax), yet its behaviour in the game is usually reasonable. This simplicity is because, in the FA, the decisions about whether to chase the ball or not, and where to throw the ball once it has been retrieved (the two main decisions made by the other fieldsmen) are left for the learning algorithm, based on the previously observed cases. Several versions of the two decisions trees (used for the situations just mentioned) are elaborated in the course of a simulation, as we explain in Section 6.4.1 on the implementation of the FA. In the algorithm below, **DecideToChase()** consults the last version of the decision tree for chasing the ball generated by the learning mechanism. This function returns either **true** or **false**, the latter also being returned in case consulting the tree returns **failure**. Similarly, **DecideWhereToThrow()** consults the last version of the decision tree generated for choosing the locations where retrieved balls should be directed. This function actually performs the action of "throwing the ball" to one of the two wickets, the wicketkeeper, or the bowler (according to the present situation), unless the consultation of the decision tree returns **failure**. In case the result from consulting either of the decision trees with the present situation of the game in the FA's perspective returns a **failure**, one of the misbehaviours of the FA (which we illustrated, in Section 6.3.3, when describing the SGS) concerned with not knowing what

---

[15] We shall comment later in Section 9.2 on our plans to use the cricket simulation as a testbed for some ideas related to a new conception of learning in DAI based on the notion of legitimate peripheral participation in communities of practice, presented in Section 4.4.

to do next will occur. (The two functions that access the decision trees are underlined in the algorithm below.)

In the terms that we have used in Chapter 4, if an informant agent—in this case the "*captain*," which we consider to be the fieldsman in the leg slip position—has given hints or instructions to the FA that after the batting act the ball is chased by one fielder (and only one, in this version of the game) and returned (by throwing, not carrying) to the wickets or one of the special players (bowler or wicketkeeper), this is all the information the FA needs, apart from the observation of cases for the learning algorithm (detailed in the next section). This simple information on the general structure of the game could also be obtained by inference of regularities in the SGS (say, by another (anthropologist) agent, or the FA agent itself) or with the use of variation on Q-learning (see, e.g., (Pebody 1997)), but we have not followed up these possibilities in the present work.

```
UpdateLearning()
if BallHitByBatsman() then
    if DontKnowIfBallIsMine() then
        if DecideToChase() then
            SetBallIsMine()
            RunToBall()
        else
            SetNotMine()
    else if BallIsMine() then
        if CaughtBall() then
            StopRunning()
            DecideWhereToThrow()
        else if Running() ∧ SomebodyElsesBall() then
            StopRunning()
            SetNotMine()
            Blush()
        else
            RunToBall()
```

Algorithm 6.3: Algorithm Producing the Complete Behaviour of the Foreign Agent

A few other functions in the algorithm above need explaining. First, the objects of class Player have a state variable indicating whether they should be responsible for chasing the ball or not (or whether they have not yet reasoned to decide whether they should do so or not). This is checked by DontKnowIfBallIsMine() and BallIsMine(), and controlled by SetBallIsMine() and SetNotMine(). Further, if the FA starts to chase a ball and realises that it was a wrong decision, given that another fielder is already pursuing the ball (SomebodyElsesBall()), it stops running and "blushes" (Blush()), i.e., the colour used for the FA in the screen changes from yellow to a pinkish colour, so as to show the user that the FA has made a mistake (this is true for all four types of FA's faults, which we mentioned earlier in Section 6.3.3). The function **UpdateLearning()** is explained in the section below.

Note that the FA does not use the sophisticated method for chasing the ball used by the other fieldsmen, who first run to a target position which is the closest retrieval point (i.e., the nearest point to a fieldsman's position that is on the ball's trajectory), and then run collinearly with the ball's trajectory, towards it (as we described in Section 6.3.2). The FA simply runs in the direction of the current position of the ball; at every simulation step the **RunToBall()** function must be called to update the direction of the FA's running to the current position of the (moving) ball; that is, it uses the simplest possible way of chasing the ball. This means that the FA misses the ball slightly more often than the other fieldsmen (which is realistic).

6.4.1 The Implementation of the Foreign Agent

The implementation of the FA is rather complex computationally. Two POSIX threads are constantly running in concurrency with the main simulator. Each one is permanently generating one of the two decision trees mentioned above: one for the decision of whether to chase the ball or not and the other for the decision of where to throw the ball once it is retrieved. At every simulation step, when the FA is given its opportunity to "reason," it checks whether either of the threads has finished execution (i.e., whether the relevant thread does not exist anymore), which means that the generation of a decision tree has finished, and if so, all the cases[16] accumulated up to that point, for that particular matter (chasing or throwing), are given to a new thread created for the generation of a new decision tree. (All this is the role of **UpdateLearning()** included in Algorithm 6.3.) This allows the main simulation to carry on while learning takes place (possibly with a large number of cases), without degrading the animation of the game as observed by the user. The last generated decision tree is permanently accessible to the FA, so whenever in its reasoning process a decision is needed, the appropriate tree can be accessed instantaneously.

Therefore, the implementation of the classes for the ID3 learning algorithm must be prepared for their use within concurrent asynchronous processes. Semaphores are used to control the access to mutual exclusion areas, thus synchronising the threads. This is necessary for the moment when a new decision tree is ready and must be recorded in the area where the FA constantly accesses it, thus replacing the old version of that tree.

The cases used for the training of the learning algorithm are accumulated in matrices in the SGS class. One of the matrices keeps all of the cases when fieldsmen decide to chase the

---

[16] Note that this means that we have not yet implemented incremental learning, which can be quite useful here.

ball and some of the cases when they decide not to run (because the number of decisions not to run is much greater than the number of decisions to do so, the ID3 algorithm would not work properly had we left all the negative cases). These cases have the form:

BallInDir	*DistBallTraj*	*BallSpeed*	***Run?***

where the attribute[17] *BallInDir* has values **yes** or **no**, stating whether the ball is moving in the direction of the player or not; attributes *DistBallTraj* and *BallSpeed* are real numbers giving the distance from the player to the nearest point in the ball trajectory and the present speed of the ball, respectively. The category ***Run?*** has possible values **yes** or **no** (besides **Failure**, intrinsic to the ID3 algorithm), according to whether the player has run or not. Attributes *DistBallTraj* and *BallSpeed* must be discretised before ID3 works on these cases. This is accomplished using the simplest algorithm given in (Quinlan 1993).

The other matrix keeps all the cases of players throwing the ball. These cases have the form:

DistToW1	*DistToW2*	*DistToWK*	*DistToBR*	*BMRunning*	***WhereThrown?***

where the four first attributes are real numbers giving the distance from the player to, respectively, W1, W2, the wicketkeeper, and the bowler; and *BMRunning* is a **yes/no** attribute stating whether the batsman was running (i.e., attempting to score runs) at that moment. The category ***WhereThrown?*** can have one of four values, namely, **w1**, **w2**, **wkpr**, **blr** (W1, W2, wicketkeeper and bowler), besides **Failure** of course. Again discretisation is needed, in this case for the four first attributes.

There is still the question of the FA arriving at the information that these are the important parameters to consider (from the several others that are, directly or not, in the SGS) for the learning process. Again, as in the case of the general information for the behaviour of the FA, this can be thought of as being taken from an informant agent (the captain, as we have mentioned earlier). Notwithstanding the plausibility of simple experiments on inference of regularities over the SGS removing the current need for that *a priori* information "given" to the FA, we are well aware that it entails a gross simplification of the migrant-agent problem. We therefore need to stress here that the aim of the experiments described in this chapter is not to test particular anthropological techniques for agent adaptation, as would be natural to suppose in the context of the thesis. Instead, we have initially chosen the idea of investigating the

---

[17] Note that in machine learning this word has a different meaning from the one used by anthropologists which we mentioned in Section 5.2.3.

suitability of traditional (general) learning techniques for some basic activity in a target society, which would allow an anthropologist agent (in the case where the FA is itself an anthropologist agent) to proceed with the investigation of the target society. Recall the idea of participant observation presented in Section 4.3; we emphasised there that for an anthropologist to gain access to the information relevant to an unexplored tribe s/he must join in the very activity in which the natives are themselves involved; and we mentioned there that the same would be needed in societies of self-interested agents. These experiments suggest that one example of a general knowledge which anthropologist agents ought to possess is basic geometry—not only for game-playing societies, but it appears to be a useful knowledge in general; and the results in the next section show that even a simple learning algorithm is of help in such situations too.

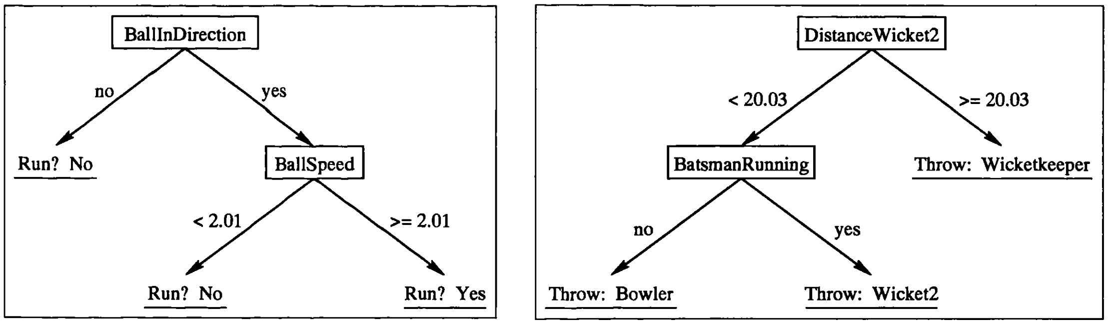
6.4.2 The Results of the Learning Algorithm

Similarly to the two other files mentioned before, a file `CricketID3<time>.txt` is created for every simulation. In this file one can find all the decision trees generated (with the ID3 algorithm) during the simulation for both chasing the ball and for the target of a throw of the ball (after it is retrieved), in a textual form. An abridged version of this file for an arbitrary simulation is given below (as before, abridgements are denoted by "..."; in discretised values, **LT** means "less than" and **GTE** means "greater than or equal"):

```
New decision tree for Chasing, ready at 4.7 seconds (7 cases):

ID3: Generated Decision Tree

BallInDir ( values order: no yes )
   Run?: no
   Run?: yes

New decision tree for Chasing, ready at 22.43 seconds (15 cases):

ID3: Generated Decision Tree

DistBallTraj ( values order: LT6.27 GTE6.27 )
   BallInDir ( values order: no yes )
      Run?: Failure
      BallSpeed ( values order: LT11.48 GTE11.48 )
         Run?: no
         Run?: yes
   Run?: no

...

New decision tree for Throwing, ready at 37.8 seconds (4 cases):
```

```
ID3: Generated Decision Tree

DistW2 ( values order: LT23.10 GTE23.10 )
   WhereToThrow: blr
   WhereToThrow: wkpr

...

New decision tree for Throwing, ready at 163.16 seconds (13 cases):

ID3: Generated Decision Tree

DistWK ( values order: LT20.39 GTE20.39 )
   WhereToThrow: wkpr
   DistW1 ( values order: LT29.12 GTE29.12 )
      WhereToThrow: blr
      WhereToThrow: w2

...

New decision tree for Throwing, ready at 4796.79 seconds (243 cases):

ID3: Generated Decision Tree

DistW2 ( values order: LT20.03 GTE20.03 )
   BMRunning ( values order: yes no )
      DistW1 ( values order: LT0.49 GTE0.49 )
         WhereToThrow: Failure
         DistWK ( values order: LT18.86 GTE18.86 )
            WhereToThrow: w1
            WhereToThrow: w2
      DistBR ( values order: LT19.06 GTE19.06 )
         WhereToThrow: blr
         WhereToThrow: Failure
   WhereToThrow: wkpr

New decision tree for Chasing, ready at 6197.82 seconds (510 cases):

ID3: Generated Decision Tree

BallInDir ( values order: no yes )
   Run?: no
   BallSpeed ( values order: LT2.01 GTE2.01 )
      Run?: no
      Run?: yes
```

As can be seen, the degree of complexity varies with the passage of time (i.e., number of cases available). Simple but effective decision trees generated in the simulations are depicted in Figure 6.2. A decision tree for chasing the ball is shown in a graphical form in Figure 6.2(a) and for throwing the ball in Figure 6.2(b). They do not match exactly the behaviour of the local agents (i.e., the expert fieldsmen), but they do produce a behaviour of the FA which is quite acceptable for the game. Note that the thresholds generated by the discretisation algorithm as shown in the trees below are quite accurate in relation to the actual behaviour of the local agents: recall that a ball speed of 2 m/s is the threshold below which the wicketkeeper is responsible for chasing the ball, so the fielders do not need to do it; and the fieldsmen always throw the ball to the wicketkeeper if they are too far away, which is considered to be more than half-way between the centre of the pitch and one of the boundaries, therefore that threshold is also sensible (it is actually somewhat lower, which incidentally is even better given the inexperience of the FA). However, as we have mentioned before, several trees are generated during the course of the simulations. Accordingly, depending on the accuracy of the last generated decision tree (which depends on the number and quality of the cases the FA had available at the time of starting the generation of that tree) the FA may or may not behave as expected, thus causing the types of (mis)behaviour we mentioned earlier when describing the SGS (in Section 6.3.3).



(a) Decision Tree for Chasing the Ball (b) Decision Tree for Throwing the Ball

Figure 6.2: Simple and Effective Decision Trees Generated in the Simulations

The decision tree in Figure 6.2(a) says that if the ball is not coming from the pitch in the direction of the location of the fieldsman (the FA), than surely it is not this player that should receive the job of chasing the ball. On the other hand, if the ball travels in the right direction, do not chase it only if it is too slow (also a proper piece of reasoning, as the wicketkeeper will take a slow-moving ball). Otherwise, the decision tree says that the player must run (to try to get the ball). Evidently, this is a safe behaviour for the FA in the sense that it will never make the mistake of not chasing a ball that is its responsibility to chase, but may quite often believe that it must chase a ball when it is actually another fieldsman's responsibility to do so, which is incorrect but not a harmful behaviour (recall that the FA stops chasing and "blushes" when it realises that such is the case, which indeed happens a great deal at certain points of the simulation).

The decision tree in Figure 6.2(b) indicates that the FA will (appropriately) throw the ball to the wicketkeeper when at a long distance from W2. (This is the same as being too far from the pitch—W2 being just one possible reference for the pitch—thus yielding a behaviour similar to the other fieldsmen.) If that is not the case, then if the batsman is running (therefore there is an opportunity to achieve a Run Out) the FA will throw the ball to W2. Of course the ideal here would be to calculate which wicket is nearer, but the limited learning algorithm used in the present work does not normally make such correlations; this sometimes happens (indirectly), as in the case of the more complex decision tree shown in the transcription of a simulation file above (the one for `Throwing`, ready at `4796.79` s, using 243 cases). Continuing with the rules given by decision tree, if the batsman is not running, the ball is to be thrown to the bowler (coincidently the right choice in the FA's fielding position), as it can then be made dead (recall that in the proper fieldsman's behaviour, it is thrown either to the bowler or the wicketkeeper, whichever is nearer).

6.5 Conclusion

We have described a simulation of the game of cricket where one the players is a "foreign agent"—it is not familiar with the game, and in future projects, neither need it be familiar with the culture of the local agents, including their language. The simulation is based on simple kinematics and geometry but displays a wide range of interesting cricket episodes, which can be observed both in the graphical interface and by means of the SGS. This gives plenty of scope for studies on learning the game; the choice of cricket is precisely because the spectrum of complications that can still be added to the game is enormous. Besides adding complications to the game itself, several further experiments can be carried out using the present state of the game, for example to study the behaviour of the FA in other fielding positions, learning the more improved method for chasing the ball that local agents use, or what happens when the FA uses other (more sophisticated) learning algorithms (or maybe even ID3 with other attributes). The SGS in its present condition can be used directly in experiments concerning an anthropologist agent (acting as the foreign agent) and the inference of regularities in the SGS in an attempt to make sense of the game, as we mentioned earlier. As for the complications to the game, anyone familiar of cricket could mention plenty; some straightforward improvements include: adding the second batsman (the non-striking one), creating a batsman that plays with some tactics (rather than randomly), devising more elaborate behaviours for the bowler and wicketkeeper, etc.

This work relates to our project on agent migration in a number of ways. It is a realistic example of the problem of cultural mismatch in social practice where one of the agents is from a distinct society of agents, although many levels of symbolic representation of the agents' mental states would have to be incorporated in the simulation for a thorough study of this problem. In this more elaborate setting of the cricket simulation, we could continue the project using our work on ascription of intensional ontologies extended with taxonomical relations (Chapter 5) and on the idea of legitimate peripheral participation, which we commented on in Section 4.4, influencing new conceptions of learning in DAI. These ideas of major future projects are discussed in Section 9.2. This work can also explore or be related to the concepts of formal cultural descriptions of societies, informant agents (to which we have alluded already in this chapter), and further anthropological notions which have been discussed in Chapter 4.

We have shown that even a simple learning algorithm can be effective enough in producing a "local" behaviour for a foreign agent: our foreign agent has been able to acquire a reasonable performance in terms of both ball retrieval and target of throws, which are fundamentally useful in cricket. Recall from Chapter 4 (more precisely Section 4.3) that an anthropologist must engage in *participant observation.* Thus, even simple kinematic knowledge plus general learning capabilities are important, insofar as they allow engagement in activity, for an agent interested in studying cultural aspects of a particular society (e.g., the ontology used therein as we have mentioned already). In particular, if the target society is comprised of self-interested agents, engaging in the activity of the community is a *sine qua non* for an anthropologist agent to gather information from that society—it helps in persuading the agents to reveal the information needed by the anthropologist agent for a more elaborate cultural description.

Part III

The Future

(*When Emotions Shall be a Sufficient Cross-Cultural Link*)

« ... Ἄτροπον δὲ τὰ μέλλοντα. καὶ τὴν μὲν Κλωθὼ τῇ δεξιᾷ χειρὶ ἐφαπτομένην συνεπιστρέφειν τοῦ ἀτράκτου τὴν ἔξω περιφοράν, διαλείπουσαν χρόνον, τὴν δὲ Ἄτροπον τῇ ἀριστερᾷ τὰς ἐντὸς αὖ ὡσαύτως· τὴν δὲ Λάχεσιν ἐν μέρει, ἑκατέρας ἑκατέρᾳ τῇ χειρὶ ἐφάπτεσθαι. »

Πλάτωνος, Πολιτεὶα

" ... *and Atropos the things that are to be. And Clotho with the touch of her hand helped to turn the outer circumference of the spindle, pausing from time to time. Atropos with her left hand in like manner helped to turn the inner circles, and Lachesis alternately with either hand lent a hand to each.* "

Plato, The Republic [617C–D]

(Shorey 1935)

Chapter 7

Universal Aspects of Agencies: Towards an Emotional Stance in MAS

We present a simulation of a society of agents where some of them have "moral sentiments" towards the agents that belong to the same social group. The Iterated Prisoner's Dilemma is used as a metaphor for the social interactions. Besides the well-understood phenomenon of short-sighted, self-interested agents performing well in the short term but ruining their chances of such performance in the long run in a world of reciprocators, the results suggest that, where some agents are more generous than that, these agents have a positive impact on the social group to which they belong, without compromising too much their individual performance (i.e., the group performance improves). The inspiration for this project comes from a discussion on Moral Sentiments by M.Ridley. We describe various simulations where conditions and parameters over determined dimensions were arranged to account for different types and compositions of societies. Further, we indicate several lessons that arise from the analysis of the results and comparison of the different experiments. We argue that allowing agents to possess suitably-chosen emotions can have a decisive impact on Multi-Agent Systems in general. This implies that some common notions of agent autonomy (and related concepts) should be reexamined. The motive behind this work on advocating the use of emotions in agent architectures is the idea taken from anthropology that emotions are an important cultural link among all human cultures, thus it has an important role to play in the future of anthropological work on interoperability of MAS.

7.1 Introduction

We have recently come to appreciate the importance, for this project, of including issues of *emotions* or *sentiments* (we use these words interchangeably here) in the modelling of MAS in general, as we shall discuss later. Inspired by Ridley's ideas on Moral Sentiments (1996) we also wish to argue here that emotions are a missing factor for agents to display social behaviour; it indicates why autonomous agents do not have to be necessarily selfish (as claimed, e.g., in (d'Inverno and Luck 1996b)). Further, notions of goal adoption and social norms (Conte and Castelfranchi 1995) also need to be revisited, based on the *emotional stance* that we propose informally for MAS. (These points are discussed in Section 8.3.) We began the work reported here, whose initial result were presented in (Bazzan, Bordini, and Campbell 1997), by introducing (metaphorically) an emotional aspect into simulations using the Iterated Prisoner's Dilemma[1] (IPD). Ironically, IPD originated in the field of Game Theory, an area concerned with *rational* decisions and self-interest. In (Bazzan, Bordini, and Campbell 1999) we presented further results and conclusions, and (Bordini, Bazzan, and Campbell 1998) extended their analysis and discussion.

It is known that in the IPD, mutual defection is not the only solution, unlike the situation for the one-shot Prisoner's Dilemma (PD), where it is the rational one (Brembs 1996). This was verified in Axelrod's computer tournament (1984), by the use of a tactic referred to as Tit-For-Tat (TFT), whose success is due to its being, in Axelrod's words, nice, retaliatory, forgiving, and clear. These ideas have been employed in the field of MAS in order to enable and explain the achievement of cooperation and coordination among self-interested agents (e.g. in (Rosenschein and Zlotkin 1994)). Nevertheless, little work has been done using the IPD in interactions among social agents, particularly agents with emotions. One reason for this may be the failure of theories based on rational choice to account for *social* action and choice, as suggested by Conte and Castelfranchi (1995). Game theory tends to treat social agents' goals in a process of choice for each agent in isolation and exclusively from its own point of view; for example, agents do not attempt to modify the mental state of an opponent. This unusual setting for the IPD (i.e., considering what happens in the game if some agents are not at all rational) is arguably the source for the new insights we present here.

There has been marked opposition from game theorists to the widespread use of IPD and TFT as the basis for explaining complex social interactions among humans (Binmore 1998). What we have done here is extend the rules of the game so that it can relate to various issues of social agents, including moral and philosophical aspects such as why people are able to keep their promises once they agree to cooperate and why people behave (truly) altruistically. We have used the PD simply as a metaphor for social interaction, without the intention of further colonisation of game theorists' territory.

The ideas on moral sentiments that inspired this work are presented in the next section.

---

[1] We provide the basic ideas on the IPD, for the reader not familiar with them, later in this section.

What Is Done		What Happens	
By You	By Your Accomplice	To You	To Your Accomplice
Give Evidence	Stay Silent	Released	Long Jail Sentence
Give Evidence	Give Evidence	Short Jail Sentence	Short Jail Sentence
Stay Silent	Stay Silent	Fined	Fined

Table 7.1: A Version of the Prisoner's Dilemma Story

We then describe, in Section 7.3, the simulation that assesses some of those ideas, and show the results and some analysis in Section 7.4. In Chapter 8 (Section 8.3) we shall present several discussions of relevance to our project based on these results and contrast our conclusions with some widespread conceptions of agent autonomy and related notions.

(N.B., the text below is intended for the reader with no prior knowledge of the Prisoner's Dilemma. Those familiar with it should proceed to Section 7.2. It is, however, a very simplistic introduction; for further details see, e.g., (Brembs 1996).)

The Basics of the Prisoner's Dilemma

The Prisoner's Dilemma (PD) is a simple two-person game[2] in which players can either cooperate with or defect each other. The game is usually presented by means of a story about prisoners (see below), hence its name. It is considered as a standard model for the study of the evolution of cooperation, particularly based on the theory of *reciprocal altruism* (Trivers 1971). As Brembs (1996) puts it: "The 'Prisoner's Dilemma' is used as the standard metaphor to conceptualise the conflict between mutual support and selfish exploitation among interacting non-relatives in biological communities." It is used not only in biology but in many areas of research: there is a vast literature that makes use of the PD metaphor.

Our version of the story about the two prisoners, based on some of the stories available in the literature, goes like this. Two persons are charged with a joint violation of law and are held by the police in separate cells so that they cannot communicate (e.g., to discuss a common plan of action). Also, the players are supposed not to be acquainted with each other and think that, in all likelyhood, they will not be faced with this "dilemma" (i.e., whether to cooperate or defect that person) again in the future. The police then tell each of the prisoners that their situation is as given in Table 7.1.

From the choices given in the table, clearly to be released is better than to be fined, which

---

[2] We do not provide here the elements of game theory (a mathematical theory of optimal choice of strategy in face of conflicting interests), which influences a whole paradigm of MAS; we have given references to this approach to MAS above and also in Section 2.3.

P1/P2	C	D
C	R/R	S/T
D	T/S	P/P

- $T > R > P > S$
- $R > (S+T)/2$
- e.g., $T = 5, R = 3,$ $P = 1, S = 0$

Table 7.2: The Prisoner's Dilemma Payoff Matrix

is better than to have a short sentence, which is better than to have a long sentence. (The idea is that even if one is found guilty, having at least helped the police reduces one's sentence; whereas without further testimonial evidence they cannot fully charge anyone with the violation of law.) Now, what should one do in such a situation: stay silent (i.e., *cooperate* (C) with one's accomplice) or give evidence against the other (i.e., *defect* (D) one's accomplice)?

In mathematical or computational versions of the game, it is traditionally represented by the payoff matrix given in Table 7.2 (the payoffs of both prisoners P1 and P2 are given there, separated by "/"). In the table, T stands for *Temptation to defect*, R stands for *Reward for mutual cooperation*, P stands for *Punishment for mutual defection*, and S stands for *Sucker's payoff*. If the two conditions to the left of the table are satisfied, there is no better solution in the one-shot version of the game than to defect (for both players, given that they are assumed to be rational and therefore both reason similarly). The usual values for the payoff matrix are also given to the left of the table (these are also the values used in the simulations presented in this chapter).

When played repeatedly, however, there is no trivial answer for the best strategy for a player in the game (given that individuals may meet again in the future to play the game, and therefore retaliation may occur). In this setting, the game is called the Iterated Prisoner's Dilemma (IPD). The extraordinary interest in this game started when Axelrod announced that the strategy called Tit-For-Tat (TFT) had won his computer tournament (Axelrod 1984). This strategy simply starts by cooperating and then copy the opponent's last move. As we have mentioned, Axelrod says the strategy's strength is in the fact that it is "nice" (starts by cooperating), "retaliatory" (punishes a previous defection with defection), "forgiving" (returns to cooperation after the opponent "repents" and cooperate), and "clear" (it is easy to spot a player following this strategy, thus allowing others to adequate their strategy for mutual benefit).

7.2 Moral Sentiments: a Prolific Source of Ideas

In a recent publication on "The Origins of Virtue" (Ridley 1996), particularly in its Chapter 7 entitled "Theories of Moral Sentiments"[3], which draws significantly on the ideas of (Frank 1988), Ridley makes the point that *moral sentiments* (emotions like compassion towards others and guilt for not having played fair with someone) prevent us from being *rational fools*. These are short-sighted, self-interested people who fail to consider the effects that their actions have on others. They act to maximise their earnings in the short term but spoil their chances of doing well in the long run because people do not reciprocate with those who have proven selfish in the past.

Moral sentiments lead us to sacrifice rational decisions, yet they are of fundamental importance to social relations inasmuch as they allow us to create a reputation as altruistic people. Altruism, which most people praise as a virtue, will lead a kind person to have a good reputation, hence paying off in future interactions (see comments below on the role of trust in PD situations). However, these same emotions drive us to want those who belong to the same social group to be somewhat self-interested, which is better for the group too: we are particularly concerned with the welfare of people from the same social group or those who "share our genes."

In other words, moral sentiments are decisive in the dilemma between getting the maximum out of an opportunity and being cautious about the future. They are instinctive[4], an intrinsic part of our highly social nature—remarkably, distinct sciences have arrived at this conclusion from completely different sources of evidence, as Ridley insightfully points out[5]. In effect, they are the *guarantee of our commitments*, which makes complex, stable social relations possible, and this stands to our long-term advantage. They alter the reward of problems in which one must be committed to cooperation (like the PD) and somehow bring to the present distant costs that would not have arisen in a rational calculation. Furthermore, the fact that emotions are *universally recognisable* allows the virtuous to get together and take advantage of cooperation (thinking in terms of the PD helps here too), isolating the selfish rationalists: people can actually

---

[3] The reader will notice that this is an important influence on this chapter. This does not mean, of course, that we share all the views expressed in that book, in particular the political implications drawn towards the end of the text.

[4] This is not to say that we are not self-interested and that we do not have other "darker" instincts too.

[5] However, Ridley seems not to be aware of how similar his "moral sentiments" are to what Plato classifies as a third part of the *psuchē*, which is, in the traditional translation cited below, referred to as "high spirit" (from the Greek *thūmos*). He therefore misses a very good (and almost 2.4 millenia old) piece of support for his argument (not only in terms of the role of moral sentiments in support of rationality but also in what concerns its innateness). As Socrates says in the Republic of Plato: "For then we supposed it [*thūmos*] to be a part of the appetitive, but now, far from that, we say that, in the factions of the soul, it much rather marshals itself on the side of the reason." (Shorey 1930; 440e). Plato also compares the high spirit to the "helpers" which are part of the structure of his idealised *polis*: "... or just as in the city there were three existing kinds that composed its structure, the moneymakers, the helpers, the counsellors, so also in the soul there exists a third kind, this principle of high spirit, which is the helper of reason by nature unless it is corrupted by evil nurture..." (Shorey 1930; 441a).

avoid "playing" with those who do not reciprocate in real life. To summarise the workings of emotions in social life: "Rage deters transgressors; guilt makes cheating painful for the cheat; envy represents self-interest; contempt earns respect; shame punishes; compassion elicits reciprocal compassion. And love commits us to a relationship."(Ridley 1996).

When, however, people perform altruistic acts that are not rational and do not pay off even in the long run (which we shall refer to as *true altruism*), they are falling prey to sentiments originally designed (through natural selection) to obtain other people's trust, which is convenient for "real life's prisoner's dilemmas"[6]. This is a remarkable insight from Frank's theory (1988), as it admits some light into the discussion on altruism, which has become so paradoxical and with dangerous consequences since the wide acceptance (followed by misinterpretations) of the "selfish gene" theory (Dawkins 1989).

In terms of MAS, we expect that the representation of emotions in agents' architectures, leading agents to have moral sentiments, could account for a type of long-term coordination which would not be possible otherwise. The nature of the beneficial effect of altruism is such that *it cannot be calculated in advance*: not by human beings, let alone computational agents. This is a direct consequence of the principle of bounded rationality, and MAS should borrow from the ingenious mechanism designed by natural selection to solve this problem (i.e., our innate ability to empathise, to feel guilty, etc.; in other words, our moral sentiments). In more practical terms, agents who seek only to maximise a utility function or are overconcerned with self-interest (Rosenschein and Zlotkin 1994) can miss good opportunities for themselves that they cannot foresee. In order to make our point clearer, it is important to emphasise the characterisation of (true) altruism*true altruism—ii*: *an altruist acts without an anticipation of increased payoff at any time* even if such reward is likely to happen. Contrarely, the most usual type of "altruism" referred to is the type of altruism that is an investment in trustworthiness, which eventually repays itself, and can, therefore, be considered ultimately as selfish, according to some schools of thought; it is not necessarily the "misuse" of sentiments that allows what we referred to earlier as true altruism.

To all this, a utilitarian MAS researcher might well counterargue that, if humans fall prey to these sentiments and end up doing things that are not to their own advantage even in the long run, then surely there is the chance of the same problem befalling agents if they are endowed with emotions. To show that, even if this happened, it would not be a "problem," is the main point of the present work. The results of our experiments show clearly that even this type of truly altruistic behaviour has an indirect, unwitting benefit if a group of agents is considered

---

[6] It should be noted that, as Ridley puts it, this is only a dilemma if one does not know whether one can trust one's accomplice.

(recall that altruism that has a positive reward also for the individual can be regarded, as reciprocal altruism is (Trivers 1971), to be ultimately selfish, even if the motivation is not). Yet, the compromise for individual altruistic agents is not as great as the societal benefit.

In brief, the main lessons drawn from Ridley's book are that we keep our *commitments* (or break them) and are capable of living in complex, stable social environments because of our *emotional decisions* (based on our instinctive virtue), not because of our rationality alone. This is enough support in favour of an emotional stance for MAS, given that the whole point of this area is to profit from social action that enriches individual capabilities. If this is so, it means that we need to revisit several notions in MAS, particularly the concept of agents' *autonomy*. We concentrate, briefly and informally, on these issues in Section 8.3.

The discussions are much more elaborate in (Ridley 1996; Chapter 7), so we advise the reading of that material for a complete account. The points presented in this section comprise the general ideas that have inspired the conception of the simulations we describe below. Similar motivations are also seen behind the ideas discussed in (Simon 1990)[7] and to some extent in the simulations in (Cesta, Miceli, and Rizzo 1996; Castelfranchi, Conte, and Paolucci 1998; Kalenka and Jennings 1998), although these motivations and ideas were explored there in rather different contexts and not with this particular issue of emotions in mind.

7.3 Description of the Experiments: Agents with Moral Sentiments in an IPD Exercise

In our simulations, agents interact by playing the IPD. Unlike other experiments reported in the literature, the overall population of agents (which we refer to as *society*) here is divided into *groups*; we also classify agents (with consequences in the strategies they follow) according to their "wealth," which is also unusual. We design them to reflect different characteristics of social groups, as we want to investigate Ridley's point about the effects of altruism for agents and their social groups. We analyse the effects of parameters such as population size, how agents are organised (i.e., small or large groups), percentage of agents playing at a time, number of egoists and altruists present in the groups, different thresholds to classify agents in terms of their performance, and the role of bankrupt agents.

At each step of the simulation the variable characteristics of an agent are assigned: whether it plays or not, a random opponent, its wealth state, etc. Once this is known, each agent decides how to play according to the tactics we shall introduce later. The points they earn by playing

---

[7] Incidentally, Ridley cites this very same paper when discussing cultural conformism in later chapters of his book, which we only recently came to realise.

are the standard amounts for the PD payoff matrix (given in Section 7.1). After each move individuals collect the points earned (as in some sort of bank account). The average value of points earned in a certain number of simulation steps is used to determine the *wealth state* of the individual, which influences how it will play. According to certain thresholds on this average, an agent's state in each simulation step can be *Wealthy* (W), *Medium* (M), or *Straitened* (S). An agent is in state W if its average score is greater than a certain threshold T_W; it is in state S if its average score is below a certain lower threshold T_S; and it is in state M otherwise. These thresholds vary in the different types of societies we study, and are better detailed next when we explain these differences. To represent the effort one puts into interacting socially, agents pay P points to play.

In all experiments conducted, individual agents playing the IPD can be either egoistic or altruistic. The former type of agent defects in all interactions (ALLD), and the latter plays either Tit-For-Tat (TFT) when the opponent is an agent from another group, hence displaying a fair (reciprocating) behaviour, or plays with Moral Sentiments (MS) when interacting with agents from the same social group (the truly altruistic behaviour). We use the word *partner* to refer to an agent who belongs to the same social group as another. Pairs of agents (from all groups) are chosen randomly to play each round of the game.

The MS strategy for an altruistic agent means, basically, that it cooperates with its partners unless it is in a straitened state and the opponent is wealthy. To understand this, we must first think of the most altruistic strategy that is possible in the IPD, namely ALLC (always cooperate). Our MS strategy is close to that, apart from the fact that the altruistic behaviour is dropped when the agent is itself in need of help. Note that the straitened altruist will only defect from a wealthy partner, so its defection will not be too harmful for the other agent (metaphorically speaking, of course); the straitened altruist does not try to recover by taking advantage of a partner who is not in a wealthy state. It is important to emphasise that, when using the MS strategy, the opponent (in terms of the PD) is necessarily a partner (i.e., from the same group), but may be an egoistic agent. In this case, the straitened altruist avoids being exploited by its partner since it cannot afford this (usually it cooperates even with egoists, except when the agent itself is straitened). If the opponent is an altruistic partner agent (and, recall, wealthy), then the straitened altruist will clearly be helped (as it will get T points). In the perspective of the wealthy altruistic agent, there is nothing wrong with this: remember that by the fact that they both belong to the same social group we mean that there is an emotional liaison between them.

To summarise, altruists in state W or M always cooperate, and a straitened altruist cooperates if the opponent is either in state S or M, and defect if the opponent is in state W. This is the social mechanism to compensate for the short-sightedness of some and to allow agents to recover from a history of bad performance, which is possible through the altruism of those from the same group or by avoiding overexploitation (recall Ridley's mention on one not being *too* generous). These tactics are clearly stated in Table 7.3, where it is shown how an agent plays (Ag. Plays)—it either *cooperates* (C) or *defects* (D)—in each situation according to: the opponent's last move (Op. Last M.), the agent's state (Ag. State), and the opponent's state (Op. State).

Strategies	Op. Last M.	Ag. State	Op. State	Ag. Plays
ALLD (E)	*	*	*	D
TFT (A)	C	*	*	C
(non-partners)	D	*	*	D
MS (A)	*	S	W	D
(partners)	*	S	M,S	C
	*	W,M	*	C

Table 7.3: The Tactics Used by Altruistic (A) and Egoistic (E) Agents

As usual in IPD experiments, agents have the means of remembering the last move of any opponents they have met in the past, if needed. This has no meaning for "pure" egoists since in our design they always defect. However, if they are to learn from past experiences (e.g. that this behaviour does not pay off and that egoistic may not mean "always defect" but rather "only defect while others still cooperate with you"), then that capacity to remember may be useful for egoists as well as others[8]. For altruistic individuals, the memory of the past move influences their current one in the following way: if the opponent is not a partner and has defected in the last move, TFT will lead to defection as well. If the opponent is a partner, then again the last move is irrelevant since the determining factor here is the wealth of both individuals.

We have investigated many scenarios by combining specific values for the parameters mentioned earlier. We design each situation to model characteristics that we can relate to human societies and/or multi-agent systems. We report here the most significant ones.

There are four types of society[9]: *Long Memory* (LM), *Generous Middle Class* (GM), *Polarised Society* (PS), and *Fair Shares* (FS). In the LM the strong characteristic is that the whole history of agents' performance in their past plays is remembered when classifying them as W, M or P. On

---

[8] But note that proper short-sighted agents are needed in these experiments, for the point is to show the role of true benevolence in the presence of really egoistic agents that do end up bankrupt. Besides, if the egoists ever stopped defecting, the altruistic agents would not be true altruists, as they could be compared to reciprocal cooperatores, the ones that are interested in altruism as a form of investment.

[9] We have used brief illustrative names to characterise the types of society we have simulated, so that readers can interpret them easily in the text.

the other hand, in the other three types of society, the states of the agents are a function of their performance over a fraction of this time. We have experimentally set this *History Length* to be $HL = 10$ steps for the societies with a limited length of history[10].

As for the GM society, the aim is to check the performance of the groups when agents in state S also take advantage of the MS strategy when playing with agents in state M (according to Table 7.3, this only happens when playing with wealthy agents). In fact, the role of agents in state M is not the one of the middle class, but the role of the wealthy altruists of the society (one may think of them as really "compassionate:" they help even though they cannot afford it as much as the wealthy ones). Therefore, in practice there are only two states of wealth in this society. The PS has been designed (by setting different thresholds for the classification) so that only a small proportion of agents can be considered to belong to the middle class of the society, but again straitened altruistic partners do not borrow from the middle class (i.e., they play according to the table of tactics).

In the FS the wealth classification of each individual agent depends on the points earned by every agent in the group in the recent history. Each group has its own thresholds which vary at each simulation step. Let $k_1 = P/2$ and $k_2 = T/2$ be two constants (defined experimentally). We identify the wealthiest agent and the poorest agent in a group Gr, so as to calculate a *difference of wealth* (DW) in that group, according to:

$$DW_{Gr} = max(AvgScs_{HL}(Gr)) - min(AvgScs_{HL}(Gr))$$

where $AvgScs_{HL}(Gr)$ are the average scores for each of the agents in group Gr in the last HL steps.

If the difference of wealth in the group is substantial, i.e. $DW_{Gr} > (P - k_1)$, then there will be three different classes of agents (as described above) and MS is played when an altruistic straitened agent and a wealthy partner are chosen to interact. Otherwise (i.e., DW_{Gr} is not substantial), no agent plays MS (there is no point in an agent being helped by a partner who has almost identical past earnings), which is to say that all altruists play *cooperate* (with partners) here. When there is a considerable difference in the recent earnings (i.e., the past HL steps) of the agents in the group being analysed, new thresholds are calculated for that group at this simulation step. For this, we employ a factor that reflects the difference of wealth (DWf) in the agents of the particular group (Gr):

$$DWf_{Gr} = DW_{Gr}/k_2$$

The thresholds for the current group (Gr) are then defined as:

---

[10] We have also performed simulations with smaller values of HL, which did not introduce significant changes in the results. A detailed study of the effect of the value of HL in the computation of agents' states is envisaged.

$$T_{W_{Gr}} = max(AvgScs_{HL}(Gr)) - DWf_{Gr} \text{ and } T_{S_{Gr}} = min(AvgScs_{HL}(Gr)) + DWf_{Gr}$$

where T_W and T_S are as described previously.

Each of the above types of societies was tested under the following conditions:

Bankrupt-Excluded (BE): if an agent runs too low in points (because, e.g., it has recently played with several egoists) it is said to be *bankrupt*, in which case neither itself nor its opponent is allowed to play; instead the bankrupt agent is awarded P points so that it can at least afford the price to play in the next step of simulation. (This unearned award is an easy solution—implemented in the earliest version–to deal with the problem of bankruptcy; although the next condition is more elaborate, we maintain the results of this version, as the comparison between the two scenarios is interesting.) Another important characteristic here is that all agents are scheduled to play.

Some-Play (SP): in this version of the simulation, agents are not given points to recover but are allowed to play even with negative balances in their accounts. Also, only a variable percentage of the agents are selected to play at each time. This aims at representing societies with different opportunities for agents to do "business" (in the form of IPD interactions); that is, some societies may have greater financial activity than others. The selection of agents to play does not discriminate among their states: even agents in a straitened state can play and in this case they can only hope that the opponent is a wealthy partner playing the MS strategy to help it to recover. Another peculiarity here is that agents that are not selected to play pay $P/2$ points, representing the costs for "living," which is half the resources spent to interact socially.

We have simulated the SP cases with $p = 60\%$ and $p = 80\%$ of agents playing. Finally, each one of the twelve combinations above (i.e. the four types of societies, each for BE, SP60%, and SP80%) was simulated for the following combinations of parameters:

- 3 groups, each with 4 agents: group G1 is formed by altruists only, group G3 by egoists only, and G2 is mixed (the composition of groups G1, G2 and G3 apply to all cases); three variations were simulated: with 1, 2, and 3 egoists out of the 4 agents in G2;
- 3 groups, each with 20 agents, G2 having 10 egoists and 10 altruists;
- 3 groups, each with 40 agents; again three variations were simulated here: with 10, 20, and 30 egoists in G2;

- 3 groups, each with 80 agents; also three variations: with 20, 40, and 60 egoists in G2;
- 3 groups, each with 100 agents, G2 having 50 egoists;
- 15 groups (5 groups of each type of composition, as for G1, G2 and G3), each with 4 agents, mixed groups having 2 egoists each;
- 15 groups, each with 20 agents, mixed groups having 10 egoists;
- 15 groups, each with 60 agents, mixed groups having 30 egoists.

That is, we have variations on few and many agents, in few and many groups. Each case was repeated 1000 times, enough to nullify variations in individual runs of the simulation. The simulation horizon was $t = 500$ steps for populations of 12 agents; $t = 1000$ for 60 agents; $t = 2000$ for 120, 240, and 300 agents; and $t = 5000$ for more than 300 agents in the society.

Altogether 168 cases were studied, from which we were able to verify the effects of the relevant parameters and draw several conclusions. These appear to be significant for MAS in general, in the quest to use human societies as a model for designing more efficient societies of agents, as we discuss in Section 8.3. The most significant of these conclusions are presented next.

7.4 Results and Analysis: the Unwitting Benefit of Altruism

Our main measure of performance is whether the *altruists* perform better than the *egoists*. This can be translated, initially, into whether the group G1 accumulates more points than G3 and, if it does, how long it takes for this to happen (normally the performance of G1 is poor near the beginning of the simulation, whilst G3 has a bad performance by the end). Since G2 is not a homogeneous group, we have to distinguish the two types of agents belonging to it, namely altruists and egoists, who clearly have different levels of performance: we call AM the subgroup of altruists and EM the subgroup of egoists in the mixed group (G2); the altruists in an homogeneous group (G1) are referred to as AH and egoists in G3 are referred to as EH. Hence, we also compare the performance of G1 to those of the AM and EM subgroups (and examine the time needed for G1 to surpass AM and EM in performance). Another important comparison is between AM and EM using the same criteria as for comparing G1 and G2. These are the comparisons that matter, because G3 has, as expected, a very bad performance.

Besides the time it takes for altruists to overtake the egoists, other measurements of performance were made using a snapshot in the simulation regarding the amount of points (per capita, so as to account for differences in the size of the groups) accumulated by the best group at that moment. This gives us a measure of the wealth of the society. Finally, we have also measured the number of times the MS strategy was selected and look at the effect this has on the wealth of the group and on the performance of the altruists.

Before we proceed with the material in this section, we provide a lookup table (see Table 7.4) with the abbreviations we have introduced so far, in order to facilitate the reading of the analysis of the results that the section presents.

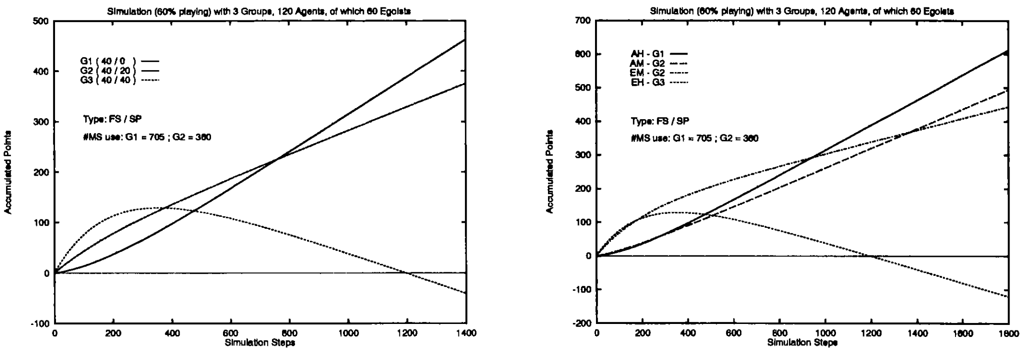
Ab.	**Definition**	**Ab.**	**Definition**
G1	A homogeneous group of altruists	MS	Moral Sentiments (the strategy we have introduced)
G2	A mixed group (i.e., one that has both egoists and altruists)	LM	Long Memory (all past earnings count in wealth classification)
G3	A homogeneous group of egoists	GM	Generous Middle Class (agents in state M also play MS)
AH	Altruist agents in a Homogeneous group	PS	Polarised Society (society with an attenuated middle class)
AM	Altruist agents in a Mixed group	FS	Fair Shares (society with variable wealth classification thresholds)
EM	Egoist agents in a Mixed group	BE	Bankrupt-Excluded
EH	Egoist agents in a Homogeneous group	SP	Some-Play (not all agents are chosen to play the IPD each step)

Table 7.4: A Lookup Table with the Abbreviations Used Below

There are two basic types of graphs that we have produced for the analysis of the results. The first shows the performance of each of the social groups. As for the second type of graph, we were able to observe the behaviour of each individual agent in every case in which the number of agents was less than 20 (e.g., as in Figure 7.5(b) which we introduce later). Otherwise, for the second type of graph, we only depict the behaviours of the four different types of agents mentioned above (namely, AH, AM, EM, EH). In the figures presented here the notation $G_n(a/e)$ means that group n has a agents of which e are egoists. Figure 7.1 has sample graphs showing our general results (for the performance of social groups and the types of agents in graphs 7.1(a) and 7.1(b), respectively).

Despite the various configurations of conditions and parameters introduced in the present set of simulations, we were able to confirm the initial conclusions previously reported in (Bazzan, Bordini, and Campbell 1997). These major lessons are:

1. The more egoists in a group, the faster the group collects points initially, but the worse is its performance after some time—which means that, as Ridley predicts, rational fools max-



(a) Performance of the Groups

(b) Performance of the Types of Agents

Figure 7.1: Graphs Illustrating the General Results

imise their earnings in the short term but compromise their future performance (a consequence of the reciprocating character of TFT, which is played among groups). Groups of altruists accumulate more points than the others in the long run.

2. The more egoists in the society:
   (a) the less time it takes for G1 to surpass G2, no matter what percentage of agents is playing; in most cases studied, this is also valid regarding the time for AM to surpass EM;
   (b) the fewer points it collects as a whole (regardless of the percentage of agents playing).

Thus, the presence of the egoists is harmful for all members of the group in the long run, since they would all be performing better if the egoists would stop being rational fools.

Next, we enumerate more recent conclusions, considering the whole range of parameters and types of simulations, as reported in (Bazzan, Bordini, and Campbell 1999) and (Bordini, Bazzan, and Campbell 1998). One very important observation which we make explicit here is that:

3. The generosity of the MS strategy yields a better performance for the group than mere reciprocity—recall that reciprocity is what accounts for the long-term success of individual agents and is ultimately the reason for the conclusions above—without compromising at the same proportion the individual performance of the altruistic agents that use the strategy. (This is explained in detail below.)

In Figure 7.2 we have added a fourth line to the group graphs (a) and (c) which is the average of the performance of G1 and G3 (**`Avg(G1,G3)`**, in the figure). This is to make it easier to visualise how much better G2 is doing because of the generosity of its altruists. Note that if the altruists in G2 were playing TFT within the group (i.e., a situation in which playing with a partner makes no difference as far as the tactics for altruists are concerned), the performance of G2 would be exactly that of the average of G1 and G3's performances, rather than what it is in the figure[11]. This is a remarkable finding, whose message can be understood better in terms of an analogy. If one gives money to a homeless person and nobody else gets to know about it, this is clearly the case where altruism has no personal reward. It is the situation that we mention in Section 7.2 where one is falling prey to the sentiments that are an important mechanism in social life, and it happens that they bring great personal rewards in normal circumstances through trustworthiness. In circumstances where no personal reward ensues, one interesting side-effect occurs: the improvement of the social group! In the homeless analogy, it is as if one is not better off (indeed one is worse off) by one's altruistic act, even in the long run, but one's city as a whole is, indirectly, doing much better (compare, in graph 7.2(c), G2 with `Avg(G1,G3)`), i.e., it does not reduce significantly one's own performance (compare, in graph 7.2(d), AM with AH). Regarding MAS, this implies a better performance of the overall system, which is important. In terms of the PD, this happens because, for the group, a joint reward of $S + T$ earned when an egoist plays against a wealthy/medium altruist partner, is better than $P + P$, received when TFT is played against ALLD. It is a similar effect to the (Adam) Smithian finding that the division of labour leads to society being *more than the sum of its parts*. The lesson here is: even the individual drawback of being driven by an emotional response when it will not repay itself allows a group that is encumbered with rational fools to perform much better than it would if emotions were used only for the purpose for which they evolved in our species (i.e., the virtuous purely maximising their individual earnings in the long run through reputation and reciprocation).

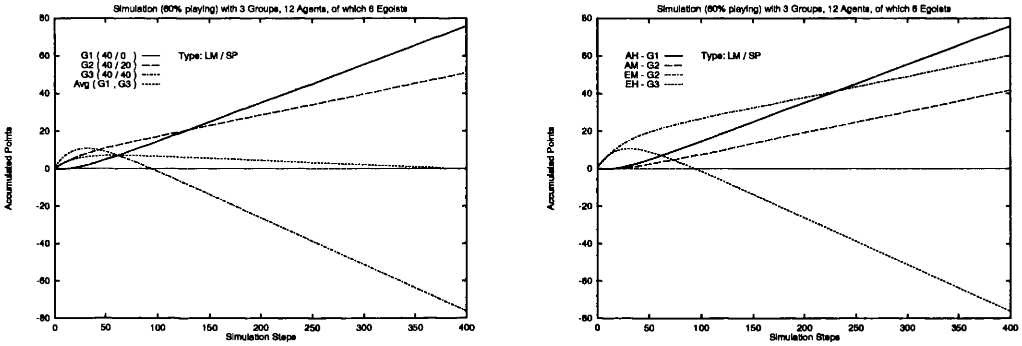
Besides the group improvement due to altruism, Figure 7.2 also shows the effect of the use of the strategy we have devised (MS) as opposed to other altruistic strategies (e.g., ALLC). Graphs (a) and (b) are from the LM type of society; that is, the one in which there is no limited history length. In reality, an unlimited history length leads to all agents being classified as M, thus the MS strategy was not used (except for a negligible number of times in the beginning of the simulation). Specifically, the altruistic agents have all been playing ALLC there. On the other hand, graphs (c) and (d) are from the FS society, where the use of MS is best managed (recall that the thresholds were variable there). A very interesting effect is seen here. Although the graphs for the groups (7.2(a) and 7.2(c)) have very similar shapes[12], one can see easily that

---

[11] This is so because, if a group has half of its agents performing as well as agents in G1 and the other half performing as badly as the ones in G3, the group as a whole has the exact average of G1 and G3.

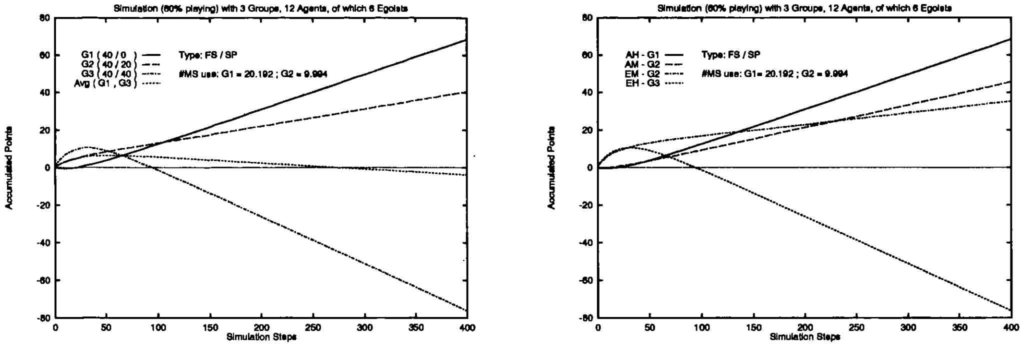
[12] The side-effect of the total amount of points earned in the society being slightly reduced with the use of MS is

this is not so in the graphs for the types of agents (7.2(b) and 7.2(d)). While in the LM society, where no MS is being used, EM agents (egoists in a mixed group) are doing consistently better than AM (the altruists in the same group), the same does not happen in the FS society, where MS is being used! The reason is that with ALLC the altruists are being too much exploited by their selfish partners. The MS strategy is "kind" enough to keep the performance of the mixed group higher than with pure reciprocation (remember the similarity of the curves for G2 in (a) and (c)), yet being fairer to those who are generous and actually responsible for the good performance of the group (i.e., the improved performance of the group does not rely on a disastrous performance for the altruists).



(a) Altruists as ALLC (Groups)

(b) Altruists as ALLC (Types of Agents)



(c) Altruists Using MS (Groups)

(d) Altruists Using MS (Types of Agents)

Figure 7.2: Effects of Altruism and of the Use of the MS Strategy

Concerning the various sizes of populations and their compositions, for both the BE and SP conditions, we conclude that:

4. The larger the number of agents in the groups:
    (a) the longer it takes for G1 to surpass G2 and for AM to surpass EM (although, in some cases of LM, G1 already begins by performing better than G2);

---

discussed later.

   (b) the greater the total use of MS (though this is not always true regarding its use per capita).

5. The reverse of items 4a and 4b is true if the number of agents is kept fixed but distributed in a larger number of groups. (See comment on this below.)

6. In general, the larger the number of agents in the groups, the more points the group collects, no matter what number of groups is involved.

7. The effect of the number of egoists on the amount of use of the MS strategy is not coherent among the different types of societies. While the latter increases with the former for the society FS, it decreases for GM, and has no monotonic relation in PS.

8. For the case where there is more use of MS (society GM), the time for AM to overtake EM is lower. However, the use of the MS strategy slightly decreases the number of points that are accumulated. (This is discussed near the end of this section.)

9. The more groups, the fairer the picture is: the EM subgroups of groups G6–G10 (mixed groups) perform closer to their homogeneous counterparts (G11–G15, egoists only), as do AM subgroups with respect to G1–G5 (see Figure 7.5(a), given later). As a consequence, the mixed groups perform worse than in the case with less groups; it is closer here to the average between groups of altruists and groups of egoists only, although still better than that level. In other words, the surplus performance of mixed groups that we discussed earlier is not as conspicuous here.

From items 4b, 5, and 9, we can conclude that the effect of the MS strategy (see Item 3 and the explanation for Figure 7.2) is more conspicuous in societies with few and large groups, where partners have more chance to interact. Before one tries any detailed comparison with human societies (which is rather outside the scope of this thesis), one must take into account that we have assumed here that the probability of interacting with a partner is the same as for interacting with agents from other groups, which may not be always true in real life.

By comparing the BE and SP conditions, we can conclude that:

10. Irrespective of the other parameters, when agents are playing under BE, A outperforms E by an even greater margin and the society as a whole collects more points. This is explained by the fact that in BE all agents play (except the small proportion that is bankrupt): the more opportunities for "business" (in the form of PD interaction) exist at each instant

in the society, the more rapidly the simulation stabilises (see items below about percentage of agents playing).

11. In general (but see below), under BE agents use the MS strategy more often, and this helps the altruists (recall the comparison between MS and ALLC).

The following summarises the conclusions from the comparison of the different types of societies and conditions concerning the use of the MS strategy:

12. Agents make use of MS more often under the BE condition than under SP, except for the FS society, where the number of times MS is played in BE is a figure between those for SP60% and SP80%.
13. Under the SP condition, there is more use of MS in the GM society, followed by the FS, and the PS.
14. Under the BE condition, there is more use of MS in the GM society, followed by the PS, and the FS.
15. Visual comparison of the graphs for GM and FS show that they are very similar; sometimes AM overtakes EM faster in GM, but FS has slightly greater accumulated points (the decrease is due to the current side-effect of the use of MS; see discussion below). We believe this is due to the use of MS being better managed in FS (i.e., it happens only when it is really necessary).

Comparing the *performance* of the different types of societies, we have:

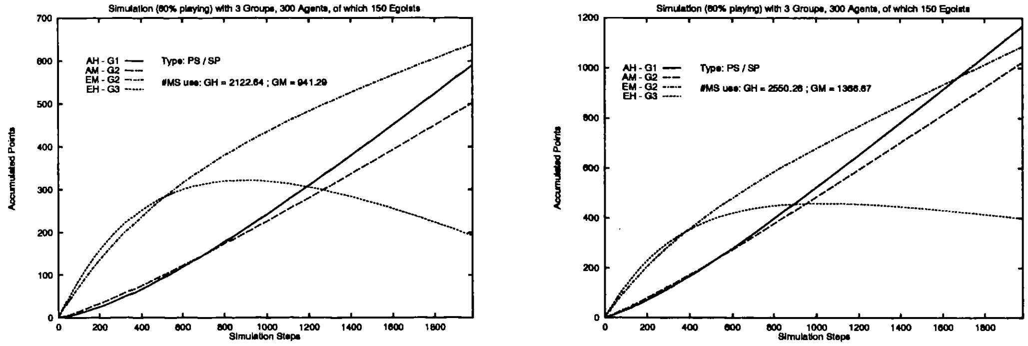
16. In all of them the number of points collected is almost the same, except for the LM society. This presents a clear differentiation among SP60%, SP80% and BE: in the first and the last, LM performance is much superior to that of the other societies (even 9 times better for SP60%), while under SP80% it is much inferior to the others.
17. LM also distinguishes itself from the other societies regarding the time G1 needs to overtake G2. This is lower than the others under SP60%, higher under SP80%, and it is higher than the others up to 120 agents but lower after that for BE.

Regarding the percentage of agents playing at each time, the basic lesson is, as expected, that the greater this is, the sooner the results are obtained (the fastest the simulation stabilises). Compare graphs (a) and (b) in Figure 7.3. We have confirmed that, the larger the percentage of agents playing:

18. The less time it takes for G1 to surpass G2 in performance, regardless of the number of egoists, except for the LM society, in which either the time needed increases with the percentage, decreasing for the BE condition, or it happens that G1 already starts better than G2 (considering BE as 100%).

19. The more points the group collects, except for the LM society in which actually the number of points decreases sharply from SP60% to SP80%, with BE achieving a value intermediate between these.

20. The greater the frequency of the use of the MS strategy (in SP only).



(a) Some-Play at 60%

(b) Some-Play at 80%

Figure 7.3: Effects of the Percentage of Agents Playing

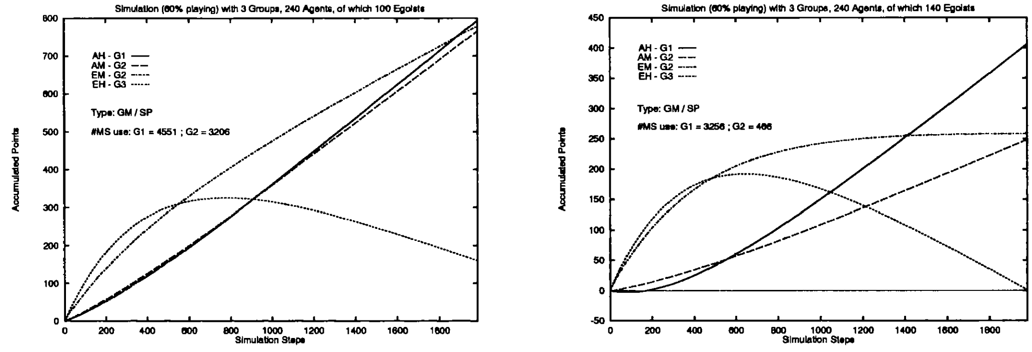
By comparing the cases of 12 and 120 agents (other parameters being the same), we have seen that:

21. It takes from 6 to 8 (according to the type of society and conditions) times longer for G1 to surpass G2 in performance in the latter case.

Finally, comparing the cases for which we have defined three variations on the number of egoists in the mixed group (keeping the other parameters fixed) we conclude that:

22. Every increase in the number of egoists in G2 by 25% causes the performance of those egoists to decrease by 36% of what it was before (in terms of accumulated points by the end of the simulation).

The numbers in the item above are averages of the ratios for all types of societies. There are variations for each type of society but the ratios are very similar when, in the same type of society, the total number of agents is increased. Based on this and other parameters still to be considered, one could, for example, create heuristics for an egoist agent deciding which group to join and seeking the one where its attitude would be most profitable[13]. The effect of increasing the number of egoists in the mixed group can be seen in Figure 7.4. Compare Graph (a), where the mixed group (G2) has only $\frac{1}{4}$ of egoists, with Graph (b) where G2 has $\frac{3}{4}$ of egoists, all other parameters remaining unchanged.



(a) Small Percentage of Egoists

(b) Large Percentage of Egoists

Figure 7.4: Effects of the Percentage of Egoists in Mixed Groups

The horizon of the simulations shown in some of the graphs was not always sufficient for AH and/or AM to perform better than EM. Nevertheless, the trend in those graphs (e.g., the ones in Figures 7.3 and 7.4) shows clearly that this would happen soon afterwards. In some cases with very few EM, these really have the best performance of all (and in some cases when MS was not being used, they performed better than AM). This is because the few egoists can count on the generosity of the (many) altruists in their group, who cooperate with them despite the fact that they are selfish, for the sake of group welfare.

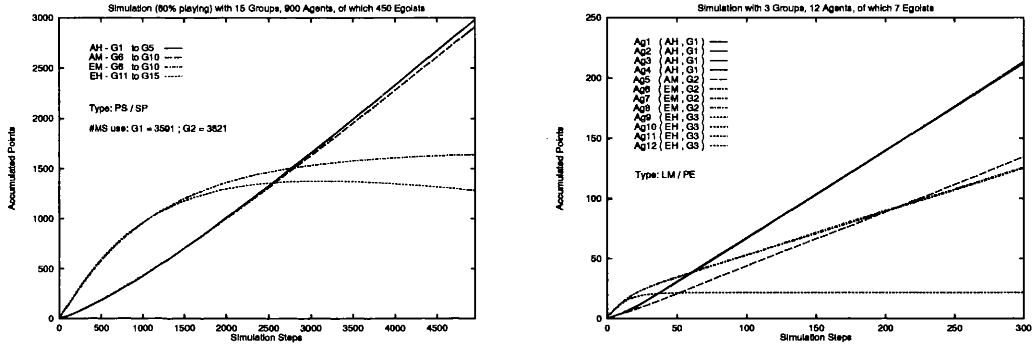
A current drawback is the fact that the use of MS slightly decreases the amount of accumulated points in comparison with ALLC (see Figure 7.2), because this "defect to be cared for" mechanism results in a lower total of points to the group than a mutual cooperation. Recall that for the group $S + T$ was better than $P + P$, but it is not better than $R + R$ (due to the condition of the PD that $R > (S + T)/2$ must hold) which could happen in the case of two altruists from that group playing ALLC being chosen to interact. However, it is important to remember too that this mechanism is essential to allow the altruists to discern when they can afford to cooperate with partners (who may be egoists), and at any rate allows for individuals to recover from bad histories of performance; we have discussed this in our comments on Figure 7.2. Therefore, future extensions should also include some sort of penalty for societies with straitened agents, which would reverse the current side-effect of decreased accumulated points due to agents helping their partners. This is not just artificial modelling, for we see this phenomenon in human

---

[3] We did not consider here the time factor, for example, which could mean that for short stays certain kinds of groups could be better than others for the opportunistic agents.

societies: the more capital a group has, the greater the amount of money that can be earned from the financial interactions in which it engages. Note that, if we decided that altruists would always cooperate among themselves (even when not wealthy) in order to prevent the drawback, there would not be a clear way of distinguishing altruist and egoist partners without changing the rules of the PD and agents being informed of the opponents' strategy artificially. (Interestingly, it is not possible to distinguish these strategies merely through behaviour.)

Consider also Figure 7.5 below. Graph 7.5(a) shows the situation for many groups, and 7.5(b) the LM/BE society with a line for each individual agent, a type not shown in the graphs before.



(a) Many Groups and Agents

(b) The LM/BE Society / Individual Agents

Figure 7.5: Supplementary Graphs

In summary, the results show clearly that, in the long run, groups of altruistic agents (G1 and AM) accumulate more points than any other. The altruists are not rational fools: they compromise their present possibilities of gain to make sure they will do well in the future. The whole group performs well because individual failures are compensated by the generosity of those doing well, avoiding bankruptcy. In other words, to reciprocate pays off in the long run and pure altruism improves the performance of the group! Also, restrained altruism (rather than unconditional) produces more sensible results as far as the altruists are concerned (they at least perform better than their egoistic partners, while still keeping an improved group performance). We have also verified that homogeneous groups of egoists (G3) perform very well only in the short term. Their selfishness in the game compromises their reputation: once the (sensible) agents in a society have found out about their character, they suffer retaliation (characteristic of the TFT). At this point, egoists either only earn enough points to survive, that is, pay to play in the next step (in the BE condition), or go to complete bankruptcy (in SP). Figure 7.5(b), where the curves for agents in G3 (EHs) are logarithmic and eventually stabilise, confirms the former case, while the latter case was seen in Figure 7.1(b), where the curve falls sharply, even to negative values. Mixed groups (G2) have an intermediate performance, but they do not exhibit the "catastrophic" effect seen in a group where there are no altruists. The presence of some altruists there assures the relative development of the group.

7.5 Conclusions

Our results suggest strongly that rational fools maximise their earnings in the short term but compromise their performance in the long run, while altruists may not have the best performance at the beginning of the simulations, but normally end up much better than the others. The results also show clearly that the more altruists there are in a group, the better they, and the group as a whole, perform; more importantly, their generosity, although somewhat "irrational" from an individual point of view, implies that the group as a whole performs significantly better than when pure reciprocation is used. Accordingly, we conclude that to behave rationally (in the classical sense in game theory) may not be the best attitude in the long run or as far as the group is concerned. We believe that a present "missing point" in MAS that would yield societal surplus (in a sense resembling the Smithian notion of the society being more than the sum of its individuals) is an emotional stance to be amalgamated with the present rational/intentional one. Further, we shall remark on the consequential view that the notion of agent autonomy and others related to it deserve reevaluation. Both points are argued for in Section 8.3. That discussion is an important consequence of the work in this chapter, thus fundamental for a proper understanding of the motivations for it. In short, convincing the DAI research community of the value of moral sentiments for agents is important in opening ways for the feasibility, in the future, of an ambitious use of anthropologically-based adaptation of migrant agents in which the only (or at least a fundamental one) cross-cultural link, as in human societies, is formed by emotional attitudes and motivations.

Chapter 8

Discussion

This chapter aims at presenting all the discussions related to the main points made in the thesis; it encompasses our views on several aspects of MAS in the perspective of the work described here and at times mentions the limitations of our approach. It also relates our main ideas to other work in the area, including some critiques of that work. In Section 8.1 we defend a form of interoperation of MAS which entails adaptation to different societies rather than making them all conform to a single standard; it is a contrast to current approaches to interoperation, on the basis of our anthropological ideas (Chapter 4). Section 8.2 discusses issues of ontological evolution and agent autonomy in that respect—this is related to our proposals in Chapter 5. Finally, Section 8.3 argues in favour of including emotional aspects in architectures of autonomous agents and against current conceptions of agent autonomy, on the grounds of the work presented in Chapter 7.

8.1 A Contribution to the Critique of KQML

We suggest that the idea of allowing an agent to join a society that was conceived in a different way than the society the agent was originally designed to belong to can be useful for interoperability (for issues on interoperation of systems see (Genesereth and Ketchpel 1994; Wiederhold 1994)). Justifications for interoperability are that an agent can profit from knowledge or other agents' capabilities that are not present in its own particular society and, conversely, that a society can profit from the visit of a foreign agent if it has any knowledge or is able to perform actions not available to the member agents of that society before the migration. We thus envision that Multi-Agent Systems (MAS) can interoperate through agent migration among disparate agencies. Further, we think that it is important from a social simulation (Castelfranchi 1990; Conte and Castelfranchi 1995) or a cognitive science (Green and others 1996) point of view that each society is allowed some freedom to evolve its own particular language and culture, rather than to oblige MAS just to comply with some kind of engineering standard in order to allow interoperation: cultural variety among agents creates a context for experimentations of interest in those areas, insofar as it reproduces several phenomena that are important in human cognition, e.g., learning of foreign languages. Accordingly, research on conceiving agents that can adapt to the intrinsically very different kinds of "cultures" of agent societies that can possibly exist (we have provided some examples in Section 4.2) is needed.

Further, if we consider that agent societies are already being designed by different people using varied approaches, and that societies that were originally designed in similar ways may evolve autonomously[1], we cannot expect conventional ways of social interactions to be a sufficient means to achieve interoperability of multi-agent systems, as intended with the Agent Communication Language (ACL) from the ARPA Knowledge Sharing Effort (KSE) (Genesereth and Ketchpel 1994); recall that we briefly introduced its main components in Section 2.2.3. However, the opinion that its component KQML could be used as a "universal interaction language" (Labrou and Finin 1994) has been almost a consensus[2] (cf. also (Cohen and Levesque 1995)). In contrast, an anthropological approach to MAS, a substantial enterprise to which this thesis aims at making small contributions, would establish the grounds for agents to migrate between *different* societies, which is clearly a way towards MAS interoperability without the need for standardisation. In this section we present a critique of the current major approach to interoperability in MAS, the KSE's ACL (Genesereth and Ketchpel 1994).

Undoubtedly, KSE's ACL is by far the most advanced proposal for agent communication at present. Many projects are being developed using this approach to agent communication and it is on its way to being standardised by American organisations. More recently, there has been another large effort to create standards for agent technology in the form of a non-profitable organisation called Foundation for Intelligent Physical Agents (FIPA) (FIPA 1998); for the specification on the communication aspect particularly, see (FIPA 1997). Some of FIPA's ideas can also be found in (Steiner 1998). These and The Agent Society (The Agent Society 1998) were all mentioned in Section 2.2.3.

Nevertheless, Cohen and Levesque (1995) argue that KQML needs further semantic clarification (which is actually agreed by the KQML authors (Mayfield, Labrou, and Finin 1995; Labrou and Finin 1994)), so that designers can have a precise idea of the meaning of perfor-

---

[1] Human analogues are the cases where a society that shares a culture is divided and isolated, e.g, because of environmental reasons, for a long period of time. There are examples of such cases or interpretations of historical evidence in (Munch 1971; Langdon 1975).

[2] In more general terms, researchers recognise the problem of agent migration but can only see standardisation as solution for it: "A further major impediment to the development of **open** multi-agent systems is the lack of standards for basic agent features like agent-level communication, and agent-level brokering and directory services." (Jennings *et al.* 1997; their emphasis).

matives. More recently, a semantics for KQML was given in (Labrou 1996). However, the semantics is given in the form of preconditions (for the use of performatives) and postconditions (i.e., what should be the state of the world after the use of the performative), which still has been criticised as an unsuitable form of semantics for an agent communication language.

Not only do we agree with that particular critique presented in (Cohen and Levesque 1995), but we also should like to add that we have strong reasons to believe that this semantic formalisation that KQML researchers themselves feel is necessary can be used as a means to allow agents to have quite different languages/cultures and still be able to interact. This can be done in much the same way as proposed for language adaptation using formal, meta-level information in (Bordini 1994), see Chapter 3.

Using universal meta-languages rather than universal communication languages (Bordini 1994) brings up a problem similar to that of distributed interpretations (Gasser 1995), where conveying the semantics of a message together with the actual message to be sent to some agent is not enough as it may become necessary to send the semantics of the language used to give the previous semantic interpretation, and so on[3] (this is referred to as the problem of "keeping meanings and representations stable across space and time" in (Gasser and Huhns 1989b)). Although it may seem that we are just shifting the problem "one level up" and making the standardisation at the meta-level rather than the communication level itself, we believe that this is not true because we are proposing that agents use the upper bound (apart from natural language) of that sequence of meta-levels, i.e., a level that we humans use, and the research community itself recognises as necessary for our understanding when new languages or agent theories are created (Cohen and Levesque 1995; Mayfield, Labrou, and Finin 1995; Luck and d'Inverno 1995): the level of languages for formal specification. Further, this sort of meta-language can be used to specify other parts of the target culture besides communication features (e.g., social organisation and norms), so if standardisation were unavoidable it would be more convenient to standardise at the more general level than on every specific aspect of a MAS architecture. Finally, it is important to emphasise that, in the case of agreement at the meta-level, there is not even a need for one universal standard: any formal specification language can serve as a meta-language, insofar as a computational interpreter exists for it and that agents can have access to that interpreter.

The cognitive approach to anthropology which is the basis of our ideas on anthropological migration has an interesting observation on the reasons for the differences in languages or in the

---

[3] But note that this is *not* an indefinite regression, as we see in the next sentence. The upper bound is, of course, natural language.

same language with the passage of time (which will eventually apply to MAS too): "Just as there is no inherent quality in an object that forces us to perceive it in exactly one way [this concerns the classification or grouping of objects perceived as 'similar;' refer again to Section 5.2.3], neither is there an intrinsic characteristic associating an object with its name. Consequently, with the passage of time, a class of objects may be renamed, but the class of objects denoted by this name does not change, or, conversely, the class of objects denoted by a name may change, but the name does not." (Tyler 1969). (This is, of course, additional to the obvious case of names changing due to changes in the environment). As we have mentioned in Section 4.2 (see also (Bordini and Campbell 1995)), anthropologists of the cognitive school try to provide each culture with a particular theory, which is derived from its particular formal description. Therefore, our idea of using a meta-language to describe societies of agents formally seems to be akin to cognitive anthropologists' opinions, to the extent that they claim that "Only when such particular descriptions [of cognitive systems] are expressed in a single metalanguage with known logical properties will we have arrived at a general theory of culture. Such a general theory will be equivalent to the language in which we describe cultures." (Tyler 1969).[4]

On the whole, there does not seem to be anything special in ACL that would make it the only language to be appropriate or attractive for use by all designers world-wide, despite its extensibility and portability features. There is not much difference in terms of expressive power between that and other approaches to agent communication based on speech acts (e.g., the ones in (Demazeau 1995; Campbell and d'Inverno 1990; Cohen and Levesque 1995; Singh 1994)). Furthermore, it does not seem that research in the field has reached the point that the ultimate language for agent communication can now be established in an attempt to make all designers adopt it. Due to the nature of this domain, such a target may prove unrealistic even in the future. Besides, technologies that have been made standard to fit perfectly together, in practice do not succeed in all cases; such approaches may therefore turn out to be a limitation on the conception of MAS and still not solve the problem it is intended to solve. Also, it is the immigrant agent's obligation to speak the particular language used in the target society, as happens with anthropologists (see Section 4.3.3), rather than, for example, their communications all being translated into KQML. This is important from the "social simulation" (Castelfranchi 1990) point of view that we have mentioned earlier and increases the interest of cognitive science for DAI simulations. Further, we think that having different cultures and languages increases agent

---

[4] This is precisely the point that makes this school of thought in anthropology highly controversial at present, but as we have mentioned before these ideas might make sense in computer science. To have a common language for the description of cultures would indeed be interesting, if at all possible (at least for agent systems); it would be an important step forward in MAS, as we do not have anything near that now.

believability (Bates 1994), because one could identify "local" and "foreign" behaviours of agents, and because the adaptation to the different cultures also creates believability in the sense that everyone appreciates (and can make allowances for that fact) that a foreign agent faces cultural problems, which are likely to compromise its general performance until an adaptation process has been completed.

It is at least reasonable to remark that the mediation (Genesereth and Ketchpel 1994; Wiederhold 1994; Huhns 1995) approach for interoperations of non-agent software (i.e., legacy systems) is certainly a good alternative for interoperation of this kind of software. Our own approach is particularly concerned with interoperation among different societies of autonomous intelligent agents. Further, a standardisation on KQML dialects can be useful for some current applications, e.g. information agents accessing heterogeneous information with wide resources(Klusch and Shehory 1996), as in the Internet(Decker *et al.* 1997). We claim that for the community of Multi-Agent Systems in particular we can profit from richer simulations of agents that do not agree in language and culture, which is indeed expected for the sort of complex agents mentioned in (Castelfranchi 1990; 1995).

There is an issue which we have considered in our conception of MAS that makes things rather more complicated. The fact that agents may be able to change their languages (a possibility already foreseen, in regards to protocols, in (Populaire *et al.* 1993)) and culture, as in the case of human evolution, may cause some difficulties with the formal descriptions that agents are supposed to have available when migrating, as they may not be up to date. That is the reason why we proposed in Section 4.2 the existence of *anthropologist agents* which should ideally be able to generate and update formal descriptions of languages (resembling ethnography of communication) and cultures exactly as social anthropologists do (one part of the former consists of the description of the local ontology, e.g. as seen in Chapter 5). Accordingly, we believe that ontologies cannot be merely specified by MAS designers (as in the ACL proposal) but agents should be able to modify their own ontologies in an evolutionary sort of way; if they are cognitive autonomous agents there is no reason why they should not be able to do so. Consequently, because of autonomous evolution of societies, it is highly desirable that the anthropologist agent is able to *discover* ontologies (and taxonomical relations of the terms therein) used in MAS (see next section).

Finally, by agreeing with Cohen and Levesque's criticism (1995) that KQML is in need of semantic formalisation, and with the proposal in Chapter 3, one can conclude that KQML is in an autophagous[5] situation in this point of its evolution. If, on the one hand, KQML needs further semantic formalisation in order to become widespread, providing it with such a formal description, on the other hand, may come to downgrade its claim to be a *universal language* that is thus capable of promoting interoperation (Labrou and Finin 1994; Genesereth and Ketchpel 1994), since agents can then treat it as just one more language to be learnt/used. It is of course quite possible that KQML will be one of the several different languages agent societies will use (while still being able to interoperate), perhaps the most widely used in the near future.

For all these reasons we have presented, it is not unlikely that the KQML approach (or, better said, approaches that aim at standardisation) will turn out in the future to have been the wrong way for MAS, if a gamut of mechanisms for cooperation and communication among intelligent agents is to be explored. On the other hand, we emphasize that this statement does not eliminate the merit of KSE's ACL as (probably) the first widespread technique for complex (speech-act-based) agent communication.

8.2 On Agent Autonomy for Ontological Evolution and Management

The first thing to consider in terms of comparison of our work on ontological ascription (Chapter 5) with other approaches to ontology, is that a limitation of our approach is that it concerns the use of individual constants (i.e., terms) only. The suitability of our approach for other sorts of language units (e.g., representing actions rather than individual constants) remains to be studied. Also, it could be claimed that the expressive power of traditional approaches to ontology is greater, since each term has a logical statement associated with it rather than just a set of predicates.

On the other hand, our approach is, to the best of our knowledge, the only approach to consider that ontologies are not designed by programmers or designers of MAS, but that agents may change them evolutionally. More importantly, our approach aims at allowing interoperation of MAS even if we consider that ontologies are neither fixed nor provided in advance by designers, by allowing agents to ascribe ontological descriptions to societies of agents. Further, if we did compromise on expressive power of our approach, this was so in order to provide for its generality: as can be seen, even for very simple informant agents the approach still applies. Additionally, our approach is well-founded, being based on a theory from Pragmatics.

Concerning the formal specifications we presented in Section 5.3, we have mentioned that

---

[5] By *autophagous* we mean the property of something that progresses by consuming or even destroying itself in its previous form.

an Intensional Ontology of Terms (IOT) maps a term to a set of predicates (any set), even though Definition 5.4 (see Section 5.2.2.1) clearly says that the proper definition of a term is precisely the set of predicates that are necessary and sufficient to distinguish the term from every other term in the universe of discourse of a particular society of agents (unless it has synonyms). In the formalisation in Section 5.3.2, however, it is not assured that every set of predicates in the range of an IOT conforms to this minimal-set criterion. Therefore, an IOT ascribed to a society of agents is not guaranteed to be either complete or sound by the the anthropologist agent. It is simply tentative, and the anthropologist agent should keep studying the society in search for evidence that it is correctly built[6]. This should be performed by observing the communication in the target society (i.e., the same process that we would expect to be used for the generation of the set of sentences to be given to the informant agents in the interview). Therefore, further work is needed on the generation of the sentences used in interviews of informant agents, and on their revision in case the anthropologist agent finds that a wrong definition has been ascribed. On the other hand, an ontological mismatch (between actual and ascribed ontologies) may be the consequence of an actual evolution of the ontology used by the agents, in which case a new ascription has at any rate become due.

The same situation happens in regard to the recovery of taxonomical relations as presented in Section 5.2.3 and formalised in Section 5.3.3. The taxonomical relations generated by an anthropologist agent from an intensional ontology it has ascribed to a certain society should be regarded simply as "clues" to what the actual taxonomical relations in that society are. Furthermore, the process of recovery of taxonomical relations should be followed, as is the case with ethnographers, by a confirmation of the relations that were found, e.g., by the anthropologist agent interviewing the informant agents again. This time, instead of the acceptance of sentences, the enquiries to be made should resemble the way ethnographers systematically test taxonomies, by asking questions expecting a negative reply (see Section 5.2.3). Better still, the anthropologist agent could ask whether the informant agents accept that two terms denote contrasting segregates or whether they have an inclusion relation (of course, this would increase the abilities that are required from the informant agents; they would need to be more sophisticated agents so as to understand concepts like segregates, contrast sets, inclusion, etc.). The formalisation of this confirmation process is planned as future work (see Section 9.2). As is the case with ontologies, taxonomical relations also may evolve, so that the anthropologist agents must

---

[6] In the case of synonyms (i.e., two terms that have been ascribed the same intensionality), for example, it is possible that details about one of the words are still missing, and the terms are not actually synonyms. We have assumed that such process of verification had been accomplished and therefore we were justified in assuming that terms with equal (ascribed) intensionalities were indeed synonyms.

be attentive to possible changes (in effect, changes in taxonomical relations may simply reflect changes in intensional ontologies).

Presently, our concern with ontological and taxonomical ascriptions aims at solving the problem of migrant agents being able to use the communication language of the target societies. However, it is widely accepted among anthropologists that a taxonomy of terms is quite revealing about the peculiarities of a society, as we saw in Section 5.2.3. Therefore, further research linking social anthropology and MAS should provide useful techniques for migrant agents to use the kind of information we considered in Chapter 5 for a more thorough understanding of the target societies, rather than just for linguistic compatibility.

As for the ascription of immutable ontologies, one remark should be made. The amount of time over which an anthropologist agent has been studying a particular society is not taken into account when generating immutable ontologies (in the formalisation presented in Section 5.3.2). Therefore, an immutable ontology can be ascribed even if the anthropologist agent does not have long-term information about the society, in which case the immutable ontology is not guaranteed to be completely reliable (in the sense that it has been an immutable ontology for such short time that this fact may not be worth taking into consideration, as it may come to change in the near future). In order to determine how long it takes for the (reliance on the) use of an ascribed immutable ontology to be advisable, a study with real-life applications using MAS should be conducted, but there is certainly no harm in providing this type of ontology to migrant agents even in early stages of an anthropologist agent's work, thus our provision for it.

The research method used in Chapter 5, the reader cannot have failed to observe, is basically one of providing formal specifications rather than "informative experimentation," which is more usual in research on artificial societies, which is the area where those topics are more familiar. On the same lines as (Conte and Castelfranchi 1995), *formal* here means that we deal with abstract aspects of the subject studied; further, it means "explicit, consistent, controllable and unequivocal (or almost unequivocal) ways of expressing thoughts." Although formal methods for the presentation of research ideas are well accepted in artificial intelligence, some still seem to think that they should be used only for the purpose of theorem proving. Is support of the opposite view, with which we agree, we quote again Conte and Castelfranchi 1995: "In any case, we challenge the idea that (logical) formalism is fundamentally aimed at theorem proving. ... A primary objective ought to be to provide good concepts, clear notions, and heuristic categories of description. Logical instruments are essential for concepts and categories of description to be defined in a rigorous, explicit, concise and controllable way." They further say that they are concerned with "producing computational models, and therefore constructing systems performing tasks in accordance with theoretical expectations," for which the formalisations are clearly indispensable.

Further, we support the view that research on Artificial Intelligence (and also Computer Science in general) needs both theoretical support and practical testing, and clearly the theoretical work must come first. There have been plenty of "experiments without theory" and this has not been good for the subjects. Also, formal specification is certainly more on the theoretical rather than experimental side of work, but it has also an element of experimentation in it. Not only can one refine the theory and correct mistakes in initial intuitions (those familiar with formal work know how inevitable this is), but also it points out the parts of the system being built that should be focused on in subsequent experimentation—the parts that proved difficult to produce to one's satisfaction in the specification are the ones in question. Our specification can therefore be regarded partly as an (another kind of) experimental activity.

Concluding the argument on our research method, as for (Conte and Castelfranchi 1995) the main purpose of our specification is in respect of rigorous presentation of ideas; but also, because of the use of the Z framework, we were able to provide (directly) a computational model, which is important in the context we are working on, namely MAS. As we have mentioned in Section 5.3, a simplified version of the specifications presented there has been animated using PiZA (Hewitt 1997). This allows us to verify that the specifications are correct (i.e., that the computations they generate produce results in accord with the theoretical definitions). Besides, it provides a computational tool which can be used for tests in several other domains so as to check the generality of our approach. The mechanisms used for the ascription of a cricket ontology (in Section 5.4.2) and for the derivation of the cricket taxonomies (in Section 5.4.3) do not make use of any semantic features specific to cricket, so it is reasonable to expect them to be useful in any domain. At any rate, the approach of using Z specifications not only makes our definitions precise, but it also provides us with such a tool for further verification of the (generality of the) approach.

Another reason for our using, in particular, Luck and d'Inverno's formal framework (1995) based on Z, is that we can rely on specification of other features of autonomous agents (which are necessarily complex systems) being made by other researchers, allowing us to concentrate on the particular problems we have set out to investigate. d'Inverno and Luck aim at providing a unifying framework where effort in different sub-fields of MAS can be harmonised (d'Inverno and Luck 1996b; 1996a); an ambitious task that is arguably an important current issue in MAS. They claim that Z's modularity and abstraction can be helpful in that enterprise (Luck and d'Inverno 1995; d'Inverno and Luck 1996a; 1998). Further, on other general advantages of

Z itself, they mention its sufficient expressiveness, suitability for moving from specifications to implementation of computational prototypes (as we have done, as commented above), and the availability of supporting books and tools, among others.

As for the case study presented here (Section 5.4), even though the peculiarities of the cricket ontology allow us to demonstrate the nature of our approach to ascription of ontologies and recovery of taxonomical relations, it has, of course, been scaled down in order to serve this purpose here—it is, as usual in AI work, a "toy problem." Issues of scalability to real-world MAS application remain to be addressed. Also, one could consider in the future issues of scalability to human languages (rather than agent communication languages as intended here): anthropologist agents could be conceived to work as "*lexicographers*," supporting the creation of dictionaries (of human languages).

This part of our work makes direct use of anthropological techniques to solve problems in DAI. Evidently, DAI techniques can be used for studies in anthropology (as indeed for all social science)—see, e.g., (Doran *et al.* 1994; Doran and Palmer 1995), which makes use of *artificial societies* to study the emergence and perpetuation of hierarchical societies. This thesis aims at showing that the converse also holds: anthropology can provide the basis for resolving open research issues in DAI. Fortunately, Doran's own work presents further evidence in support of our claim. His idea that *collective misbeliefs* can be beneficial for a society of agents (Doran 1998a; 1998b), which clearly was inspired by his longstanding work on computer simulation of societies supporting anthropological and archaeological models, can have a decisive impact in future MAS. Although the idea has not been used in other (more general) multi-agent applications, it seems to be a contribution to MAS design that has a quite general appeal, and is therefore relevant to this thesis. Doran's work recognises, as ours does, the importance of emotional aspects of agents in explaining certain social phenomena (Doran 1997): this is connected to our discussion in the next section.

We remark, *en passant*, that the activity of an anthropologist agent ascribing ontologies to communities of agents could be regarded as the computational counterpart of work on the multidisciplinary subject of ethnography of communication (see, e.g., Hymes 1977 and Bauman and Sherzer 1989). Therefore, yet another field of the social sciences which has been neglected so far with respect to DAI, despite being of relevance to the area, as the most recent work on the "contribution" this section refers to demonstrates, is ethnography of communication.

In addition, it is interesting to note that the theories present in both strands of social thought used in Chapter 5 (on the anthropological and the semantic/pragmatic sides) follow philosophical principles that are familiar to *logical positivism* (or *logical empiricism*). It is currently argued that such theories are dead for social sciences because they fail to comply with the intrinsic "human" aspects of the problems involved. However, it seems they still have contributions to make to the computational counterpart problems in DAI, where the formal aspect is essential. Thus, peripherally, our thesis also suggests that it is helpful to make further investigation of fairly old work on social sciences as sources of inspiration for what have been called (Gasser and Huhns 1989b; Gasser 1991) "open problems in DAI." However, recall that we have agreed with Gasser (1991), as we stated in (Bordini and Campbell 1995) and in Section 4.2, in that theories from the social sciences more recent than those provide the basic principles underlying the appropriate conception of DAI as an inherently social one; these are clearly inspired by *philosophical pragmatism*. This philosophical compromise between opposing schools of thought (logical positivism and philosophical pragmatism) seems to indicate that in MAS it is a matter of finding the best inspirations for each problem in the most relevant approaches of the social sciences.

8.3 Let Agents be Benevolent

In Chapter 7 we have argued for the positive effects of altruistic behaviour, which in autonomous agents can only happen (without philosophical objection as to whether true altruism is involved) through the inclusion of emotional aspects in their architectures (towards which we will argue in this section). One of the motivations for that part of our work is our belief that, if in the future MAS makes extensive use of concepts related to agents' emotions, agreement on emotions (and perhaps perceptions is another interesting level of inter-cultural homogeneity) would allow one to further the studies on forms of interoperability of MAS that avoid the need to standardise on other aspects of agency. We draw this lesson from social anthropology too: one of the things that an anthropologist can be sure is present in all cultures is the whole spectrum of human emotions; Ridley (1996) says that they are *universally recognisable*, and even motives such as avoiding guilt are common across all cultures. Therefore, an *anthropologist agent* (i.e., an agent responsible for supporting migrant agents in their adaptation to "strange" societies) would benefit from being able to assume that emotional motives are present in agents from all *target societies*. This is because emotions would influence the behaviours they display, their social conventions, etc., which in turn are of interest to the anthropologist agent. That is, a culture would be comprehensible to the anthropologist agent because an understanding of those peculiar conventions would be possible at the emotional level.

Further, our view is that an *emotional stance*[7] is yet another "missing point" (Castelfranchi

---

[7] With this expression we wish to convey our position that emotions, and not only rationality (on which the work on MAS so far has been based), are an important part of the conception of agents. Dennett says of his *Intentional*

1990) in present MAS. It should not be considered only for its role in believability, which is the point that most works that consider emotions as being relevant to agents make (see, for example, (Rizzo *et al.* 1997; Bates 1994; Hayes-Roth, Brownston, and van Gent 1995)). It is evident from the results shown here that to behave altruistically (even in a prisoner's dilemma context, where *rationality* has always been paramount) can prevent agents (of whatever paradigm) being *rational fools*, for it secures a good performance in the long run for the agents and in the more extreme cases for their social groups, at least. (Note that, metaphorically, a social group can be seen as a complete MAS, wherever the notion of group is not available.) This leads us to question the current widespread idea in MAS that autonomous agents should necessarily be selfish. Ridley wittily says that sociobiology "caught the self-interest virus" in the 1960s. It can be said that MAS has caught the same virus too.

Apart from the whole tradition of work on utility maximisation in the game-theoretic approach (Rosenschein and Genesereth 1985; Rosenschein and Zlotkin 1994) where agents are by definition self-interested, among the representatives of the "autonomy as selfishness" outlook in MAS are d'Inverno and Luck (1996b) who have claimed:

> Cooperation will occur between two parties only when it is considered advantageous to each party to do so. Autonomous agents are thus selfish agents. A goal (whether traditionally viewed as 'selfish' or 'altruistic') will always be adopted so as to satisfy a 'selfish' motivation. (page 529)
>
> The effects of benevolence are possible, but only through self-serving motivations. (page 533)

In terms of the history of MAS, this line of thought seems to have started with the the discussion on *social power* in (Castelfranchi 1990). The paper was in the right direction at that stage, when benevolence was being taken for granted. It is time now to relocate benevolence, not as something taken for granted, but as an important phenomenon which may *evolve* in societies of autonomous intelligent agents from explorations of agent emotions; we also argue that it is in the agents' best interest to do so. This issue in particular should be of interest at least for those concerned with agent autonomy via the cognitive (as opposed to utilitarian) view of the field (cf. (Conte and Castelfranchi 1995)).

The misconception about autonomy and benevolence goes together with the absence of an explicit emotional component in present MAS theories. Motivations do not have to be nec-

---

Stance (1987): "Once the intentional strategy is in place, it is an extraordinary tool in prediction..." and "The first answer to the question of why the intentional strategy works is that evolution has designed human beings to be *rational*, to believe what they ought to believe and want what they ought to want" (our italics). Without any presumption to have a philosophical theory about this, we think it is possible that the intentional strategy in some cases works (i.e., it allows us to predict the behaviour of *intentional systems*) because humans have moral sentiments, not because they are rational. Dennett is an important philosophical influence on the foundations of MAS, and rightly so. Our point is only that the role of emotions has been overlooked in MAS.

essarily self-serving. Consider, e.g., the idea of terminal interest adoption defined by Conte and Castelfranchi (1995). They mention the possibility of autonomous agents adopting others' interests out of affection, although they do not concentrate on the emotional aspects of agents in that book. On the other hand, they say that the usual means by which an agent can act in a selfless way, i.e. adopt a goal of another agent, is through an individual (personal) goal of being benevolent (compassionate) towards that agent. Further, they state that truly benevolent agents are those who undertake a mode of goal adoption they call *terminal*, which they claim to be the type of adoption common since the early days of MAS (i.e., it is *assumed* that agents will adopt each other's goals; essentially, these agents are not fully autonomous). This is, therefore, the perfect ground for d'Inverno and Luck's definition (quoted above). Indeed in this context there is a point in saying that when an agent is being benevolent, either it is not autonomous or it is pursuing an individual goal and thus its motivation is ultimately selfish[8]. The problem is exactly in the artificial mechanism of modelling benevolence as an individual goal *per se*, due to the lack of representation and processing of "emotion." Conte and Castelfranchi themselves have mentioned that a selfless action could be derived from an impulsive (reactive) behaviour[9]. Clearly, we need to distinguish emotional processing from the (now) orthodox rational/intentional one, and as a consequence the cynical[10] definition of autonomy can be dropped. Recall that an important consequence of the emotional stance is the ability to truly exploit societal interaction in MAS, according to our discussion on the role of emotions in human social nature (see Section 7.2 and references given there). In brief, we suggest that emotional[11] processing should be balanced with rational processing in the decision-making process in autonomous agents (recall our referring to Plato's idea that what he calls "high spirit," which relates closely to emotions, is reason's "natural helper"). We are proposing, in effect, a more comprehensive view of rationality in MAS.

---

[8] Ridley comments that this idea is present in philosophy at least since the work of the philosophers of the "Edinburgh" school. If the motivation for an action is selfish (e.g., one is charitable because it makes one feels better), then surely one cannot speak of benevolence? Nonetheless, this is not always accepted: one might consider the effect of the action itself, rather than the motivation.

[9] We agree with this idea, but note that although emotional processing is of a more "impulsive" nature than the rational one, this does not mean to say, in our view, that it is reactive in the usual sense in MAS, or even "hybrid" for that matter; emotions are more naturally a part of cognitive rather than reactive agents.

[10] The word is used here as it is used by Ridley for the selfish-gene approach to sociobiology, which (correctly) takes the altruism out of altruism as far as animals are concerned, particularly in the (often romanticised) case of "altruism" among hymenoptera.

[11] It is important to alert the reader to the fact that, clearly, when we refer to "emotion" in a computational agent, we mean it just as a metaphor, in the same way that intelligence is as metaphor: the interest there rests in producing agents that behave *as if* they were in fact intelligent (whether this is or is not intelligence is a long, unresolved philosophical debate), and the same applies to emotions. The particularities of whatever computational mechanisms the research community devises that indeed produce that behaviour are not significant for the argument here. It is plausible that it will turn out to be an *extended* logical model. The fact that it is computationally viable is demonstrated by the examples of agent architecture involving emotional processing that we cite later on (even though it is true that there are only a few and preliminary attempts so far).

We appreciate that social conventions (or *norms* as Conte and Castelfranchi call them) are culture-specific. "But our cultures are not random collections of arbitrary habits. They are canalized expressions of our instincts. ... That is why, for all their superficial differences of language and custom, foreign cultures are still immediately comprehensible at the deeper level of motives, emotions and social habits." (Ridley 1996). Accordingly, Conte and Castelfranchi's notion of norm and their significant attempt to solve the micro-macro link problem (1995) could profit from an emotional stance as much as our anthropological approach to migration of agents does. It is possible that our emotional stance may, if properly pursued, help to resolve the paradox of individuals being self-interested yet displaying societal behaviour: it may be an alternative for the dichotomy between the utilitarian approach (which is how Conte and Castelfranchi refer to the game theory approach to MAS, e.g. (Rosenschein and Zlotkin 1994)) we have already criticised here and the Hobbesian approach (e.g., (Conte and Castelfranchi 1995)) which has to resort to "Leviathan" to ensure cooperation and resolve disputes.

To round off the argument, agent autonomy is not necessarily concerned with fulfilling an agent's own selfish motivations; it has to do with its freedom to *choose* how to behave (or to have the resources to do something by itself, depending on the context)—even if that means choosing what is not best for its own (present) goals, e.g. under the influence of emotional decisions (as we have remarked, this requires a specific part of the agent's architecture to deal with emotions). We observe in the results of our simulation that altruism prevents agents from being rational fools. As Ridley puts it, when being truly altruistic, (i.e., doing something in the interests of others at one's own expense and even with no future reward at all), we are giving way to emotions which are an important mechanism behind the complex brand of social interactions that humans experience. It is certainly worth recovering this notion for the benefit of MAS.

One of the advantages of the simulations we present here is that it allows us to verify some of Ridley's propositions; it also yields some insights into the issues discussed above. Note that, whilst the quantitative results of the simulations presented here apply to MAS where the IPD metaphor (with the further assumptions we have made here, e.g., the existence of groups) makes sense, Ridley's line of argument, presented in Section 7.2, seems to apply quite generally; it is a high-level argument in support of our suggestion that designers of general[12] MAS architecture should consider an emotional aspect as well as the traditional rational one, as we assert in this thesis. We do not aim at proposing a specific implementation of emotions in an agent architecture. There have been only a few attempts to do so in general agent architectures so

---

[12] By general we mean architecture for all sorts of agents and not only those that have a specific relation to "believability"; see references for this area given above.

far, but see (Aubé and Senteni 1996; Sloman and Poli 1996). (Sloman and Poli's SIM_AGENT and Tok, an architecture for believable agents, are mentioned in (Müller 1999); we have given further references in Section 2.3.)

Our ideas on moral sentiments relate closely to issues of *formation* of norms (conventions) among autonomous agents (Conte, Castelfranchi, and Dignum 1999). It is easy to see that emotions are important in attaching agents to social norms—Doran mentioned in (1998b) that emotions may be an important mechanism for designers of societies to "manipulate" agents, e.g., to maintain their collective misbelief in "cults" (Doran 1998a). However, in a contrary view, it is plausible that just as emotions are the right basis for seeing that autonomous agents can be *autonomous* and still behave (truly) altruistically as we have argued above, studying emotions in agent architectures can help us understand how autonomous agents themselves form and perpetuate conventions, which are essential for social behaviour. These are interesting issues to be addressed in further research. Similar motivations to (Conte, Castelfranchi, and Dignum 1999) can be found in (Ossowski and García-Serrano 1999), which relies on both a sociological and an economic approach; they are concerned with social action without compromising on agents' autonomy. This question too seems to be connected intrinsically to some notions of emotions or moral sentiments. In summary, moral sentiments are involved both in: (i) how to ensure or encourage adherence of agents to social norms (i.e., conventions), given that emotions can be instrumental in arranging that suitable rewards (for a reputation for compliance with norms) and penalties (for non-compliance) are applied to individual agents (Castelfranchi, Conte, and Paolucci 1998); and (ii) the very formation of norms by autonomous agents (Conte, Castelfranchi, and Dignum 1999).

In a recent paper, Castelfranchi, de Rosis, and Falcone (1997) recognise the importance of emotions in domains other than the usual believable-agents one, which seems to support our argument. When commenting on the several reasons for agents needing personalities they include *Social/Cognitive Modelling*, as follows:

> Since in nature and in society agents have personality and this seems an important construct in psychology, one might aim at modelling personality in agents (or emotions or cognitive biases) to reproduce relevant features of human interaction. (page 16)

We believe this should be an additional source of guidance for the work on MAS as originally defined as the field of Distributed Artificial Intelligence concerned with coordinated intelligent behaviour among a collection of autonomous intelligent agents (Gasser 1987). In other words, if we build agents' rationality inspired by the human counterpart but fail to provide them with other important human mechanisms such as emotions, we shall find that we have built rationally foolish agents, which will be no more useful to their collaborators than rational fools are to human societies (with analogues of all the undesirable consequences to the rational fools individually too).

Chapter 9

Conclusion

9.1 Summary of the Thesis

We present our anthropological approach to migration of agents as a new way of dealing with some basic questions of DAI. It implies the relevance for the foundations of DAI of a discipline not normally considered in this context, namely social anthropology. We suggest that it should be among the disciplines of interest to MAS if promoting interoperation of disparate agencies (which we argue to be more suitable for the area than standardisation) is to grow into a substantial MAS topic. This is our underlying thesis here. We now summarise the main points and results from this thesis.

In Chapter 4, we presented some concepts from Cognitive Anthropology and techniques used by field workers in Social Anthropology which contribute, in several levels of abstraction, to the conception of agents that can migrate among different societies. We have, accordingly, recommended that Anthropology should figure among the DAI disciplines when one considers the problem of interoperability among potentially very different multi-agent systems. We also discussed a conception of situated learning which involves anthropology, as well as sociology and psychology, as a comprehensive conception of learning that has important implications for MAS.

In Chapter 5 we presented a new way of specifying ontologies used in societies of agents based on a theory of intensionality, which matches our anthropological approach to migration of agents. It allows intensional ontologies to be ascribed to societies of agents, and is thus instrumental for the agent adaptation approached in this thesis. Further, inspired by work on ethnography, we presented a means for an anthropologist agent to recover taxonomical relations from the intensional ontologies it has ascribed to societies of agents, and we have formalised it along with the ontological ascription process (recall that the generated taxonomies are augmented with the classification criteria for each segragate). This is also an important aspect to figure in anthropological descriptions of MAS. We have ascertained our approach through a case study on an ontology and taxonomies from a non-trivial ball game. The simplicity of the approach is important in terms of generality: in order to allow interoperability among the largest possible number of different approaches, the less that is required from the informant agents, the better. Still, the approach proved to be expressive enough to be effective even for the exotic culture of cricket.

In Chapter 6, we described a simulation of the game of cricket where one the players is a "foreign agent." The simulation is based on simple kinematics and geometry but displays a wide range of interesting cricket episodes, which can be observed both via its graphical interface and by means of a symbolic game summary. The foreign agent adapts by means of the ID3 learning algorithm, and its behaviour in the game is reasonable (or at least "safe," in the sense that it, e.g., minimises the risks of the foreign agent not acting when action from it is necessary). The foreign agent has displayed a reasonable behaviour related to both ball-trajectory interception and targets of throws, which are fundamentally useful in cricket. This suggests that even the oldest and most basic learning algorithm (plus some general kinematic knowledge) can have results that are acceptable from the point of view of both the foreign and the "captain" agents (i.e., migrant and local agents), and this is an important support for our "thesis" about the feasibility of foreign agent adaptation (as a means towards interoperation).

One situation where to engage reasonably in the practice of a community (i.e., target society) can be important for the task of a foreign agent itself is that of an anthropologist agent: recall the concept of *participant observation* from anthropology, on which we commented in Section 4.3.1 and is particularly important for the task of an anthropologist agent when studying a society of autonomous agents, given that they may (specially in current conception of autonomous agents—cf. our discussion in Section 8.3) refuse to cooperate (as informants) with the task of the anthropologist agent if it does not make itself useful for the community in some way. Further, a good aspect of our choice of cricket for this simulation is the variety of possible levels of complexity, hence a great deal of experimentation can ensue, including experiments on the ideas of learning as legitimate peripheral participation (this idea is briefly expounded in the next section).

In Chapter 7 we have presented some simulations based on the iterated prisoner's dilemma and on ideas on moral sentiments where some agents behave altruistically. Our results suggest strongly that rational fools maximise their performance in the short term but compromise their performance in the long run, while altruists show exactly the opposite behaviour. The results also show clearly that the more altruists there are in a group, the better they, and the group as a whole, perform; more importantly, their generosity, although somewhat "irrational" from an individual point of view, implies that the group as a whole performs significantly better than when pure reciprocation is used.

All this work produced plenty of scope for discussion, which was presented in Chapter 8 (alongside our views on several issues of MAS in the perspective of this thesis, current limitations of our approach, and its relations to some other work in DAI). We have argued that an "emotional stance" is a missing factor in MAS which can improve the performance of societies of agents in face of their bounded rationality. Convincing the DAI research community of the value of moral sentiments for agents is important in opening ways for the feasibility, in the future, of an ambitious use of anthropologically-based adaptation of migrant agents in which the main cross-cultural link, as in human societies, is through agents' emotional attitudes and motivations. This has an impact on present (widespread) notions in MAS, especially the notion of autonomy. One of the main purposes of Chapter 8 has been to argue in favour of a form of interoperation of MAS which entails adaptation to different societies (to this end, the work on learning of protocol languages overviewed in Chapter 3 is instrumental) rather than making them all conform to standards: our anthropological approach contrasts with current approaches to interoperation.

9.2 Future Work

There are many ideas that occurred to us during the work on this thesis. It would be impractical to mention all, but we give some pointers to a few of them. First, the idea of using Case-Based Reasoning techniques (Campbell and Wolstencroft 1990) is a quite straightforward one; it is relevant to many of the problems of the anthropologist agent (we mentioned the creation of the initial theory needed for the interview for the ascription of ontological descriptions in Section 5.2.2.2; further comments are made below). Second, it would be interesting to investigate the necessity of minimal universals, possibly influenced by Schank's Conceptual Dependency Theory (1975) which has been somewhat neglected in the DAI community, despite its fame elsewhere in the history of AI. Last, there are the several types of agents, e.g. the "novice master" agent[1], mentioned in Section 4.3.2, which need to be tested in real applications to assess their practical usefulness.

We now proceed to mention specific projects for future work on some of the individual

---

[1] Although we have not made thorough use of the idea of the novice master agent in this thesis, we have in effect looked at some of its properties here, to the extent that the novice master agent shares in the characteristics of both informant and anthropologist agents.

"contributions" (Chapters 5–7).

Ascription of Intensional Ontologies and Retrieval of Taxonomical Relations

There are many improvements that we envisage for the process, formalised in Chapter 5, of ascription of intensional ontologies and retrieval of taxonomical relations of the terms in the ontology. We mention some below:

- Further investigation is needed for the process of observation of language use in a target society in order to gather the relevant predicates associated with the terms, so that the initial theory can be built by the anthropologist agent before the interview. It too should be inspired by anthropological methods in similar activities. This is an example problem where the use of case-based reasoning, mentioned above, can be investigated.

- Additional use of anthropological methods is also required for a formalisation of the process of confirmation that inferred taxonomical relations in an ascribed intensional ontology of terms is correct, as we mentioned in Section 8.2. We also mentioned there a study of the amount of time over which the anthropologist agent needs to engage in "participant observation" before an immutable ontology of terms become reliable. (We also mentioned, in Section 4.3.1, an interesting experiment on the time needed for an agent to generate descriptions for computational societies through participant observation, and thus on the relation between time-scales for the human and the computational processes.)

- Further studies in pragmatics are necessary to examine whether our approach can deal with components of a language other than "terms."

- More complicated types of interview can also be researched. If we make further demands on the cognitive capabilities of informant agents[2], then anthropologist agents can use more elaborate types of interview than merely asking for an "acceptance" (if methods to deal with complex queries were developed also for anthropologist agents accordingly). For example, all the so-called "wh" questions (what, where, who, why, when, how) can be relevant for an anthropologist agent, and a set of minimal useful question templates and conditions governing their use should be worth developing.

- We have mentioned *en passant* that, because anthropologist agents keep the whole history of ascribed intensional ontologies and taxonomical relations, "historian agents," or

---

[2] Recall that an advantage of the simple form of interview we devised is that it is so general as to be applicable to all sorts of agents.

"linguist agents" interested in "agent archaeology," could, in principle, study the evolution of societies' ontologies, which could be revealing about traits and progresses of those societies. Research on means of doing so would be an attractive multidisciplinary project.

- Clearly, testing our approach in other environments (i.e., besides our reduced cricket ontology) is an important step if it is to be used in real systems and not just for the sort of experimentation we report here (but the animated version of our Z specifications can be useful in doing so). We also mentioned, in Section 8.2, that besides testing our approach further in the context of agent communication languages, there is the potential for applications in the scale of human languages (e.g., building anthropologist agents that can help the task of human lexicographers—that of compiling dictionaries).

The Cricket-Playing Society

Appart from the straightforward extensions of the cricket simulation that we have mentioned in Section 6.5—e.g., adding further complications to the game and inferring regularities in the symbolic game summary—two major ideas of future projects were mentioned there, on which we elaborate briefly here:

- An exercise using ascription of intensional ontologies extended with taxonomical relations in the cricket society. In order to accomplish this, first the cricket simulation must be extended to account for a fairly complex type of communication between local agents. Then an anthropologist agent studying that society (e.g. to help a foreign agent), or the foreign agent in the game itself if endowed with such capabilities, could study the vocabulary used in the cricket society and eventually produce an (intensional) ontological description to that society, and extend it with taxonomical relations, by means of the technique we introduced in Chapter 5. Note that the "cricket ontology" is quite peculiar (recall that that is the reason we used it as a case study in Section 5.4), thus there is quite a domain of curiosities to explore in this research (again in interdisciplinary mode: it involves anthropology and linguistics, besides DAI).

- The creation of a new learning mechanism inspired by the idea of legitimate peripheral participation (see Section 4.4). The cricket society seems to be a good setting for the study of any such learning mechanism, insofar as the situated character of it, and the idea of increased access to practices involving greater responsibilities, both make sense in the cricket context. In accordance with Lave's arguments (1988), this would be a learning curriculum (see Page 61) specific to cricket, as learning is different in every single community, even though it has this general form of legitimate peripheral participation. The local agents (captain and other fieldsmen) would gradually instruct the foreign agent to move from one fielding position to another, progressing in terms of expected complexity of the local task, until the foreign agent had experienced all the roles in the game (or at least the roles in the fielding team), which would give the foreign agent a higher level of understanding of the game as a whole.

We have also considered in Section 6.5 the possibility of an anthropologist agent studying the relation of the behaviours in the game and the general culture of the agents in the cricket society (this is the ultimate objective of this part of our work). However, for the existence of a complex "local culture" that would serve such purpose, the agents playing the game would have to be fully autonomous, cognitive agents, implementing all the research concepts created in the area. These are some distance from complete and functional implementations in the present state of the art.

Moral Sentiments in the IPD and MAS

There are various other hypotheses to be tested and many variations and extensions of the simulations presented in Chapter 7 to consider. One example we mentioned there was the strategy for egoists attempting to join the group that maximises their own earnings. Also, we mentioned in Section 7.4 that future extensions should be able to account for the burden that it is to have straitened agents in a social group, which would reverse the current side-effect of decreased accumulated points in the use of the MS strategy. Further, egoistic agents could learn by reinforcement that their behaviour is not appropriate and try to reverse it, if there was a known possibility that they could regain trustworthiness in the society. We also plan to introduce some mechanisms to allow agents to discover the "characters" of others and subsequently exercise the freedom to accept or refuse them as partners in future social interactions, as Ridley argues to be the case among humans.

Further, we plan to investigate the robustness of our strategy in an "artificial life" type of simulation. For the particular purpose of this research, the fitness function to be used in the selection mechanism must be relative to groups and not individuals as is usual in this type of simulation at present. Although the concept of group selection is inconsistent with the selfish-gene theory (Dawkins 1989), we could, in principle, consider the idea of cultural group selection (Soltis, Boyd, and Richerson 1995; cited in (Ridley 1996)) to test our approach in an evolutionary setting. For humans, there are conditions under which group selection could occur, even though the evidence elicited so far accounts only for a mild form of cultural group selection, as mentioned in (Ridley 1996). Since the analogies on which we base our simulations regard traits peculiar to humans (in particular emotions), the experiment would be worth trying.

9.3 Final Remarks

We envision that multi-agent systems will be able to interoperate and still have diverging artificial cultures and communication languages, as opposed to current standardising approaches to interoperability of MAS. There is arguably a broad interdisciplinary interest in Multi-Agent Systems of this particular brand. In other words, we believe that, with this proposed type of interoperability, any computational systems can potentially be part of a *global MAS* which is:

- *complex*, in the sense it comprises limitless independent societies or groups (all of which can interoperate, no matter what their discrepancies);
- *heterogeneous*, in the sense that societies have their own "culture," yet this will not be a hindrance to interoperability;
- *open* or *dynamic*, since agents can migrate between societies (thus engendering interoperability);
- *adaptive* both in the sense that societies adapt to whatever number of agents are present and agents adapt to whatever traits they face in the societies to which they migrate;
- apt for *autonomous evolution*, as agents can make use of their intelligence and autonomy to decide on changes over their own "culture," again without causing problems of interoperability.

The approach to interoperability of MAS discussed in this thesis relies on cognitive agents which migrate to MAS of whatever flavour (cognitive, reactive, game-theoretical, market-oriented, etc.). The advantage of this anthropological approach as opposed to the usual approach to interoperability is twofold, in particular for the kind of MAS that is founded on complex aspects of social and cognitive sciences (Conte and Castelfranchi 1995). Providing agents that can learn to interact with different communities not only allows interoperation of disparate MAS without the need for standardisation on models and languages of agents (which is not likely to succeed given the actual history and peculiarities of the MAS area), but also brings to light several issues of interest in the disciplines that form the foundations of MAS. Further, there is evidence from the social sciences (Lave 1988) that the understanding of cognition, which is fundamental in DAI, cannot be accomplished without anthropological considerations. Accordingly, we have endeavoured to show the importance of placing social anthropology among the disciplines of interest to an appropriate conception of open MAS. It is well known that MAS can be helpful in the foundations of the social sciences—including anthropology (Doran *et al.* 1994; Doran and Palmer 1995)—as argued by (Gasser 1987; Castelfranchi 1997). We have aimed at showing that the converse relation also holds good: DAI can profit from work on cultural anthropology in connection to some open research issues concerning, in particular, interoperability of societies of agents.

Appendix A

Communication Features of a Sample Target Society

This appendix is related to the material in Chapter 3.

A.1 Protocol Description Language

We give below the syntax of the Protocol Description Language (PDL) as a BNF grammar. We have used the following conventions: terminal symbols are underlined (this applies to the whole appendix) and non-terminal symbols are enclosed in "<" and ">". Apart from slight improvements, this is the language presented in (Populaire *et al.* 1993). For the basic concepts used in the language (e.g., transitions, states, protocols, laws) see (Populaire *et al.* 1993) or (Bordini 1994).

Protocols

```
<Table_of_Laws> ::=  <TLaw> <Table_of_Laws> |
                     <TLaw>
<TLaw>          ::=  version <Date> <Law> |
                     <Law>
<Law>           ::=  law <Name> { <LProtocols> }
<Date>          ::=  <Day> / <Month> / <Year>
<LProtocols>    ::=  protocol <Name> { <LStates> } <LProtocols> |
                     protocol <Name> { <LStates> }
<LStates>       ::=  <State> <LStates> |
                     <State>
<State>         ::=  state <Name> { <LTransitions> }
```

Transitions

```
<LTransitions> ::=  [ <KTransition> ] | <LTransitions> |
                    [ <KTransition> ] ;
<KTransition>  ::=  if (<condition>) <Transition> |
                    <Transition> |
                    end
<Transition>   ::=  <Message> -> <Stage> |
                    -> <Stage>
<Message>      ::=  (<Type>) (<Route>) (<Force>) (<Content>)
<Type>         ::=  inform |
                    request |
                    answer
<Route>        ::=  you |
```

broadcast |
<Group_of_Agents>
<Stage> ::= law <Name> protocol <Name> state <Name> |
protocol <Name> state <Name> |
state <Name>

(Note that the syntax of the Protocol Execution Language (PEL) was given in Section 3.3.2.)

A.2 Formal Semantics of the Protocol Languages

For the sake of simplicity, we do not consider the syntactic levels above that of a *protocol* from the PDL.

A.2.1 Semantics in the Structural Approach to Operational Semantics

For details of the Plotkin's structural approach to operational semantics (which is based on *terminal labelled transition systems*), see (Plotkin 1981).

Protocol Description Language

Before we give semantics to the PDL, we provide some common basic sets which will be used in several definitions:

- $\mathcal{Q} = \{q_0, \ldots, q_i, \ldots\}, i \in \mathbb{N}$

  – recursively enumerable set of protocol states
- $\Sigma = \{\sigma_0, \ldots, \sigma_i, \ldots\}, i \in \mathbb{N}$

  – recursively enumerable set of message interpretations
- $\Delta = ((\Sigma \cup \{\varepsilon\}) \times \mathcal{Q}) \cup \{\varepsilon\}$

  – set of all alternative protocol transitions from a certain state
- $\Psi = \mathcal{Q} \times \mathbb{P}(\Delta)$

  – set of all available states and all available protocol transitions from each of them (i.e., the "next-state" relation for each state)

One particular feature of the definitions given below which is potentially misleading is the fact that we use a transition system (the basis of the operational approach) to give semantics to a language used to describe systems where transitions (between configurations, in this case called states) also occur. It is important to note that protocols described by the PDL can themselves be represented by transitions systems. Therefore, to avoid confusion, we use the term *protocol transition* to denote the sort of transitions described by the PDL, and *state* to denote the configurations which these transitions transform.

Now we are able to provide the semantics for the PDL. First we give its abstract syntax, which is the adequate format of syntax for the approach to semantics used here. In fact, we give the abstract syntax of the configurations of each terminal transition system used: note the introduction, in the abstract syntaxes, of meta-variables ranging over the structures that are the terminal configurations of the terminal transition system giving the semantics for each syntactic category. This means (as in (Plotkin 1981)) that the meta-variables are part of the abstract syntax of the corresponding *configuration* rather than the syntactic category itself. The formal definitions follow:

Syntax

Basic Sets

States	$q \in Q$
Messages	$\sigma \in \Sigma$
Protocol Transitions	$\delta \in \Delta$
List of Protocol Transitions	$TL \in \mathbb{P}(\Delta)$
Definitions of State	$\psi \in \Psi$
Definitions of List of States	$SL \in \mathbb{P}(\Psi)$

Derived Sets

Protocol Transitions	$t \in \mathit{Trans}$, where $t ::= \sigma \; \underline{\texttt{->}} \; q$ $\mid \; \underline{\texttt{->}} \; q$ $\mid \; \underline{\texttt{end}}$ $\mid \; \delta$
List of Protocol Transitions	$tl \in \mathit{TransList}$, where $tl ::= \underline{[} \; t \; \underline{]} \; \underline{\mid} \; tl$ $\mid \; \underline{[} \; t \; \underline{]} \; \underline{;}$ $\mid \; TL$
Definitions of State	$s \in \mathit{StateDef}$, where $s ::= \underline{\texttt{state}} \; q \; \underline{\{} \; tl \; \underline{\}}$ $\mid \; \psi$
Definitions of List of States	$sl \in \mathit{StateDefList}$, where $sl ::= s \; sl$ $\mid \; s$ $\mid \; SL$

Transition Rules

Protocol Transitions

$\langle \mathit{Trans}, \rightarrowtail, \Delta \rangle$

Normal Transition	$\sigma \; \underline{\texttt{->}} \; q_0 \rightarrowtail (\sigma, q_0)$
Empty Transition	$\underline{\texttt{->}} \; q_0 \rightarrowtail (\varepsilon, q_0)$
Final Transition	$\underline{\texttt{end}} \rightarrowtail \varepsilon$

List of Transitions

$\langle \mathit{TransList}, \rightarrowtail, \mathbb{P}(\Delta) \rangle$

1. $\dfrac{t \rightarrowtail \delta}{\underline{[}\ t\ \underline{]}\ \underline{;} \rightarrowtail \{\delta\}}$

2. $\dfrac{t \rightarrowtail \delta}{\underline{[}\ t\ \underline{]}\ \underline{|}\ tl \rightarrowtail \underline{[}\ \delta\ \underline{]}\ \underline{|}\ tl}$

3. $\dfrac{tl \rightarrowtail tl'}{\underline{[}\ \delta\ \underline{]}\ \underline{|}\ tl \rightarrowtail \underline{[}\ \delta\ \underline{]}\ \underline{|}\ tl'}$

4. $\underline{[}\ \delta\ \underline{]}\ \underline{|}\ TL \rightarrowtail \{\delta\} \cup TL$

Definitions of State

$\langle \mathit{StateDef}, \rightarrowtail, \Psi \rangle$

1. $\dfrac{tl \rightarrowtail tl'}{\underline{state}\ q_0\ \underline{\{}\ tl\ \underline{\}} \rightarrowtail \underline{state}\ q_0\ \underline{\{}\ tl'\ \underline{\}}}$

2. $\underline{state}\ q_0\ \underline{\{}\ TL\ \underline{\}} \rightarrowtail (q_0, TL)$

Definitions of List of States

$\langle \mathit{StateDefList}, \rightarrowtail, \mathbb{P}(\Psi) \rangle$

1. $\dfrac{s \rightarrowtail s'}{s\ sl \rightarrowtail s'\ sl}$

2. $\dfrac{sl \rightarrowtail sl'}{\psi\ sl \rightarrowtail \psi\ sl'}$

3. $\psi \rightarrowtail \{\psi\}$

4. $\psi\ SL \rightarrowtail \{\psi\} \cup SL$

Protocol Execution Language

Below, we give semantics to the *execution* of protocols (i.e, execution of transitions of protocols). To distinguish from protocol description (rather than execution) we use a different symbol for the transition relation (of the labelled transitions system used to give semantics in the structural approach to operational semantics): we use "$\twoheadrightarrow$" instead of "$\rightarrowtail$". Below, there is a label used in the transition rules, which is $\mathbf{P} \in \mathbb{P}(\Psi)$—it expresses the fact that the execution of protocol transitions depends on the interpretation of the protocol according to the semantics given above (i.e., the protocol as "learned" by the agent).

First, we introduce the following auxiliary notation for a "stack of conversations" (i.e., each of the parallel parts of the conversational structure in Figure 3.2):

Notation for the Stack of Conversations

$\mathit{emptyconvstack} : \quad \rightarrow \mathit{CStack}$
$\mathit{convstack} \quad : Q \times \mathit{CStack} \rightarrow \mathit{CStack}$

where Q is the same set of states given above.

We also need a recursively enumerable set for the names to the conversation, denoted by *CName*. Further, the domain of a conversation C is the power set of:

$$\mathit{Conv} = \{(cn, cs) \mid cn \in \mathit{CName}, cs \in \mathit{CStack}\}$$

We now give the semantics itself.

Syntax

Basic Sets

Conversation Names	$cn \in \mathit{CName}$
States	$q \in Q$
Messages	$m \in (\Sigma \cup \{\varepsilon\})$

Derived Sets

Execution of Protocol Transitions $ex \in \mathit{Exec}$, where

$$\begin{array}{rl} ex ::= & ex_0 \ \underline{;}\ ex_1 \\ \mid & cn\ \underline{:}\ \underline{\text{new}}\ m\ \underline{\text{->}}\ \text{q} \\ \mid & cn\ \underline{:}\ \underline{\text{nested}}\ m\ \underline{\text{->}}\ \text{q} \\ \mid & cn\ \underline{:}\ \underline{\text{resume}} \\ \mid & cn\ \underline{:}\ t \\ \mid & \underline{\text{nil}} \end{array}$$

where $t \in \mathit{Trans}$, as before.

(Labelled) Transition System

$\Gamma = \{\langle ex, C\rangle \mid ex \in \mathit{Exec}, C \in \mathbb{P}(\mathit{Conv})\} \cup \{C\}$
$\mathrm{T} = \{C\}$
$\langle \Gamma, \mathbb{P}(\Psi), \twoheadrightarrow, \mathrm{T}\rangle$

Transition Rules

Nil

$\langle \underline{\text{nil}}, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C$

Composition

1. $$\frac{\langle ex_0, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} \langle ex_0', C'\rangle}{\langle ex_0 \ \underline{;}\ ex_1, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} \langle ex_0' \ \underline{;}\ ex_1, C'\rangle}$$

2. $$\frac{\langle ex_0, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}{\langle ex_0 \ \underline{;}\ ex_1, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} \langle ex_1, C'\rangle}$$

Simple:

- Normal

$$\frac{(cn, \mathit{convstack}(q', ps)) \in C \land (q', \delta) \in \mathbf{P} \land (\sigma, q) \in \delta}{\langle cn\ \underline{:}\ \sigma\ \underline{\text{->}}\ q, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = (C \setminus \{(cn, \mathit{convstack}(q', ps))\}) \cup \{(cn, \mathit{convstack}(q, ps))\}$

- Empty

$$\frac{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps})) \in C \wedge (q', \delta) \in \mathbf{P} \wedge (\varepsilon, q) \in \delta}{\langle \mathit{cn}\ \underline{:}\ \varepsilon\ \underline{\texttt{->}}\ q, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = (C \setminus \{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps}))\}) \cup \{(\mathit{cn}, \mathit{convstack}(q, \mathit{ps}))\}$

- Final

$$\frac{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps})) \in C \wedge (q', \delta) \in \mathbf{P} \wedge \varepsilon \in \delta}{\langle \mathit{cn}\ \underline{:}\ \underline{\text{end}}, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = (C \setminus \{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps}))\}) \cup \{(\mathit{cn}, \mathit{ps})\}$

Concurrent

$$\frac{(q', \delta) \in \mathbf{P} \wedge (m, q) \in \delta}{\langle \mathit{cn}\ \underline{:}\ \underline{\text{new}}\ m\ \underline{\texttt{->}}\ q, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = C \cup \{\mathit{convstack}(q, \mathit{emptyconvstack})\}$

Nested

$$\frac{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps})) \in C \wedge (q', \delta) \in \mathbf{P} \wedge (m, q) \in \delta}{\langle \mathit{cn}\ \underline{:}\ \underline{\text{nested}}\ m\ \underline{\texttt{->}}\ q, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = (C \setminus \{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps}))\}) \cup \{(\mathit{cn}, \mathit{convstack}(q, \mathit{convstack}(q', \mathit{ps})))\}$

Resuming

$$\frac{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps})) \in C}{\langle \underline{\text{resume}}, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow} C'}$$

where $C' = (C \setminus \{(\mathit{cn}, \mathit{convstack}(q', \mathit{ps}))\}) \cup \{(\mathit{cn}, \mathit{ps})\}$

Finally, the *behaviour* of the execution of protocol transitions is given by the function *Execute* : $\mathit{Exec} \times \mathbb{P}(\mathit{Conv}) \times \mathbb{P}(\Psi) \to \mathbb{P}(\mathit{Conv})$ defined by:

$$\mathit{Execute}(\mathit{ex}, C, \mathbf{P}) = C' \iff \langle \mathit{ex}, C\rangle \overset{\mathbf{P}}{\twoheadrightarrow}{}^{*} C'$$

Note that, although the protocols are non-deterministic, *Execute* is defined as a function rather than a relation because the protocol transition(s) to be executed (parameter *ex* to *Execute*) determine completely the next conversational state.

A.2.2 Semantics in the Denotational Framework

The notation used here for the denotational semantics is the one given in (Watt 1991) for giving semantics to programming languages (and is based on λ-calculus). The symbols '⟦' and '⟧' are used to emphasise syntactic phrases that are arguments to semantic functions. The only notation that is not standard and was used below is the following. A definition of the type **let** $\delta' = (q, \delta) \in \mathbf{P}$ **in** *e* **end** denotes some sort of instantiation (==) of free variables. For example, if $q == q_0$ and $\mathbf{P} == \{(q_0, \{(m_0, q_1), (\epsilon, q_1)\}), (q_1, \{\epsilon\})\}$, then, in *e*, the following holds: $\delta == \{(m_0, q_1), (\epsilon, q_1)\}$ and $\delta' == (q_0, \{(m_0, q_1), (\epsilon, q_1)\})$.

Protocol Description Language

Semantic Domains

(The same sets Q, Σ, Δ and Ψ defined in Section A.2.1, in this appendix.)

Semantic Functions

Interpretation of Messages

Semantic Function
$\mathcal{MI} : \mathit{Message} \rightarrow \Sigma$

$\mathcal{MI}$ is a bijection from messages to message interpretations.

The function $\mathcal{MI}$ binds every message interpretation in Σ to syntactic messages m.

Interpretation of State Names

Semantic Function
$\mathcal{NI} : \mathit{Name} \rightarrow Q$

$\mathcal{NI}$ is a bijection from state names to states.

The function $\mathcal{NI}$ binds every state in the set of states Q to every state name n (where the state name is the token used to represent that state in the protocol).

Interpretation of Protocol Transitions

Semantic Function
$\mathcal{TI} : \mathit{Transition} \rightarrow \Delta$

Semantic Equations

$\mathcal{TI}\,[\![\, m\ \underline{\texttt{->}}\ \underline{\text{state}}\ n \,]\!] =$
  let $\sigma = \mathcal{MI}\ m$ **in**
  let $q = \mathcal{NI}\ n$ **in**
  (σ, q)
$\mathcal{TI}\,[\![\ \underline{\texttt{->}}\ n \,]\!] =$
  let $q = \mathcal{NI}\ n$ **in**
  (ε, q)
$\mathcal{TI}\,[\![\, \underline{\text{end}} \,]\!] = \varepsilon$

Normal transitions are interpreted as pairs containing a message and a state (obtained by the interpretation of its state name in the syntactic phrase being interpreted). A transition can be: a pair (σ, q), which is the normal one; a pair (ε, q), which is an empty transition; and it can also be ε, which is the final one. Each of these interpreted transitions will be part of the next-state relation associated to the state of whose definition they are part.

Interpretation of Lists of Transitions

Semantic Function
$\mathcal{TLI} : \mathit{LTransitions} \rightarrow \mathbb{P}(\Delta)$

Semantic Equations

$\mathcal{TLI}\,[\![\, \underline{[}t\underline{]} \mid tl \,]\!] =$
  let $\delta = \mathcal{TI}\ t$ **in**
  let $TL = \mathcal{TLI}\ tl$ **in**
  $TL \cup \{\ \delta\ \}$
$\mathcal{TLI}\,[\![\, \underline{[}t\underline{]}\ \underline{;} \,]\!] =$
  let $\delta = \mathcal{TI}\ t$ **in**
  $\{\ \delta\ \}$

The meaning of lists of transitions is, clearly, a member of $\mathbb{P}(\Delta)$. A state is defined by its list of transitions, i.e., a set whose members are of the types given above; it has the information about the next state where to move on the receipt or transmission of a message.

Interpretation of State Definitions

Semantic Function
$\mathcal{SI} : State \rightarrow \Psi$

Semantic Equations
$\mathcal{SI}\, [\![\ \underline{\text{state}}\ n\ \underline{\{}\ tl\ \underline{\}}\ ]\!] =$
  let $q = \mathcal{NI}\ n$ **in**
  let $TL = \mathcal{TLI}\ tl$ **in**
  (q, TL)

When a syntactic phrase for a complete state definition is interpreted, it yields a pair formed of the state in Q which represents unequivocally that state and a set giving all alternative transitions from *that state.*

Interpretation of Lists of State Definitions

Semantic Function
$\mathcal{SLI} : LStates \rightarrow \mathbb{P}(\Psi)$

Semantic Equations
$\mathcal{SLI}\, [\![\ s\ sl\ ]\!] =$
  let $\psi = \mathcal{SI}\ s$ **in**
  let $SL = \mathcal{SLI}\ sl$ **in**
  $SL \cup \{\ \psi\ \}$
$\mathcal{SLI}\, [\![\ s\ ]\!] =$
  let $\psi = \mathcal{SI}\ s$ **in**
  $\{\ \psi\ \}$

Finally, we give semantics to a list of state definitions. Note that this is exactly a protocol *(a set of which form each* law *being defined). The interpretation of a protocol is in* $\mathbb{P}(\Psi)$ *(recall that* Ψ *is the domain of all state definitions).*

Although we have omitted part of the PDL in the formal semantics above, the extension is trivial: protocols are given names and a list of them forms a law that is also given a name; a set of such laws would then be the semantics of a table of laws, which completes the PDL as given above in Section A.1.

Protocol Execution Language

Before we introduce the main semantic function here which gives semantics to the whole execution of protocols, we will need one more auxiliary semantic function.

Interpretation of Conversation Names

Semantic Function
$\mathcal{CI} : Name \rightarrow CName$

$\mathcal{CI}$ is a bijection from names to conversation identifiers.

The function $\mathcal{CI}$ *binds every conversation identifier in the set CName to its given name n (where the state name is the token used to represent that state in the protocol).*

We now provide the semantic function for the execution of protocols and all the semantic equations giving its definition. Below, *ispfm* denotes the <u>i</u>nitial <u>s</u>tate of the <u>p</u>rotocol <u>f</u>ailure <u>m</u>eta-protocol, which is part of the global protocol (presented in the next section).

Semantic Function

$Execute : Exec \times \mathbb{P}(Conv) \times (\Psi) \rightarrow \mathbb{P}(Conv)$

Semantic Equations

$$Execute\,[\![\, \underline{\mathrm{nil}} \,]\!]\; C\, \mathbf{P} = C$$

The execution of a transition nil *does not affect the state of an agent's conversations.*

$$\begin{array}{l} Execute\,[\![\, ex_0 \;\underline{;}\; ex_1 \,]\!]\; C\, \mathbf{P} = \\ \qquad Execute\,[\![\, ex_1 \,]\!]\; (Execute\,[\![\, ex_0 \,]\!]\; C\, \mathbf{P})\; \mathbf{P} \end{array}$$

The execution of the composition *of protocol transition means the execution of the second transition in the context of the conversation generated by the first when executed in the context of the given conversation structure. The protocol to be used in the executions is the same.*

$$\begin{array}{l} Execute\,[\![\, n \,\underline{:}\, m \;\underline{\texttt{->}}\; \underline{\mathrm{state}}\; s \,]\!]\; C\, \mathbf{P} = \\ \qquad \mathbf{let}\; cn = \mathcal{CI}\; n\; \mathbf{in} \\ \qquad \mathbf{let}\; C' = (cn, convstack(q, ps)) \in C\; \mathbf{in} \\ \qquad \mathbf{let}\; \delta' = (q, \delta) \in \mathbf{P}\; \mathbf{in} \\ \qquad \mathbf{let}\; \sigma = \mathcal{MI}\; m\; \mathbf{in} \\ \qquad \mathbf{let}\; q' = \mathcal{NI}\; s\; \mathbf{in} \\ \qquad \mathbf{if}\, (\sigma, q') \in \delta \\ \qquad\quad \mathbf{then}\; (C \setminus \{C'\}) \cup \{(cn, convstack(q', ps))\} \\ \qquad\quad \mathbf{else}\;\; (C \setminus \{C'\}) \cup \{(cn, convstack(ispfm, convstack(q, ps)))\} \end{array}$$

The semantics of the execution of a simple protocol transition of type normal *is simply the change of the current state of that particular conversation. Instead, if there is a protocol failure, ispfm is pushed onto the top of the previous stack for that conversation, i.e., the next state given in that transition is not considered (this is true for the equations below too).*

$$\begin{array}{l} Execute\,[\![\, n \,\underline{:}\; \underline{\texttt{->}}\; \underline{\mathrm{state}}\; s \,]\!]\; C\, \mathbf{P} = \\ \qquad \mathbf{let}\; cn = \mathcal{CI}\; n\; \mathbf{in} \\ \qquad \mathbf{let}\; C' = (cn, convstack(q, ps)) \in C\; \mathbf{in} \\ \qquad \mathbf{let}\; \delta' = (q, \delta) \in \mathbf{P}\; \mathbf{in} \\ \qquad \mathbf{let}\; q' = \mathcal{NI}\; s\; \mathbf{in} \\ \qquad \mathbf{if}\, (\varepsilon, q') \in \delta \\ \qquad\quad \mathbf{then}\; (C \setminus \{C'\}) \cup \{(cn, convstack(q', ps))\} \\ \qquad\quad \mathbf{else}\;\; (C \setminus \{C'\}) \cup \{(cn, convstack(ispfm, convstack(q, ps)))\} \end{array}$$

The semantics of the execution of a simple protocol transition of type empty *is also simply the change of the current state of the corresponding conversation; however, in this case, the change of state occurs on the presence of a special message* ε *(i.e., one with no content).*

$$\begin{array}{l} Execute\,[\![\, n \,\underline{:}\; \underline{\mathrm{end}} \,]\!]\; C\, \mathbf{P} = \\ \qquad \mathbf{let}\; cn = \mathcal{CI}\; n\; \mathbf{in} \\ \qquad \mathbf{let}\; C' = (cn, convstack(q, ps)) \in C\; \mathbf{in} \\ \qquad \mathbf{let}\; \delta' = (q, \delta) \in \mathbf{P}\; \mathbf{in} \\ \qquad \mathbf{if}\, \varepsilon \in \delta \\ \qquad\quad \mathbf{then}\; (C \setminus \{C'\}) \cup \{(cn, ps)\} \\ \qquad\quad \mathbf{else}\;\; (C \setminus \{C'\}) \cup \{(cn, convstack(ispfm, convstack(q, ps)))\} \end{array}$$

To execute a simple protocol transition of type final *means to finish that particular conversation (resuming the interrupted conversation immediately preceding that one, if there is any).*

```
Execute $[\![\, n \underline{:}\ \underline{\mathrm{new}}\ m\ \underline{\texttt{->}}\ \underline{\mathrm{state}}\ s \,]\!]\ C\ \mathbf{P} =$
    let $cn = \mathcal{CI}\ n$ in
    let $\sigma = \mathcal{MI}\ m$ in
    let $q' = \mathcal{NI}\ s$ in
    if $\exists\, q.(q, \delta) \in \mathbf{P} \wedge (\sigma, q') \in \delta$
      then $C \cup C'$
      else $C \cup C''$
```

where $C' = \{(cn, convstack(q', emptyconvstack))\}$
and $C'' = \{(cn, convstack(ispfm, emptyconvstack))\}$

This is the case where the conversation control directive new *creates a new parallel conversation (the first being created in parallel with no other). A new stack containing only the current state of that new conversation is added to the set of conversations. Note that our semantics does not considered levels of detail such as checking whether q, besides existing, is also an initial state of that protocol.*

```
Execute $[\![\, n \underline{:}\ \underline{\mathrm{nested}}\ m\ \underline{\texttt{->}}\ \underline{\mathrm{state}}\ s \,]\!]\ C\ \mathbf{P} =$
    let $cn = \mathcal{CI}\ n$ in
    let $C' = (cn, convstack(q, ps)) \in C$ in
    let $\delta' = (q, \delta) \in \mathbf{P}$ in
    let $\sigma = \mathcal{MI}\ m$ in
    let $q' = \mathcal{NI}\ s$ in
    if $(\sigma, q') \in \delta$
      then $(C \setminus \{C'\}) \cup \{(cn, convstack(q', C'))\}$
      else $(C \setminus \{C'\}) \cup \{(cn, convstack(ispfm, C'))\}$
```

The conversation to which the directive nested *refers is found in the set of conversation and the new current state is pushed onto that stack.*

```
Execute $[\![\, n \underline{:}\ \underline{\mathrm{resume}} \,]\!]\ C\ \mathbf{P} =$
    let $cn = \mathcal{CI}\ n$ in
    let $C' = (cn, convstack(q, ps)) \in C$ in
    $(C \setminus \{C'\}) \cup \{(cn, ps)\}$
```

Finally, it is possible to cancel a nested conversation started previously by means of resume. *This can be accomplished by simply popping its current state off the relevant stack. Note that, although resuming is meant for* nested *conversations, we do not actually check whether that is the case for the given conversation.*

A.3 Meta-Protocols

The meta-protocols included in the communication features of a hypothetical society of agents, which we considered in Chapter 3, are given in Figure A.1. For an informal description of this global protocol, see Section 3.3.3.

```
law Global
{ protocol Protocol_Failure
  { state Initial
    { [ (inform)(you)(informing)(unexpected(Previous_Transition)) -> state Inform_Version ];
    }
    state Inform_Version
    { [ (inform)(you)(informing)(version(Number)) -> state Check_Version ];
    }
    state Check_Version
    { [ if(newer) (request)(you)(information_seeking)(transition(Previous_Transition)) ->
        state Inform_Transition ] |
      [ if(older) (inform)(you)(informing)(invalid_transition(Previous_Transition)) ->
        state Request_State ];
    }
    state Inform_Transition
    { [ (reply)(you)(informing)(transition(Previous_Transition)) -> state Final ];
    }
    state Request_State
    { [ (request)(you)(information_seeking)(state(Previous_Transition)) ->
        state Inform_State ];
    }
    state Inform_State
    { [ (reply)(you)(informing)(state(Previous_Transition_State)) -> state Final ];
    }
    state Final
    { [ end ];
    }
  }

  protocol Protocol_Exchange
  { state Initial
    { [ (request)(you)(information_seeking)(protocol(P) ∧ purpose(P, R)) ->
        state Inform_if_Possible ];
    }
    state Inform_if_Possible
    { [ if(known(P))    (reply)(you)(informing)(protocol(P) ∧ purpose(P, R)) -> state Final ] |
      [ if(¬ known(P)) (reply)(you)(informing)(unknown(P)) -> state Final ];
    }
    state Final
    { [ end ];
    }
  }
}
```

Figure A.1: Sample Global Protocol

Appendix B

Notation and Animation of the Z Specifications

This appendix is related to the material in Chapter 5.

B.1 Elements of the Z Notation

This appendix is intended to help the reader who is unfamiliar with the Z notation, although it is probably not helpful to the reader who is completely unfamiliar with formal specification. The Z notation is based on set theory and first order logic. Further, as Spivey (1992) puts it: "The other main ingredient in Z is a way of decomposing a specification into small pieces called *schemas*." He further remarks that: "A Z specification document consists of interleaved passages of formal, mathematical text and informal prose explanation. The formal text consists of a sequence of paragraphs which gradually introduce the schemas, global variables and basic types of the specification, each paragraph building on the ones which come before it." The notion of *schema* is very clearly explained by d'Inverno and Luck (1998): "Z schemas have two parts: the upper declarative part, which declares variables and their types, and the lower predicate part, which relates and constrains those variables. The type of any schema can be considered as the Cartesian product of the types of each of its variables, without any notion of order, but constrained by the schema's predicates. Modularity is facilitated in Z by allowing schemas to be included within other schemas." It also provides for the specification of systems at different levels of abstraction (i.e., details of the system can be added by means of a process of *refinement*). A last comment that is worth making here is again quoted from (Spivey 1992):

> In Z, schemas are used to describe both static and dynamic aspects of a system. The static aspects include:
>
> - the states it can occupy;
> - the invariant relationships that are maintained as the systems moves from state to state.
>
> The dynamic aspects include:
>
> - the operations that are possible;
> - the relationships between their inputs and outputs;
> - the changes of state that happen.

We next introduce in a concise way the subset of the Z notation that we have used in this paper. It is given in Table B.1 which is partly borrowed from d'Inverno and Luck (1998). It has been improved and restructured (partly based on (Jia 1995) and (Spivey 1992)), and also adapted to the particular subset of the Z notation that has been used in this paper.

Conventions Used Below

a	Identifier
d	Declaration
e, x, y	Expressions
p, q	Predicates
A, B	Sets
P	Ordered Pair
R, R_1, R_2	Relations
S, T	Schemas
s	Schema Type Expression

Declarations

$[a]$	Given Set
$a == x$	Equivalence Definition
$A ::= b\langle\langle B \rangle\rangle$	Free Type Definition

Axiomatic Descriptions

$\begin{array}{\|l} d \\ \hline p \end{array}$	(introduce global variables)

Schema Definitions

$\begin{array}{\|l} \hline S \\ d \\ \hline p \\ \hline \end{array}$	Vertical Schema
$\begin{array}{\|l} \hline S \\ T \\ \hline p \\ \hline \end{array}$	Schema Inclusion
$\begin{array}{\|l} \hline S \\ \Delta T \\ \hline p \\ \hline \end{array}$	Operations

Schema Conventions

ΔS	Change of State ($S \wedge S'$)
ΞS	No Change of State (i.e., $\Delta S \mid \theta S' = \theta S$, where θ is Binding Formation)

Decorations

$a?$	Input to an Operation
d'	Component After Operation
S'	Schema After Operation

Sequences

$\mathrm{iseq}\, A$	Injective Sequences

Ordered Pairs

$x \mapsto y$	Maplet
first P	Projection Function returning the First Coordinate
second P	Projection Function returning the Second Coordinate

Sets

$\emptyset$	Empty Set
$\{x, y, \ldots\}$	Set Display
$\{d \mid p \bullet e\}$	Set Comprehension
$A \times B \times \ldots$	Cartesian Product
$\mathbb{P} A$	Power Set
$\mathbb{P}_1 A$	Non-empty Power Set
$A \subseteq B$	Subset Relation
$A \cup B$	Set Union
$A \setminus B$	Set Difference
$\bigcup A$	Generalised Union
$\bigcap A$	Generalised Intersection
$\# A$	Size of a Finite Set

Relations

$A \leftrightarrow B$	Binary Relations
$\mathrm{dom}\, R$	Domain of a Relation
$\mathrm{ran}\, R$	Range of a Relation
$R^{\sim}$	Inverse Relation
$R_1 \oplus R_2$	Overriding

Functions

$A \rightarrow B$	Total Functions
$A ⇸ B$	Partial Functions
$A ⤔ B$	Partial Injections
$A ⤖ B$	Bijections

Expressions

$(x, y, \ldots)$	Ordered Tuple
$\mu\, d \mid p \bullet e$	Definite Descriptions (unique value for d satisfying p)
$s.a$	Component Selection

Predicates

$x \in A$	Set Membership
$x \notin A$	Non-Membership
$x = y$	Equality
$x \neq y$	Inequality
$p \wedge q$	Conjunction
$p \Rightarrow q$	Implication
$p \Leftrightarrow q$	Equivalence
$\forall\, d \mid p \bullet q$	Universal Quantification
$\exists\, d \mid p \bullet q$	Existential Quantification
$\mathbf{let}\ a == e \bullet p$	Local Definition

In declarations, "_" gives the positions for the parameters. Therefore, the notation "$R\_$" is used to declare a prefixed relation, which allows it to be used in the usual form for predicates.

Table B.1: The Relevant Subset of the Z Notation (adapted from d'Inverno and Luck 1998)

B.2 Animated Version of the Specification of Ascription of Intensional Ontologies and Retrieval of Taxonomical Relations

This is the specifications that we provide to the Z animator PIZA (Hewitt 1997).

The Basic Setting

$$\begin{array}{rl} \mathit{Term} ::= & \mathit{player} \mid \mathit{bowler} \mid \mathit{fastblr} \mid \mathit{slowblr} \mid \mathit{batsman} \mid \mathit{wicketkpr} \\ \mid & \mathit{field} \mid \mathit{side} \mid \mathit{on} \mid \mathit{leg} \mid \mathit{off} \mid \mathit{pitch} \end{array}$$

$$\begin{array}{rl} \mathit{Pred} ::= & \mathit{HB} \mid \mathit{TB} \mid \mathit{TBF} \mid \mathit{IS} \mid \mathit{TBS} \mid \mathit{IA} \mid \mathit{CB} \mid \mathit{WG} \\ \mid & \mathit{WP} \mid \mathit{FP} \mid \mathit{LRB} \mid \mathit{RRB} \mid \mathit{CF} \end{array}$$

$$\mathit{TimeInstant} ::= \mathit{ti}1 \mid \mathit{ti}2$$

$$\mathit{Sent} == (\mathit{Pred} \times \mathit{Term})$$

$$\mathit{TSocId} ::= \mathit{CSoc}$$

$$\mathit{InfAgId} ::= \mathit{IAg}1 \mid \mathit{IAg}2$$

$$\mathit{AcceptanceRelation} == \mathbb{P}(\mathit{Sent} \times \mathit{TimeInstant})$$

$$\begin{array}{|l} \mathit{accepts} : \mathit{InfAgId} \nrightarrow \mathit{AcceptanceRelation} \\ \hline \mathit{accepts} = \{ \\ \quad \mathit{IAg}1 \mapsto \{((\mathit{HB}, \mathit{player}), \mathit{ti}1), ((\mathit{HB}, \mathit{bowler}), \mathit{ti}1), ((\mathit{TB}, \mathit{bowler}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{fastblr}), \mathit{ti}1), ((\mathit{TB}, \mathit{fastblr}), \mathit{ti}1), ((\mathit{TBF}, \mathit{fastblr}), \mathit{ti}1), \\ \qquad ((\mathit{IS}, \mathit{fastblr}), \mathit{ti}1), ((\mathit{HB}, \mathit{slowblr}), \mathit{ti}1), ((\mathit{TB}, \mathit{slowblr}), \mathit{ti}1), \\ \qquad ((\mathit{TBS}, \mathit{slowblr}), \mathit{ti}1), ((\mathit{IA}, \mathit{slowblr}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{batsman}), \mathit{ti}1), ((\mathit{CB}, \mathit{batsman}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{wicketkpr}), \mathit{ti}1), ((\mathit{WG}, \mathit{wicketkpr}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{field}), \mathit{ti}1), ((\mathit{WP}, \mathit{side}), \mathit{ti}1), ((\mathit{FP}, \mathit{side}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{on}), \mathit{ti}1), ((\mathit{FP}, \mathit{on}), \mathit{ti}1), ((\mathit{LRB}, \mathit{on}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{leg}), \mathit{ti}1), ((\mathit{FP}, \mathit{leg}), \mathit{ti}1), ((\mathit{LRB}, \mathit{leg}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{off}), \mathit{ti}1), ((\mathit{FP}, \mathit{off}), \mathit{ti}1), ((\mathit{RRB}, \mathit{off}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{pitch}), \mathit{ti}1), ((\mathit{CF}, \mathit{pitch}), \mathit{ti}1)\}, \\ \quad \mathit{IAg}2 \mapsto \{((\mathit{HB}, \mathit{player}), \mathit{ti}1), ((\mathit{HB}, \mathit{bowler}), \mathit{ti}1), ((\mathit{TB}, \mathit{bowler}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{fastblr}), \mathit{ti}1), ((\mathit{TB}, \mathit{fastblr}), \mathit{ti}1), ((\mathit{TBF}, \mathit{fastblr}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{slowblr}), \mathit{ti}1), ((\mathit{TB}, \mathit{slowblr}), \mathit{ti}1), ((\mathit{TBS}, \mathit{slowblr}), \mathit{ti}1), \\ \qquad ((\mathit{IA}, \mathit{slowblr}), \mathit{ti}1), ((\mathit{HB}, \mathit{batsman}), \mathit{ti}1), ((\mathit{CB}, \mathit{batsman}), \mathit{ti}1), \\ \qquad ((\mathit{HB}, \mathit{wicketkpr}), \mathit{ti}1), ((\mathit{WG}, \mathit{wicketkpr}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{field}), \mathit{ti}1), ((\mathit{WP}, \mathit{side}), \mathit{ti}1), ((\mathit{FP}, \mathit{side}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{on}), \mathit{ti}1), ((\mathit{FP}, \mathit{on}), \mathit{ti}1), ((\mathit{LRB}, \mathit{on}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{leg}), \mathit{ti}1), ((\mathit{FP}, \mathit{leg}), \mathit{ti}1), ((\mathit{LRB}, \mathit{leg}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{off}), \mathit{ti}1), ((\mathit{FP}, \mathit{off}), \mathit{ti}1), ((\mathit{RRB}, \mathit{off}), \mathit{ti}1), \\ \qquad ((\mathit{WP}, \mathit{pitch}), \mathit{ti}1), ((\mathit{CF}, \mathit{pitch}), \mathit{ti}1)\}\} \end{array}$$

$$\begin{array}{|l}
the\_time : \operatorname{iseq} TimeInstant \\
\hline
the\_time = \langle ti1, ti2 \rangle
\end{array}$$

$$\begin{array}{|l}
\_\ before\_eq\ \_ : TimeInstant \leftrightarrow TimeInstant \\
\hline
\forall t\_1, t\_2 : TimeInstant \bullet \\
\qquad t\_1\ before\_eq\ t\_2 \Leftrightarrow the\_time^{\sim}(t\_1) \leq the\_time^{\sim}(t\_2)
\end{array}$$

$$\begin{array}{|l}
most\_recent : \mathbb{P}_1\ TimeInstant \rightarrow TimeInstant \\
\hline
\forall tis : \mathbb{P}_1\ TimeInstant \bullet \\
\qquad most\_recent(tis) = (\mu\ ti : TimeInstant \mid \\
\qquad\qquad ti \in tis \land \\
\qquad\qquad (\forall t : TimeInstant \mid t \in tis \bullet t\ before\_eq\ ti) \bullet \\
\qquad ti)
\end{array}$$

Ascription of Intensional Ontologies

$$IntensionalOntologyOfTerms == Term \nrightarrow \mathbb{P}_1\ Pred$$

$$\begin{array}{l}
SubjectiveQuasiIntension == \\
\qquad (Term \times TimeInstant \times AcceptanceRelation) \rightarrow \mathbb{P}\ Pred
\end{array}$$

$$\begin{array}{l}
IntersubjectiveQuasiIntension == \\
\qquad (Term \times TimeInstant) \rightarrow \mathbb{P}\ Pred
\end{array}$$

$$\begin{array}{|l}
subjective\_quasi\_intension : SubjectiveQuasiIntension \\
\hline
\forall t : Term;\ ti : TimeInstant;\ ar : AcceptanceRelation \bullet \\
\qquad subjective\_quasi\_intension(t, ti, ar) = \\
\qquad\qquad \{ p : Pred \mid ((p, t), ti) \in ar \}
\end{array}$$

$$\begin{array}{|l}
intersubjective\_quasi\_intension : IntersubjectiveQuasiIntension \\
\hline
\forall t : Term;\ ti : TimeInstant \bullet \\
\qquad intersubjective\_quasi\_intension(t, ti) = \\
\qquad\qquad \bigcap \{ ia : InfAgId \bullet \\
\qquad\qquad\qquad subjective\_quasi\_intension(t, ti, accepts(ia)) \}
\end{array}$$

$$\begin{array}{|l}
co\_intensive : IntensionalOntologyOfTerms \rightarrow \mathbb{P}(Term \times Term) \\
\hline
\forall iot : IntensionalOntologyOfTerms \bullet \\
\qquad co\_intensive(iot) = \{ t\_1, t\_2 : Term \mid t\_1 \neq t\_2 \land \\
\qquad\qquad t\_1 \in \operatorname{dom} iot \land t\_2 \in \operatorname{dom} iot \land iot(t\_1) = iot(t\_2) \}
\end{array}$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{AnthropologistAgent}} \\
\hline
\mathit{AutonomousAgent} \\
\mathit{generate\_sentences} : (\mathit{TSocId} \times \mathit{TimeInstant}) \rightarrow \mathbb{P}\,\mathit{Sent} \\
\mathit{history\_of\_intensional\_ontologies} : \\
\qquad (\mathit{TSocId} \times \mathit{TimeInstant}) \mathrel{+\!\!\!\!\rightarrow} \\
\qquad\qquad \mathit{IntensionalOntologyOfTerms} \\
\mathit{known\_societies} : \mathbb{P}\,\mathit{TSocId} \\
\mathit{all\_versions} : \mathit{TSocId} \mathrel{+\!\!\!\!\rightarrow} \mathbb{P}\,\mathit{TimeInstant} \\
\mathit{current\_ontology} : \mathit{TSocId} \mathrel{+\!\!\!\!\rightarrow} \mathit{IntensionalOntologyOfTerms} \\
\mathit{current\_synonyms} : \mathit{TSocId} \mathrel{+\!\!\!\!\rightarrow} \mathit{Term} \leftrightarrow \mathit{Term} \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{InitialAnthropologistAgent}} \\
\hline
\mathit{AnthropologistAgent}' \\
\hline
\mathit{history\_of\_intensional\_ontologies}' = \varnothing \\
\mathit{generate\_sentences}' = \{(\mathit{CSoc}, \mathit{ti}1) \mapsto \mathit{Sent}\} \\
\mathit{known\_societies}' = \varnothing \\
\mathit{all\_versions}' = \varnothing \\
\mathit{current\_ontology}' = \varnothing \\
\mathit{current\_synonyms}' = \varnothing \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{AscribeIntensionalOntologyOfTerms}} \\
\hline
\Delta\mathit{AnthropologistAgent} \\
\mathit{ts}? : \mathit{TSocId} \\
\mathit{ti}? : \mathit{TimeInstant} \\
\hline
\mathit{history\_of\_intensional\_ontologies}' = \\
\qquad \mathit{history\_of\_intensional\_ontologies} \oplus \{(\mathit{ts}?, \mathit{ti}?) \mapsto \\
\qquad\qquad \{\, t : \mathit{Term} \mid \\
\qquad\qquad \mathit{intersubjective\_quasi\_intension}(t, \mathit{ti}?) \neq \varnothing \bullet \\
\qquad\qquad t \mapsto \mathit{intersubjective\_quasi\_intension}(t, \mathit{ti}?) \,\}\} \\
\mathit{known\_societies}' = \{\, s : (\mathit{TSocId} \times \mathit{TimeInstant}) \mid \\
\qquad s \in \operatorname{dom} \mathit{history\_of\_intensional\_ontologies}' \bullet \mathit{first}\ s \,\} \\
\mathit{all\_versions}' = \{\, \mathit{ts} : \mathit{TSocId} \mid \mathit{ts} \in \mathit{known\_societies}' \bullet \mathit{ts} \mapsto \\
\qquad \{\, \mathit{ti} : \mathit{TimeInstant} \mid (\mathit{ts}, \mathit{ti}) \in \\
\qquad\qquad \operatorname{dom} \mathit{history\_of\_intensional\_ontologies}' \,\}\,\} \\
\mathit{current\_ontology}' = \{\, \mathit{ts} : \mathit{TSocId};\ \mathit{ti} : \mathit{TimeInstant} \mid \\
\qquad \mathit{ts} \in \mathit{known\_societies}' \wedge \mathit{ti} = \mathit{most\_recent}(\mathit{all\_versions}'(\mathit{ts})) \bullet \\
\qquad\qquad \mathit{ts} \mapsto \mathit{history\_of\_intensional\_ontologies}'(\mathit{ts}, \mathit{ti}) \,\} \\
\mathit{current\_synonyms}' = \{\, \mathit{ts} : \mathit{TSocId} \mid \mathit{ts} \in \mathit{known\_societies}' \bullet \\
\qquad \mathit{ts} \mapsto \{\, \mathit{t\_1}, \mathit{t\_2} : \mathit{Term} \mid (\mathit{t\_1}, \mathit{t\_2}) \in \\
\qquad\qquad \mathit{co\_intensive}(\mathit{current\_ontology}'(\mathit{ts})) \,\}\,\} \\
\mathit{generate\_sentences}' = \mathit{generate\_sentences} \\
\hline
\end{array}
$$

$$
\begin{array}{l}
\mathit{IntertemporalQuasiIntension} == \\
\qquad (\mathit{Term} \times \mathit{AcceptanceRelation}) \rightarrow \mathbb{P}\,\mathit{Pred}
\end{array}
$$

$$
\begin{aligned}
&\mathit{ObjectiveQuasiIntension} == \\
&\quad \mathit{Term} \to \mathbb{P}\, \mathit{Pred}
\end{aligned}
$$

$$
\begin{array}{|l}
\mathit{intertemporal\_quasi\_intension} : \mathit{IntertemporalQuasiIntension} \\
\hline
\forall t : \mathit{Term};\ ar : \mathit{AcceptanceRelation} \bullet \\
\quad \mathit{intertemporal\_quasi\_intension}(t, ar) = \\
\quad\quad \{\, p : \mathit{Pred} \mid \forall\, ti : \mathit{TimeInstant} \bullet \\
\quad\quad\quad ((p, t), ti) \in ar \,\}
\end{array}
$$

$$
\begin{array}{|l}
\mathit{objective\_quasi\_intension} : \mathit{ObjectiveQuasiIntension} \\
\hline
\forall t : \mathit{Term} \bullet \\
\quad \mathit{objective\_quasi\_intension}(t) = \\
\quad\quad \bigcap \{\, ia : \mathit{InfAgId} \bullet \\
\quad\quad\quad \mathit{intertemporal\_quasi\_intension}(t, \mathit{accepts}(ia)) \,\}
\end{array}
$$

$$
\begin{array}{|l}
\hline
\mathit{ExperiencedAnthropologistAgent} \\
\hline
\mathit{immutable\_intensional\_ontology} : \mathit{TSocId} \mathrel{⇸} \\
\quad \mathit{IntensionalOntologyOfTerms} \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\hline
\mathit{InitialExperiencedAnthropologistAgent} \\
\hline
\mathit{ExperiencedAnthropologistAgent}' \\
\Xi \mathit{AnthropologistAgent} \\
\hline
\mathit{immutable\_intensional\_ontology}' = \\
\quad \{\, ts : \mathit{TSocId} \mid ts \in \mathit{known\_societies} \bullet ts \mapsto \{\, t : \mathit{Term} \mid \\
\quad\quad \mathit{objective\_quasi\_intension}(t) \neq \varnothing \bullet \\
\quad\quad\quad t \mapsto \mathit{objective\_quasi\_intension}(t) \,\}\,\} \\
\hline
\end{array}
$$

Retrieval of Taxonomical Relations

$$
\begin{aligned}
\mathit{Taxonomy} ::= {} & \mathit{segregate}\langle\!\langle \mathbb{P}_1\, \mathit{Term} \times \mathbb{P}_1\, \mathit{Pred} \rangle\!\rangle \\
\mid {} & \mathit{contrastsets}\langle\!\langle \mathit{Taxonomy} \times \mathbb{P}_1\, \mathit{Taxonomy} \rangle\!\rangle
\end{aligned}
$$

$$
\begin{array}{|l}
\mathit{find\_segregate} : \mathit{IntensionalOntologyOfTerms} \to \\
\quad (\mathit{Taxonomy} \times \mathit{IntensionalOntologyOfTerms}) \\
\hline
\forall\, iot : \mathit{IntensionalOntologyOfTerms} \bullet \mathit{find\_segregate}(iot) = \\
\quad (\mathbf{let}\, P == \{\, p : \mathit{Pred} \mid (\forall t : \mathit{Term} \mid \\
\quad t \in \mathrm{dom}\, iot \bullet p \in iot(t)) \,\} \bullet \\
\quad\quad (\mathbf{let}\, T == \{\, t : \mathit{Term} \mid t \in \mathrm{dom}\, iot \wedge iot(t) = P \,\} \bullet \\
\quad\quad\quad (\mathbf{let}\, siot == \{\, t : \mathit{Term} \mid t \in \mathrm{dom}\, iot \wedge t \notin T \bullet \\
\quad\quad\quad t \mapsto iot(t) \setminus P \,\} \bullet (\mathit{segregate}(T, P), siot))))
\end{array}
$$

$$
\begin{array}{|l}
\mathit{segregate\_size} : (\mathit{IntensionalOntologyOfTerms} \times \mathit{Pred}) \to \mathbb{N} \\
\hline
\forall\, iot : \mathit{IntensionalOntologyOfTerms};\ p : \mathit{Pred} \bullet \\
\quad \mathit{segregate\_size}(iot, p) = \#\{\, t : \mathit{Term} \mid t \in \mathrm{dom}\, iot \wedge p \in iot(t) \,\}
\end{array}
$$

$$
\begin{array}{|l}
find\_contrast\_sets : IntensionalOntologyOfTerms \rightarrow \\
\qquad \mathbb{P}_1\ IntensionalOntologyOfTerms \\
\hline
\forall\, iot : IntensionalOntologyOfTerms \bullet \\
\qquad (iot = \varnothing \Rightarrow find\_contrast\_sets(iot) = \varnothing) \land \\
\qquad (iot \neq \varnothing \Rightarrow find\_contrast\_sets(iot) = \\
\qquad\qquad (\mathbf{let}\, ps == (\mu\, p : Pred \mid p \in \bigcup(\mathrm{ran}\, iot) \land \\
\qquad\qquad\qquad (\forall\, q : Pred \mid q \in \bigcup(\mathrm{ran}\, iot) \bullet \\
\qquad\qquad\qquad segregate\_size(iot, p) \geq segregate\_size(iot, q)) \bullet p) \bullet \\
\qquad\qquad\qquad\quad (\mathbf{let}\, csiot == \{\ t : Term \mid t \in \mathrm{dom}\, iot \land ps \in iot(t) \bullet \\
\qquad\qquad\qquad\quad t \mapsto iot(t)\ \} \bullet \{csiot\} \cup find\_contrast\_sets(iot \setminus csiot))))
\end{array}
$$

$$
\begin{array}{|l}
find\_taxonomical\_relations : IntensionalOntologyOfTerms \rightarrow Taxonomy \\
\hline
\forall\, iot : IntensionalOntologyOfTerms \bullet \\
\qquad (second(find\_segregate(iot)) = \varnothing \Rightarrow find\_taxonomical\_relations(iot) = \\
\qquad\qquad first(find\_segregate(iot))) \land \\
\qquad (second(find\_segregate(iot)) \neq \varnothing \Rightarrow find\_taxonomical\_relations(iot) = \\
\qquad\qquad contrastsets(first(find\_segregate(iot)), \\
\qquad\qquad\qquad \{\ csiot : IntensionalOntologyOfTerms \mid csiot \in \\
\qquad\qquad\qquad\quad find\_contrast\_sets(second(find\_segregate(iot))) \bullet \\
\qquad\qquad\qquad\qquad find\_taxonomical\_relations(csiot)\ \}))
\end{array}
$$

$$
\begin{array}{|l}
generate\_taxonomical\_relations : \\
\qquad (IntensionalOntologyOfTerms \times Pred) \rightarrow Taxonomy \\
\hline
\forall\, iot : IntensionalOntologyOfTerms;\ p : Pred \bullet \\
\qquad generate\_taxonomical\_relations(iot, p) = (\mathbf{let}\, piot == \\
\qquad\qquad \{\ t : Term \mid t \in \mathrm{dom}\, iot \land p \in iot(t) \bullet t \mapsto iot(t)\ \} \bullet \\
\qquad\qquad\quad find\_taxonomical\_relations(piot))
\end{array}
$$

$$
\begin{array}{|l|}
\hline
TaxonomistAnthropologistAgent \\
\hline
generate\_subjects : TSocId \rightarrow \mathbb{P}\ Pred \\
taxonomical\_relations : TSocId \rightarrow \mathbb{P}\ Taxonomy \\
\hline
\end{array}
$$

$$
\begin{array}{|l|}
\hline
InitialTaxonomistAnthropologistAgent \\
\hline
TaxonomistAnthropologistAgent' \\
\Xi AnthropologistAgent \\
\hline
generate\_subjects' = \{CSoc \mapsto \{HB, WP\}\} \\
taxonomical\_relations' = \{\ ts : TSocId \bullet ts \mapsto \{\ p : Pred \mid \\
\qquad p \in generate\_subjects'(ts) \bullet \\
\qquad\qquad generate\_taxonomical\_relations(current\_ontology(ts), p)\ \}\ \} \\
\hline
\end{array}
$$

Migration of Agents

$$
\begin{array}{|l}
\hline
\multicolumn{1}{l}{\mathit{MigrantAgent}} \\
\hline
\mathit{AutonomousAgent} \\
\mathit{ts} : \mathit{TSocId} \\
\mathit{intensional\_ontology} : \mathit{IntensionalOntologyOfTerms} \\
\mathit{immutable\_ontology} : \mathit{IntensionalOntologyOfTerms} \\
\mathit{synonyms} : \mathit{Term} \leftrightarrow \mathit{Term} \\
\mathit{taxonomies} : \mathbb{P}\, \mathit{Taxonomy} \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\bot : \mathit{TSocId}
\end{array}
$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{InitialMigrantAgent}} \\
\hline
\mathit{MigrantAgent}' \\
\hline
\mathit{ts}' = \bot \\
\mathit{intensional\_ontology}' = \varnothing \\
\mathit{immutable\_ontology}' = \varnothing \\
\mathit{synonyms}' = \varnothing \\
\mathit{taxonomies}' = \varnothing \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{Migrate}} \\
\hline
\Delta \mathit{MigrantAgent} \\
\Xi \mathit{AnthropologistAgent} \\
\Xi \mathit{ExperiencedAnthropologistAgent} \\
\Xi \mathit{TaxonomistAnthropologistAgent} \\
\mathit{ts}? : \mathit{TSocId} \\
\hline
\mathit{ts}' = \mathit{ts}? \\
\mathit{intensional\_ontology}' = \mathit{current\_ontology}(\mathit{ts}?) \\
\mathit{immutable\_ontology}' = \mathit{immutable\_intensional\_ontology}(\mathit{ts}?) \\
\mathit{synonyms}' = \mathit{current\_synonyms}(\mathit{ts}?) \\
\mathit{taxonomies}' = \mathit{taxonomical\_relations}(\mathit{ts}?) \\
\hline
\end{array}
$$

B.3 Results of the Animation for the Cricket Case Study

Given the specifications above, we provide PiZA with the following sequence of transactions (*cs* is the PiZA directive for "schema calls").

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{cs}1} \\
\hline
\mathit{InitialAnthropologistAgent} \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{cs}2} \\
\hline
\mathit{AscribeIntensionalOntologyOfTerms} \\
\hline
\mathit{ts}? = \mathit{CSoc} \\
\mathit{ti}? = \mathit{ti}1 \\
\hline
\end{array}
$$

cs3

InitialExperiencedAnthropologistAgent

cs4

InitialTaxonomistAnthropologistAgent

cs5

InitialMigrantAgent

cs6

Migrate

$ts? = CSoc$

Finally, the results that PIZA produces for those transactions are as given below. (Note that whenever the set *Sent* was displayed exhaustively, we changed it to simply *Sent* for the sake of readability.)

```
%piza-logfile
  cs1 =======================================================================
Results of calling schema InitialAnthropologistAgent
all_versions' = {}
current_ontology' = {}
current_synonyms' = {}
generate_sentences' = { ( ( CSoc, ti1 ), Sent ) }
history_of_intensional_ontologies' = {}
known_societies' = {}
==========================================================================

  cs2 =======================================================================
Results of calling schema AscribeIntensionalOntologyOfTerms
all_versions = {}
all_versions' = { ( CSoc, { ti1 } ) }
current_ontology = {}
current_ontology' = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_synonyms = {}
current_synonyms' = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
generate_sentences = { ( ( CSoc, ti1 ), Sent ) }
generate_sentences' = { ( ( CSoc, ti1 ), Sent ) }
history_of_intensional_ontologies = {}
history_of_intensional_ontologies' = { ( ( CSoc, ti1 ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
known_societies = {}
known_societies' = { CSoc }
ti? = ti1
ts? = CSoc
==========================================================================
```

```
  cs3 ======================================================================
Results of calling schema InitialExperiencedAnthropologistAgent
all_versions = { ( CSoc, { til } ) }
all_versions' = { ( CSoc, { til } ) }
current_ontology = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_ontology' = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_synonyms = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
current_synonyms' = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
generate_sentences = { ( ( CSoc, til ), Sent ) }
generate_sentences' = { ( ( CSoc, til ), Sent ) }
history_of_intensional_ontologies = { ( ( CSoc, til ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
history_of_intensional_ontologies' = { ( ( CSoc, til ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
immutable_intensional_ontology' = { ( CSoc, {} ) }
known_societies = { CSoc }
known_societies' = { CSoc }
==========================================================================

  cs4 ======================================================================
Results of calling schema InitialTaxonomistAnthropologistAgent
all_versions = { ( CSoc, { til } ) }
all_versions' = { ( CSoc, { til } ) }
current_ontology = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_ontology' = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_synonyms = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
current_synonyms' = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
generate_sentences = { ( ( CSoc, til ), Sent ) }
generate_sentences' = { ( ( CSoc, til ), Sent ) }
generate_subjects' = { ( CSoc, { HB, WP } ) }
history_of_intensional_ontologies = { ( ( CSoc, til ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
history_of_intensional_ontologies' = { ( ( CSoc, til ), {
```

```
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
known_societies = { CSoc }
known_societies' = { CSoc }
taxonomical_relations' = { ( CSoc, {
   contrastsets( segregate( { field }, { WP } ), {
      contrastsets( segregate( { side }, { FP } ), {
         segregate( { leg, on }, { LRB } ), segregate( { off }, { RRB } ) } ),
      segregate( { pitch }, { CF } ) } ),
   contrastsets( segregate( { player }, { HB } ), {
      contrastsets( segregate( { bowler }, { TB } ), {
         segregate( { fastblr }, { TBF } ),
         segregate( { slowblr }, { IA, TBS } ) } ),
      segregate( { batsman }, { CB } ),
      segregate( { wicketkpr }, { WG } ) } ) } ) }
=========================================================================

   cs5 ==================================================================
Results of calling schema InitialMigrantAgent
immutable_ontology' = {}
intensional_ontology' = {}
synonyms' = {}
taxonomies' = {}
ts' = ⊥
=========================================================================

   cs6 ==================================================================
Results of calling schema Migrate
all_versions = { ( CSoc, { ti1 } ) }
all_versions' = { ( CSoc, { ti1 } ) }
current_ontology = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_ontology' = { ( CSoc, { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
current_synonyms = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
current_synonyms' = { ( CSoc, { ( leg, on ), ( on, leg ) } ) }
generate_sentences = { ( ( CSoc, ti1 ), Sent ) }
generate_sentences' = { ( ( CSoc, ti1 ), Sent ) }
generate_subjects = { ( CSoc, { HB, WP } ) }
generate_subjects' = { ( CSoc, { HB, WP } ) }
history_of_intensional_ontologies = { ( ( CSoc, ti1 ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
history_of_intensional_ontologies' = { ( ( CSoc, ti1 ), {
   ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) } ) }
immutable_intensional_ontology = { ( CSoc, {} ) }
```

```
immutable_intensional_ontology' = { ( CSoc, {} ) }
immutable_ontology = {}
immutable_ontology' = {}
intensional_ontology = {}
intensional_ontology' = { ( batsman, { CB, HB } ), ( bowler, { HB, TB } ),
   ( fastblr, { HB, TB, TBF } ), ( field, { WP } ), ( leg, { FP, LRB, WP } ),
   ( off, { FP, RRB, WP } ), ( on, { FP, LRB, WP } ), ( pitch, { CF, WP } ),
   ( player, { HB } ), ( side, { FP, WP } ), ( slowblr, { HB, IA, TB, TBS } ),
   ( wicketkpr, { HB, WG } ) }
known_societies = { CSoc }
known_societies' = { CSoc }
synonyms = {}
synonyms' = { ( leg, on ), ( on, leg ) }
taxonomical_relations = { ( CSoc, {
   contrastsets( segregate( { field }, { WP } ), {
      contrastsets( segregate( { side }, { FP } ), {
         segregate( { leg, on }, { LRB } ), segregate( { off }, { RRB } ) } ),
      segregate( { pitch }, { CF } ) } ),
   contrastsets( segregate( { player }, { HB } ), {
      contrastsets( segregate( { bowler }, { TB } ), {
         segregate( { fastblr }, { TBF } ),
         segregate( { slowblr }, { IA, TBS } ) } ),
      segregate( { batsman }, { CB } ),
      segregate( { wicketkpr }, { WG } ) } ) } ) }
taxonomical_relations' = { ( CSoc, {
   contrastsets( segregate( { field }, { WP } ), {
      contrastsets( segregate( { side }, { FP } ), {
         segregate( { leg, on }, { LRB } ), segregate( { off }, { RRB } ) } ),
      segregate( { pitch }, { CF } ) } ),
   contrastsets( segregate( { player }, { HB } ), {
      contrastsets( segregate( { bowler }, { TB } ), {
         segregate( { fastblr }, { TBF } ),
         segregate( { slowblr }, { IA, TBS } ) } ),
      segregate( { batsman }, { CB } ),
      segregate( { wicketkpr }, { WG } ) } ) } ) }
taxonomies = {}
taxonomies' = {
   contrastsets( segregate( { field }, { WP } ), {
      contrastsets( segregate( { side }, { FP } ), {
         segregate( { leg, on }, { LRB } ), segregate( { off }, { RRB } ) } ),
      segregate( { pitch }, { CF } ) } ),
   contrastsets( segregate( { player }, { HB } ), {
      contrastsets( segregate( { bowler }, { TB } ), {
         segregate( { fastblr }, { TBF } ),
         segregate( { slowblr }, { IA, TBS } ) } ),
      segregate( { batsman }, { CB } ),
      segregate( { wicketkpr }, { WG } ) } ) }
ts = ⊥
ts? = CSoc
ts' = CSoc
=============================================================================
```

Appendix C

Algorithms for the Reasoning of the Cricket Players

This appendix is related to the material in Chapter 6. It is the high-level code for the general algorithms providing the behaviours for the "players" in the cricket simulation. (Note that "`this`" in C++ is a reference to the object that is executing the functions itself.)

```
/* -----------------------------------------------------------------------
   File: Reason.cc
   Contents: Functions for Players' High Level Reasoning
   Author: Rafael Heitor Bordini
   -------------------------------------------------------------------- */

/* -----------------------------------------------------------------------
                              DIRECTIVES
   -------------------------------------------------------------------- */

// this is needed with multi-threaded programs
#ifndef _REENTRANT
#define _REENTRANT
#endif

#include "Cricket.hh"
#include "ReasAF.hh"

/* -----------------------------------------------------------------------
                           Member Functions
   -------------------------------------------------------------------- */

void Bowler::reasoning()
{
   if ( NewBowling() ) {
      Bowl();
      SetBallThrown();
   }
   else { // Behave as a mixture of fieldsman and wicketkeeper
      if ( BallHit() || BallMissed() || BallToWicketkeeper() ) {
         if ( CaughtBall( this ) ) {
            if ( Caught( this ) ) {
               // Umpiring of OutCaught is here
               OutCaught( this );
            }
            else {
               StopRunning( this );
               if ( WorthWicket( this ) && !BatsmanSafe() ) {
```

```
                  ThrowBallToWicket( this );
                  SetBallToWicket();
               }
            }
         }
         else {
            if ( WorthBall( this ) && !BatsmanSafe() &&
                 NobodyInCharge() ) {
               RunToBall( this );
            }
            else if ( Running( this ) ) {
               StopRunning( this );
            }
         }
      }
      else if ( Running( this ) ) {
         StopRunning(this);
      }
   }
}

void Batsman::reasoning()
{
   if ( BallThrown() ) {
      if ( CollisionBallBat() ) {
         if ( UseLBW() ) {
            // Umpiring of LBW is transferred to here
            OutLBW();
         }
         else {
            HitBall();
            SetBallHit();
         }
      }
   }
   else {
      if ( !BatsmanRunning() ) {
         if ( WorthARun() ) {
            StartRunnings();
         }
      }
      else {
         if ( BatsmanIn() ) {
            if ( WorthARun() ) {
               RunToOppositePC();
            }
            else {
               StopRunnings();
            }
         }
      }
   }
}

void Fieldsman::reasoning()
{
   // FA failed to run when everybody expected him to
```

```
if ( BallHit() && IsNotMine( this ) && NobodyInCharge() ) {
   DecideIfBackUp( this );
   if ( IsBackUp( this ) ) {
      Complain( this );
   }
}
if ( (BallHit() || BallMissed()) && !IsNotMine( this ) ) {
   if ( IsNotKnown( this ) && BallHit() ) {
      DecideIfBallIsMine( this );
   }
   else if ( IsNotKnown(this) && BallMissed() ) {
      DecideIfBackUp( this );
   }
   // persistent goal: run to ball until get it or ball is dead
   if ( IsMine( this ) || IsBackUp( this ) ) {
      if ( CaughtBall( this ) ) {
         if ( Caught( this ) ) {
            // Umpiring of OutCaught is here
            OutCaught( this );
         }
         else {
            StopRunning( this );
            if ( !BatsmanSafe() ) {
               if ( WorthWicket(this) && !TooFar( this ) ) {
                  ThrowBallToWicket(this);
                  SetBallToWicket();
               }
               else {
                  ThrowBallToWicketkeeper(this);
                  SetBallToWicketkeeper();
               }
            }
            else {
               ThrowBallToNearer(this);
               SetBallToWicketkeeper();
            }
         }
      }
      else {
         if ( IsMine( this ) ) {
            if ( Running( this ) ) {
               if ( RunningToTarget( this ) ){
                  if ( MissedBall( this ) ) {
                     if (BallTooHigh(this)){
                        StopRunToTarget( this );
                        RunToBall( this );
                     }
                     else {
                        StopRunning( this );
                        SetBallMissed();
                        TellOthersMissedBall();
                     }
                  }
                  if ( ReachedTarget( this ) ) {
                     StopRunToTarget( this );
                  }
               }
```

```
                  else {
                     RunToBall( this );
                  }
               }
               else {
                  RunToBallTrajectory( this );
               }
            }
            else { // Backing Up
               if ( SomebodyElseIsCloser( this ) ) {
                  StopRunning( this );
                  SetBallMissed();
                  TellOthersMissedBall();
               }
               else {
                  RunToBall( this );
               }
            }
         }
      }

   }
   else if ( BallToWicketkeeper() || BallToWicket() ) {
      // first to realise that fielding was not enough tells others
      if ( !EnoughFielding() ) {
         // fielding was not enough for ball to reach wicket
         // or wicketkeeper, so keep fielding as backing up
         SetBallMissed();
         TellOthersMissedBall();
      }
   }
}

void WicketKeeper::reasoning()
{
   if ( BallHit() || BallMissed() || BallToWicketkeeper() ) {
      if ( CaughtBall( this ) ) {
         if ( Caught( this ) ) {
            // Umpiring of OutCaught is transferred to here
            OutCaught( this );
         }
         else {
            StopRunning( this );
            if ( !BatsmanSafe() ) {
               if ( !BatsmanJustLeft() && !WKNearWicket() ) {
                  RunToBreakWicket();
               }
               else {
                  ThrowBallToWicket( this );
               }
               SetBallToWicket();
            }
         }
      }
      else {
         if ( IsNotKnown( this ) && BallHit() ) {
            DecideIfBallIsMine( this );
```

```
         }
         if ( IsMine( this ) ) {
            if ( Running( this ) ) {
               if ( RunningToTarget( this ) ) {
                  if (MissedBall( this ) && !SlowBall() && !StillNearMyArea()){
                     StopRunning( this );
                     SetBallMissed();
                     TellOthersMissedBall();
                  }
                  if ( ReachedTarget( this ) ) {
                     StopRunToTarget( this );
                  }
               }
               else {
                  if ( StillNearMyArea() || SlowBall() ) {
                     RunToBall( this );
                  }
                  else if ( !StillNearMyArea() ){
                     StopRunning( this );
                     SetBallMissed();
                     TellOthersMissedBall();
                  }
               }
            }
            else {
               if ( SlowBall() ) {
                  RunToBall( this );
               }
               else {
                  RunToBallTrajectory( this );
               }
            }
         }
         else if ( WorthBall( this ) && !(BatsmanSafe() &&
              BallToWicketkeeper()) && NobodyRunning() ) {
            RunToBall( this );
         }
         else if ( !BallToWicketkeeper() &&
              RunningToTarget( this ) && ReachedTarget( this ) ) {
            StopRunning( this );
            StopRunToTarget( this );
         }
         else if ( !InStandingPosition() && !Running( this ) &&
              !BallToWicketkeeper() ) {
            RunTowardsWicket();
         }
         else if ( Running( this ) && BallToWicketkeeper() ) {
            StopRunning( this );
         }
      }
   }
   // attempting to send out by running to the wicket
   else if ( RunningToTarget( this ) && ReachedTarget(this) ) {
      StopRunning( this );
      StopRunToTarget( this );
   }
}
```

```
void ForeignAgent::reasoning()
{
   UpdateLearning();
   if ( BallHit() ) {
      if ( IsNotKnown( this ) ) {
         if ( DecideToChase( this ) ) {
            SetMine( this );
            RunToBall( this );
         }
         else {
            SetNotMine( this );
         }
      }
      else if ( IsMine( this ) ) {
         if ( CaughtBall( this ) ) {
            if ( Caught( this ) ) {
               // Umpiring of OutCaught here too
               // although unlikely, but just in case
               OutCaught( this );
            }
            else {
               StopRunning( this );
               DecideWhereToThrow( this );
            }
         }
         else if ( Running( this ) && SomebodyElsesBall() ) {
            StopRunning( this );
            SetNotMine( this );
            Blush();
         }
         else {
            RunToBall( this );
         }
      }
   }
}

/* ----------------------------------------------------------------------
                              Umpiring
   ---------------------------------------------------------------------- */

void Umpiring()
{
// Umpiring of "Out Caught" is done in Bowler, Fieldsman and Wicketkeeper's
//   Reasoning, in this file (also for Foreign Agent, just in case)
// Umpiring of "Out LBW" is done in the Batsman's Reasoning, in this file
   if ( HitWicket() ) {
      // Wicket is Down
      if ( BallToWicket() && !BatsmanSafe() ) {
         if ( OnlyWKTouchedBall() ) {
            OutStumped();
         }
         else{
            RunOut();
         }
      }
```

```
      else if ( BallHit() ) {
         BowledOut();
      }
      NextBowling();
   }
   else if ( BallNotPlayed() ) {
      CountNotPlayed();
      NextBowling();
   }
   else if ( BallOut() ) {
      // Ball reached boundaries
      FourRuns();
      NextBowling(); // Ball is dead
   }
   else if ( !BallThrown() && !BallNotStarted() ) {
      if ( BatsmanIn() && BatsmanRunning() ) {
         RunCompleted();
      }
      else if ( ( BallWithWicketkeeper() ||
               BallWithBowler() ) && BatsmanSafe() ) {
         // Ball is dead (they have the ball, but not caught)
         NextBowling();
      }
   }
}
```

Bibliography

Some of the entries in this bibliography have a URL provided at the end. This is given for the papers we are aware that one can download from the WWW. We therefore provide the URL to facilitate access to the material. Note that when URLs are enclosed in "<" and ">", those are electronic publications, therefore the URL is part of the reference rather than an additional information.

Allen, J. F., and Perrault, C. R. 1980. Analyzing intention in utterances. *Artificial Intelligence* 15(3):143–178. Amsterdam, North Holland Publishing Company.

Allen, J. F. 1987. Speech acts. In Shapiro, E., et al., eds., *Encyclopedia of A.I.*, volume II. New York: John Wiley & Sons. 1062–1065.

Aubé, M., and Senteni, A. 1995. A foundation for commitments as resource management in multi-agent systems. In *Workshop on Intelligent Information Agents, held as part of the Fourth International Conference on Information and Knowledge Management (CIKM'95).* Baltimore, MD: ACM, 28 November to 2 December, 1995. URL: `http://www.cs.umbc.edu/~cikm/iia/proc.html`.

Aubé, M., and Senteni, A. 1996. Emotions as commitements operators: a foundation for control structure in multi-agents systems. In Van de Velde, W., and Perram, J., eds., *Proceedings of the Seventh Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96), 22–25 January, Eindhoven, The Netherlands*, number 1038 in Lecture Notes in Artificial Intelligence, 13–25. London: Springer-Verlag.

Austin, J. L. 1962. *How to Do Things with Words.* London: Oxford University Press.

Avouris, N. M., and Gasser, L., eds. 1992. *Distributed Artificial Intelligence: Theory and Praxis.* Eurocourses, Computer and information science, V.5. Dordrecht: Kluwer Academic.

Axelrod, R. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* 37(7):122–125.

Bauman, R., and Sherzer, J., eds. 1989. *Explorations in the Ethnography of Speaking.* Cambridge: Cambridge University Press, second edition.

Bazzan, A. L. C., Bordini, R. H., and Campbell, J. A. 1997. Agents with moral sentiments in an iterated prisoner's dilemma exercise. In Dautenhahn, K., Masthoff, J., and Numaoka, C., eds., *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents, 8–10 November, Cambridge, MA*, 4–6. Menlo Park, CA: AAAI Press. AAAI Press Technical Report FS-97-02. UCL-CS [RN/97/74]. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bazzan, A. L. C., Bordini, R. H., and Campbell, J. A. 1999. Moral sentiments in multi-agent systems. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 113–131. Heidelberg: Springer-Verlag. UCL-CS [RN/98/29]. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bernard, H. R. 1994. *Research Methods in Anthropology: Qualitative and Quantitative Approaches.* Thousand Oaks, California: Sage Publications, second edition.

Berthet, S., Demazeau, Y., and Olivier, B. 1992. Knowing each other better. In *11th International Workshop on Distributed Artificial Intelligence.* Glen Arbor, February, 1992.

Binmore, K. 1998. Review of R. Axelrod's "The complexity of cooperation: Agent-based models of competition and collaboration". *Journal of Artificial Societies and Social Simulation* 1(1). <http://www.soc.surrey.ac.uk/JASSS/1/1/review1.html>.

Bond, A. H., and Gasser, L., eds. 1988. *Readings in Distributed Artificial Intelligence.* San Mateo, CA: Morgan Kaufmann.

Bond, A. H. 1990. A computational model for organizations of cooperating intelligent agents. In *Proceedings of the Conference on Office Information Systems (COIS90).* Cmbridge, MA, 25–27 April, 1990.

Bordini, R. H., and Campbell, J. A. 1995. Towards an anthropological approach to agent adaptation. In *Proceedings of the First International Workshop on Decentralized Intelligent and Multi-Agent Systems (DIMAS'95)*, II/74–II/83. Krakow, Poland: Dom Wydawnictwa Naukowych, 22–24 November, 1995. UCL-CS [RN/95/78]. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bordini, R. H., Bazzan, A. L. C., and Campbell, J. A. 1998. Moral sentiments in the iterated prisoner's dilemma and in multi-agent systems. Manuscript unpublished to date.

Bordini, R. H., da Rocha Costa, A. C., Hübner, J. F., and Viccari, R. M. 1995. Linguistic support for agent migration. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 441. Menlo Park, CA: AAAI Press / MIT Press. Extended Abstract. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bordini, R. H., Campbell, J. A., and Vieira, R. 1996. Ascribing intensional ontologies in anthropologically-based cultural descriptions of multi-agent systems. In Durfee, E., ed., *Proceedings of the Second International Conference on Multi-Agent Systems (ICMAS'96), 11–13 December, Kyoto, Japan*, 427. Menlo Park, CA: AAAI Press. Extended Abstract. UCL-CS [RN/96/69]. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bordini, R. H., Campbell, J. A., and Vieira, R. 1997. Ascription of intensional ontologies in anthropological descriptions of multi-agent systems. In Kandzia, P., and Klusch, M., eds., *Proceedings of the First International Workshop on Cooperative Information Agents (CIA'97), 26–28 February, Kiel, Germany*, volume 1202 of *Lecture Notes in Artificial Intelligence*, 235–247. Berlin: Springer-Verlag. UCL-CS [RN/97/1]. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Bordini, R. H., Campbell, J. A., and Vieira, R. 1998. Extending ascribed intensional ontologies with taxonomical relations in anthropological descriptions of multi-agent systems. *Journal of Artificial Societies and Social Simulation* 1(4). <http://www.soc.surrey.ac.uk/JASSS/1/4/3.html> (PDF version also available there). UCL-CS [RN/98/118]. PS version at URL: **`http://www.cs.ucl.ac.uk/staff/R.Bordini`**.

Bordini, R. H. 1994. Suporte lingüístico para migração de agentes. Master's thesis, Curso de Pós-Graduação em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Bratman, M. E., Israel, D. J., and Pollack, M. E. 1987. Toward an architecture for resource-bounded agents. Report CSLI-87-104, Center for the Study of Language and Information, Stanford, CA.

Bratman, M. E., Israel, D. J., and Pollack, M. E. 1988. Plans and resource-bounded practical reasonin. *Computational Intelligence* 4:349–355.

Brembs, B. 1996. Chaos, cheating and cooperation: Potential solutions to the prisoner's dilemma. *Oikos* 76(1):14–24.

Brodribb, G. 1952. *Next Man In: A Survey of Cricket Laws and Customs.* London: Putnam & Co. Ltd.

Campbell, J. A., and d'Inverno, M. 1990. Knowledge interchange protocols. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A.I.—Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'89), 16–18 August, Cambridge, 1989.* Amsterdam: Elsevier Science B.V. 63–80.

Campbell, J. A., and Wolstencroft, J. 1990. Structure and significance of analogical reasoning. *Artificial Intelligence in Medicine* 2:103–111.

Carley, K. M. 1998. Adaptive organizations and emergent forms. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris,* 2–3. Washington: IEEE Computer Society Press. Extended Abstract for an Invited Talk.

Castelfranchi, C., and Falcone, R. 1998. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris,* 72–79. Washington: IEEE Computer Society Press.

Castelfranchi, C., Conte, R., and Paolucci, M. 1998. Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation* 1(3). <http://www.soc.surrey.ac.uk/JASSS/1/3/3.html>.

Castelfranchi, C., de Rosis, F., and Falcone, R. 1997. Social attitudes and personalities in agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents.* Cambridge, Massachusetts, 8–10 November, 1997.

Castelfranchi, C., Miceli, M., and Cesta, A. 1992. Dependence relations among autonomous agents. In Werner, E., and Demazeau, Y., eds., *Decentralized A.I. 3—Proceedings of the Third European Workshop on Modelling Autonomous Agents in a Multi-Agent World, 5–7 August, 1991, Kaiserslautern, Germany.* Amsterdam: Elsevier Science B.V. 215–227.

Castelfranchi, C. 1990. Social power: a point missed in multi-agent DAI and HCI. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A.I.—Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'89), 16–18 August, Cambridge, 1989.* Amsterdam: Elsevier Science B.V. 49–62.

Castelfranchi, C. 1995. Commitments: From individual intentions to groups and organizations. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 41–48. Menlo Park, CA: AAAI Press / MIT Press.

Castelfranchi, C. 1997. The theory of social functions — challenges for agent-based social simulation. In *Workshop on Simulating Societies, held as part of the International Conference on Computer Simulation and the Social Sciences (ICCS&SS).* Cortona, Italy. Draft paper from an Invited Talk.

Castelfranchi, C. 1998. Through the minds of the agents. *Journal of Artificial Societies and Social Simulation* 1(1). <http://www.soc.surrey.ac.uk/JASSS/1/1/5.html>.

Cavedon, L., and Sonenberg, L. 1998. On social commitment, roles and preferred goals. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 80–87. Washington: IEEE Computer Society Press.

Cesta, A., Miceli, M., and Rizzo, P. 1996. Help under risky conditions: Robustness of the social attitude and system performance. In Durfee, E., ed., *Proceedings of the Second International Conference on Multi-Agent Systems (ICMAS'96), 11–13 December, Kyoto, Japan*, 18–25. Menlo Park, CA: AAAI Press.

Chess, D., Grosof, B., Harrison, C., Levine, D., Parris, C., and Tsudik, G. 1995. Itinerant agents for mobile computing. *IEEE Personal Communications Magazine.* URL: `http://www.cs.umbc.edu/agents/papers.html`.

Cohen, P. R., and Levesque, H. J. 1990a. Intention is choice with commitment. *Artificial Intelligence* 42(3):213–261.

Cohen, P. R., and Levesque, H. J. 1990b. Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., eds., *Intentions in Communication.* Cambridge, MA: MIT Press. chapter 12, 221–255.

Cohen, P. R., and Levesque, H. J. 1995. Communicative actions for artificial agents. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 65–72. Menlo Park, CA: AAAI Press / MIT Press.

Cohen, P. R., and Perrault, C. R. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science* 3(3):177–212.

Collet, C., Huhns, M. N., and Shen, W.-M. 1991. Resource integration using a large knowledge base in carnot. *IEEE Computer* 24(12):55–62.

Conklin, H. C. 1955. Hanunóo color categories. *Southwestern Journal of Anthropology* 11:339–344.

Conte, R., and Castelfranchi, C. 1994. Mind is not enough: the precognitive bases of social interaction. In Gilbert, N., and Doran, J., eds., *Simulating Society: The Computer Simulation of Social Phenomena.* London: UCL Press. chapter 12, 267–286.

Conte, R., and Castelfranchi, C. 1995. *Cognitive and Social Action.* London: UCL Press.

Conte, R., Castelfranchi, C., and Dignum, F. 1999. Autonomous norm acceptance. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 99–112. Heidelberg: Springer-Verlag.

Corrêa, M., and Coelho, H. 1995. Around the architectural agent approach to model conversations. In *Proceedings of the European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'93)*, number 957 in Lecture Notes in Artificial Intelligence, 172–185. Berlin: Springer-Verlag.

Crane, G. R., ed. nd. *The Perseus Project.* URL: `http://www.perseus.tufts.edu`.

da Rocha Costa, A. C., and Coelho, H. nd. On machine intelligence and the efficacy of computing agents in functional conversations. Instituto de Informática, Universidade Federal do Rio Grande do Sul, Brazil. Unpublished manuscript.

da Rocha Costa, A. C., de Castilho, J. M. V., and Claudio, D. M. 1995. Toward a constructive notion of functionality. *Cybernetics and Systems* 26(4):443–480.

da Rocha Costa, A. C., Hübner, J. F., and Bordini, R. H. 1994. On entering an open society. In *XI Brazilian Symposium on Artificial Intelligence*, 535–546. Fortaleza: Brazilian Computing Society, October, 1994. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Damasio, A. R. 1995. *Descartes's Error: Emotion, Reason and the Human Brain.* London: Picador.

Davis, R., and Smith, R. G. 1983. Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence* 20(1):63–109.

Dawkins, R. 1989. *The Selfish Gene.* Oxford: Oxford University Press, new edition.

Decker, K. S., and Lesser, V. R. 1995. Designing a family of coordination algorithms. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 73–80. Menlo Park, CA: AAAI Press / MIT Press.

Decker, K., Pannu, A., Sycara, K., and Williamson, M. 1997. Designing behaviors for information agents. In *To appear in the Proceedings of the First International Conference on Autonomous Agents (AGENTS-97).* February, 1997. URL: `http://www.cs.cmu.edu/~softagents/publications.html`.

Demazeau, Y., and Müller, J.-P. 1990. Decentralized artificial intelligence. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A.I.—Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'89), 16–18 August, Cambridge, 1989.* Amsterdam: Elsevier Science B.V. 3–13.

Demazeau, Y. 1995. From cognitive interactions to collective behaviour in agent-based systems. In *Proceedings of the European Conference on Cognitive Science.* Saint-Malo, April, 1995.

Demazeau, Y., ed. 1998. *Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris.* Washington: IEEE Computer Society Press.

Dennett, D. C. 1987. *The Intentional Stance.* Cambridge, MA: The MIT Press.

d'Inverno, M., and Luck, M. 1996a. A formal view of social dependence networks. In Zhang, C., and Lukose, D., eds., *Distributed Artificial Intelligence: Architecture and Modelling—Proceedgins of the First Australian Workshop on DAI, in conjunction with the Eighth Australian Joint Conference on Artificial Intelligence (AI'95), November 1995, Canberra, Australia*, number 1087 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 115–129.

d'Inverno, M., and Luck, M. 1996b. Understanding autonomous interaction. In Wahlster, W., ed., *Proceedings of the 12^{th} European Conference on Artificial Intelligence (ECAI'96).*

d'Inverno, M., and Luck, M. 1998. Engineering AgentSpeak(L): A formal computational model. *Journal of Logic and Computation* 8(3):1–27.

d'Inverno, M., Kinny, D., Luck, M., and Wooldridge, M. 1998. A formal specification of dMARS. In Singh, M. P., Rao, A. S., and Wooldridge, M., eds., *Intelligent Agents IV—Proceedings of the Fourth International Workshop on Agent Theories, Architectures, and Languages (ATAL-97), Providence, RI, 24–26 July, 1997*, number 1365 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 155–176.

d'Inverno, M., Kinny, D., and Luck, M. 1998. Interaction protocols in Agentis. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 112–119. Washington: IEEE Computer Society Press.

d'Inverno, M. 1998. *Agents, Agency and Autonomy: A Formal Computational Model.* Ph.D. Dissertation, University of London.

Doran, J., and Gilbert, N. 1994. Simulating societies: An introduction. In Gilbert, N., and Doran, J., eds., *Simulating Society: The Computer Simulation of Social Phenomena.* London: UCL Press. chapter 1, 1–18.

Doran, J., and Palmer, M. 1995. The EOS project: Integrating two models of palaeolithic social change. In Gilbert, N., and Conte, R., eds., *Artificial Societies: The Computer Simulation of Social Life.* London: UCL Press. 103–125.

Doran, J., Palmer, M., Gilbert, N., and Mellars, P. 1994. The EOS project: Modelling Upper Palaeolithic social change. In Gilbert, N., and Doran, J., eds., *Simulating Society: The Computer Simulation of Social Phenomena.* London: UCL Press. chapter 9, 195–221.

Doran, J. 1997. Artificial societies and emerging hierarchies. Unpublished manuscript. URL: `http://cswww.essex.ac.uk/staff/doran.htm`

Doran, J. 1998a. Simulating collective misbelief. *Journal of Artificial Societies and Social Simulation* 1(1). <http://www.soc.surrey.ac.uk/JASSS/1/1/3.html>.

Doran, J. 1998b. Social simulation, agents and artificial societies. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 4–5. Washington: IEEE Computer Society Press. Extended Abstract for an Invited Talk.

Durfee, E. H., and Lesser, V. 1991. Partial global planning: A coordination framework for distributed hypothesis formation. *IEEE Transactions on Systems, Man, and Cybernetics* 21(5):1167–1183.

Durfee, E. H., Sen, S., Xiang, Y., and Huhns, M. 1997. Multi-agent learning and teaching. In Hewitt, C., ed., *International Workshop on Multi-Agent Systems (IWMAS).* Cambridge, MA: Massachusetts Institute of Technology, 12-14 October, 1997. Unpublished to date.

Durfee, E. H. 1995. Blissful ignorance: Knowing just enough to coordinate well. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 406–413. Menlo Park, CA: AAAI Press / MIT Press. Invited Talk.

Durfee, E., ed. 1996. *Proceedings of the Second International Conference on Multi-Agent Systems (ICMAS'96), 11–13 December, Kyoto, Japan*. Menlo Park, CA: AAAI Press.

Ephrati, E., and Rosenschein, J. S. 1994. Divide and conquer in multi-agent planning. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI94), 31 July to 4 August, Seattle, WA*, volume 1, 375–380. Menlo Park, CA: AAAI Press.

Fagin, R., and Halpern, J. Y. 1988. Belief, awareness, and limited reasoning. *Artificial Intelligence* 34(1):39–76.

Fenster, M., Kraus, S., and Rosenschein, J. S. 1995. Coordination without communication: Experimental validation of focal point techniques. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 102–108. Menlo Park, CA: AAAI Press / MIT Press.

Fikes, R. E., and Nilsson, N. J. 1971. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2(3/4):189–208. Winter 1971 – Amsterdam, North Holland Publishing Company.

Finin, T., Fritzson, R., McKay, D., and McEntire, R. 1994a. KQML as an agent communication language. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*. ACM Press, November, 1994. URL: **`http://www.cs.umbc.edu/kqml/papers/`**.

Finin, T., Weber, J., Wiederhold, G., Genesereth, M., Fritzson, R., McGuire, J., Shapiro, S., and Beck, C. 1994b. Specification of the KQML agent-communication language (plus example agent policies and architectures). Draft, The DARPA Knowledge Sharing Initiative—External Interfaces Working Group. URL: **`http://logic.stanford.edu/sharing/papers/`**.

Finin, T., Fritzson, R., and McKay, D. 1992. A language and protocol to support intelligent agent interoperability. In *Proceedings of the CE & CALS Washington '92 Conference*. June, 1992. URL: **`http://www.cs.umbc.edu/kqml/papers`**.

FIPA. 1997. FIPA'97 specification part 2: Agent communication language. Foundation for Intelligent Physical Agents; c/o Jean-Pierre Jacquemoud; 2, rue Bellot; CH-1206 Genève; Switzerland. URL: **`http://drogo.cselt.it/fipa/`**.

FIPA. 1998. FIPA'98 specification. Foundation for Intelligent Physical Agents; c/o Jean-Pierre Jacquemoud; 2, rue Bellot; CH-1206 Genève; Switzerland. URL: **`http://drogo.cselt.it/fipa/`**.

Fisher, M., and Wooldridge, M. 1994. Specifying and executing protocols for cooperative action. In Deen, S., ed., *Proceedings of the Second International Working Conference on Cooperating Knowledge-Based Systems (CKBS)*. Keele, U.K.: Springer-Verlag. URL: **`http://www.doc.mmu.ac.uk/STAFF/michael/`**.

Fisher, M. 1995. Representing and executing agent-based systems. In Wooldridge, M. J., and Jennings, N. R., eds., *Intelligent Agents—Proceedings of the International Workshop on Agent Theories, Architectures, and Languages (ATAL-94), held as part of ECAI-94, Amsterdam, 8–12 August, 1994*, number 890 in Lecture Notes in Computer Science, 307–323. Berlin: Springer.

Fisher, M. 1999. Representing and implementing agent architectures. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 227–241. Heidelberg: Springer-Verlag.

Foley, J. D., van Dam, A., Feiner, S. K., and Hughes, J. F. 1990. *Computer Graphics Principles and Practice.* Reading, MA: Addison-Wesley, second edition.

Frake, C. O. 1961. The diagnosis of disease among the Subanun of Mindanao. *American Anthropologist* 63:113–132.

Frake, C. O. 1969. The ethnographic study of cognitive systems. In Tyler, S. A., ed., *Cognitive Anthropology.* New York: Holt, Rinehart and Winston Inc. 28–41.

Frank, R. H. 1988. *Passions Within Reason.* New York: Norton.

Franklin, S., and Graesser, A. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In Müller, J. P., Wooldridge, M. J., and Jennings, N. R., eds., *Intelligent Agents III—Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96), ECAI'96 Workshop, Budapest, Hungary*, number 1193 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 21–35. URL: **`http://www.msci.memphis.edu/~franklin/aagents.html`**.

Galliers, J. 1990. The positive role of conflict in cooperative multi-agent systems. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A.I.—Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'89), 16–18 August, Cambridge, 1989.* Amsterdam: Elsevier Science B.V.

Gasser, L., and Huhns, M. N., eds. 1989a. *Distributed Artificial Intelligence*, volume II of *Research Notes in Artificial Intelligence.* London / San Mateo, CA: Pitman / Morgan Kaufmann.

Gasser, L., and Huhns, M. N. 1989b. Themes in distributed artificial intelligence research. In Gasser, L., and Huhns, M. N., eds., *Distributed Artificial Intelligence*, volume II. London / San Mateo, CA: Pitman / Morgan Kaufmann. vii–xv.

Gasser, L. 1987. Distribution and coordination of tasks among intelligent agents. In *Proceedings of the First Scandinavian Conference on Artificial Intelligence.* Tromsø, Norway, March, 1987.

Gasser, L. 1991. Social conceptions of knowledge and action: DAI foundations and open system semantics. *Artificial Intelligence* 47(1–3):107–138.

Gasser, L. 1995. Introduction to multi-agent systems – part 1. Notes for Tutorial A, First International Conference on Multi-Agent Systems (ICMAS'95), San Francisco.

Genesereth, M. R., and Ketchpel, S. P. 1994. Software agents. *Communications of the ACM* 37(7):48–53. URL: **`http://logic.stanford.edu/sharing/papers/`**.

Genesereth, M. R., and Singh, N. P. 1994. A knowledge sharing approach to software interoperation. Unpublished manuscript. URL: **`http://logic.stanford.edu/sharing/papers/`**.

Genesereth, M. R., Fikes, R. E., et al. 1992. Knowledge interchange format version 3.0 reference manual. Technical Report Report Logic-92-1, Stanford University, Computer Science Department, Logic Group, Stanford, CA. URL: **`http://logic.stanford.edu/sharing/papers/`**.

Georgeff, M. P., and Lansky, A. L. 1990. Reactive reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI'97), 13–17 July, 1987, Seattle, WA,* 677–682. Manlo Park, CA: AAAI Press / MIT Press.

Georgeff, M. P. 1987. Many agents are better than one. In Brown, F. M., ed., *The Frame Problem in Artificial Intelligence.* Los Altos, CA: Kaufmann. 59–75.

Gilbert, N., and Conte, R., eds. 1995. *Artificial Societies: The Computer Simulation of Social Life.* London: UCL Press.

Gilbert, N., and Doran, J., eds. 1994. *Simulating Society: The Computer Simulation of Social Phenomena.* London: UCL Press.

Gmytrasiewicz, P. J., and Durfee, E. H. 1993. Toward a theory of honesty and trust among communicating autonomous agents. *Group Decision and Negotiations* 2:237–258. Special Issue on Distributed Artificial Intelligence. URL: `http://ai.eecs.umich.edu/diag/RMM.html`.

Gmytrasiewicz, P. J. 1996. On reasoning about other agents. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995,* number 1037 in Lecture Notes In Artificial Intelligence, 143–155. Berlin: Springer-Verlag.

Green, D. W., et al. 1996. *Cognitive Science: an Introduction.* Oxford: Blackwell.

Green, S., Hurst, L., Nangle, B., Cunningham, P., Somers, F., and Evans, R. 1997. Software agents: A review. Technical Report TCD-CS-1997-06, Trinity College, University of Dublin, Dublin.

Gruber, T. R. 1993a. Toward principles for the design of ontologies used for knowledge sharing. In Guarino, N., and Poli, R., eds., *Formal Ontology in Conceptual Analysis and Knowledge Representation.* Kluwer Academic Publishers. URL: `http://www-ksl.stanford.edu/knowledge-sharing/papers/`.

Gruber, T. R. 1993b. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5(2):199–220. URL: `http://www-ksl.stanford.edu/knowledge-sharing/papers/`.

Haddadi, A. 1995. *Communication and Cooperation in Agent Systems: A Pragmatic Theory.* Number 1056 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Hales, D. 1998a. Artificial societies, theory building and memetics. In *Proceedings of the Memetics Symposium, held as part of the 15th International Congress of Cybernetics.* Namur, Belgium, 24–28 August, 1998. URL: `http://cswww.essex.ac.uk/projects/res/daphal`.

Hales, D. 1998b. Selfish memes & selfless agents—altruism in the swap shop. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris,* 431–432. Washington: IEEE Computer Society Press. Extended Abstract. URL: `http://cswww.essex.ac.uk/projects/res/daphal`.

Hales, D. 1998c. Stereotyping, groups and cultural evolution: A case of "second order emergence"? In Sichman, J. S., Conte, R., and Gilbert, N., eds., *Proceedings of the First International Workshop on Multi-Agent Systems and Agent-Based Simulation (MABS'98), held as part of the Agents' World, Paris, 4–7 July, 1998,* number ? in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. URL: `http://cswww.essex.ac.uk/projects/res/daphal`.

Halliday, D., Resnick, R., and Walker, J. 1993. *Fundaments of Physics.* New York: John Wiley & Sons, fourth edition.

Hayes-Roth, B., Brownston, L., and van Gent, R. 1995. Multiagent collaboration in directed improvisation. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA,* 148–154. Menlo Park, CA: AAAI Press / MIT Press.

Hewitt, C. 1991. Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence* 47:79–106.

Hewitt, M. A. 1997. *PiZA: Prolog Z Animator, User Guide, version 1.0.9.* URL: `http://www.noodles.demon.co.uk/PiZA/PiZADocs.html`.

Hübner, J. F., da Rocha Costa, A. C., and Bordini, R. H. 1995. Identificação de papéis por apresentação na entrada de agentes em sociedades abertas. *Dynamis* 3(13):41–53. Blumenau, Brazil. URL: `http://www.cs.ucl.ac.uk/staff/R.Bordini`.

Hübner, J. F. 1995. Migração de agentes em sistemas multi-agentes abertos. Master's thesis, Curso de Pós-Graduação em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Huhns, M. N., and Singh, M., eds. 1997. *Readings in Agents.* San Mateo, CA: Morgan Kaufmann.

Huhns, M. N., ed. 1987. *Distributed Artificial Intelligence.* Research Notes in Artificial Intelligence. London / San Mateo, CA: Pitman / Morgan Kaufmann.

Huhns, M. N. 1995. Introduction to multi-agent systems – part 2. Notes for Tutorial D, First International Conference on Multi-Agent Systems (ICMAS'95), San Francisco.

Hymes, D. 1977. *Foundations in Sociolinguistics: An Ethnographic Approach.* London: Tavistock Publications.

Iglesias, C. A., Garijo, M., and González, J. C. 1999. A survey of agent-oriented methodologies. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998,* number 1555 in Lecture Notes in Artificial Intelligence, 317–330. Heidelberg: Springer-Verlag.

Ishida, T., Gasser, L., and Yokoo, M. 1992. Organization self-design of distributed production systems. *IEEE Transactions on Knowledge and Data Engineering* 4(2).

Jennings, N. R., Hogg, T., Grossner, C., and Lemaitre, C. 1997. Applications of multi-agent systems: Prospects and promises. In Hewitt, C., ed., *International Workshop on Multi-Agent Systems (IWMAS).* Cambridge, MA: Massachusetts Institute of Technology, 12-14 October, 1997. Unpublished to date.

Jennings, N. R., Sycara, K. P., and Wooldridge, M. J. 1998. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems* 1(1):7–38.

Jia, X. 1995. *ZTC: A Z Type Checker, User's Guide, version 2.01.* Division of Software Engineering, School of Computer Science, Telecommunication, and Information Systems, DePaul University, Chicago, Illinois. Available on anonymous ftp at `ise.cs.depaul.edu`.

Kalenka, S., and Jennings, N. R. 1998. Socially responsible decision making by autonomous agents. In *Fifth International Colloquium on Cognitive Science.* San Sebastian, Spain. (Invited Paper). URL: `http://www.elec.qmw.ac.uk/dai/pubs/`.

Kinny, D. 1999. The **Agentis** agent interaction model. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 331–344. Heidelberg: Springer-Verlag.

Klusch, M., and Shehory, O. 1996. Coalition formation among rational information agents. In Van de Velde, W., and Perram, J., eds., *Proceedings of the Seventh Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96), 22–25 January, Eindhoven, The Netherlands*, number 1038 in Lecture Notes in Artificial Intelligence, 204–217. London: Springer-Verlag. URL: `http://falbala.informatik.uni-kiel.de/inf/Kandzia/index-us.html`.

Labrou, Y., and Finin, T. 1994. A semantics approach for KQML—a general purpose communication language for software agents. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*. ACM Press, November, 1994. URL: `http://www.cs.umbc.edu/kqml/papers/`.

Labrou, Y. 1996. *Semantics for an Agent Communication Language.* Ph.D. Dissertation, Computer Science and Electrical Engineering Department, University of Maryland Graduate School, Baltimore, MD.

Lamb, W. 1925. *Plato in Twelve Volumes: III Lysis / Symposium / Gorgias—with an English translation by W.R.M. Lamb.* London: Heinemann. The Loeb Classical Library.

Langdon, R. 1975. *The Lost Caravel.* Sydney: Pacific Publications.

Lave, J., and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation.* Learning in Doing: Social, Cognitive, and Computational Perspectives. Cambridge University Press.

Lave, J. 1988. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life.* Cambridge University Press.

Lesser, V., and Gasser, L., eds. 1995. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA.* Menlo Park, CA: AAAI Press / MIT Press.

Lesser, V. R. 1998. Reflections on the nature of multi-agent coordination and its implications for an agent architecture. *Autonomous Agents and Multi-Agent Systems* 1(1):89–111.

Luck, M., and d'Inverno, M. 1995. A formal framework for agency and autonomy. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 254–260. Menlo Park, CA: AAAI Press / MIT Press.

Luck, M., and d'Inverno, M. 1996. Formalising the contract net as a goal-directed system. In Van de Velde, W., and Perram, J., eds., *Agents Breaking Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, number 1038 in Lecture Notes in Artificial Intelligence. Eindhoven: Springer-Verlag. 72–85.

Luck, M., Griffiths, N., and d'Inverno, M. 1997. From agent theory to agent construction: A case study. In Müller, J. P., Wooldridge, M. J., and Jennings, N. R., eds., *Intelligent Agents III—Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96), ECAI'96 Workshop, Budapest, Hungary*, number 1193 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 49–63.

Macdonell, A. G. 1933. *England, their England.* London: Macmillian.

Manela, M. 1993. *Contributions to the Theory and Applications of Genetic Algorithms.* Ph.D. Dissertation, University of London.

Martin, R. M. 1959. *Toward a Systematic Pragmatics.* Amsterdam: North-Holland.

Mathieu, P.-L., and Lacambe, G. 1997. *The Gustave Moreau Museum.* 49 rue Étienne Marcel, 75001, Paris: Édition de la Reunion des Musées Nationaux.

Mayfield, J., Labrou, Y., and Finin, T. 1995. Desiderata for agent communication languages. In *Proceedings of the AAAI Symposium on Information Gathering from Heterogeneous, Distributed Environments, AAAI-95 Spring Symposium.* Stanford University, Stanford, CA, March 27–29, 1995. URL: `http://www.cs.umbc.edu/kqml/papers`.

Mayfield, J., Labrou, Y., and Finin, T. 1996. Evaluation of KQML as an agent communication language. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 347–360. Berlin: Springer-Verlag.

Minsky, M. 1986. *The Society of Mind.* New York: Touchstone.

Mullen, T., and Wellman, M. P. 1996. Some issues in the design of market-oriented agents. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 283–298. Berlin: Springer-Verlag.

Müller, J. P. 1999. The right agent (architecture) to do the right thing. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 211–225. Heidelberg: Springer-Verlag.

Munch, P. 1971. *Crisis in Utopia.* London: Longman.

Norman, T. J., and Long, D. 1996. Alarms: An implementation of motivated agency. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 219–234. Berlin: Springer-Verlag.

Norman, T. J. F. 1997. *Motivation-Based Direction of Planning Attention in Agents with Goal Autonomy.* Ph.D. Dissertation, University of London.

Nye, A., ed. 1993. *Xlib Reference Manual for X11 Releases 4 and 5*, volume Two of *X Window System Series.* Sebastopol, CA: O'Reilly & Associates, third edition.

O'Hare, G. M., and Jennings, N. R., eds. 1996. *Foundations of Distributed Artificial Intelligence.* Sixth-Generation Computer Technology. Chichester: John Wiley & Sons.

Oslear, D., and Mosey, D. 1993. *The Wisden Book of Cricket Laws.* London: Stanley Paul & Co.

Ossowski, S., and García-Serrano, A. 1999. Social structure in artificial agent societies: Implications for autonomous problem-solving agents. In Müller, J. P., Singh, M. P., and Rao, A. S., eds., *Intelligent Agents V—Proceedings of the Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98), held as part of the Agents' World, Paris, 4–7 July, 1998*, number 1555 in Lecture Notes in Artificial Intelligence, 133–148. Heidelberg: Springer-Verlag.

Parunak, V. 1999. Industrial and practical applications of DAI. In Weiß, G., ed., *Multiagent Systems—A Modern Approach to Distributed Artificial Intelligence.* Cambridge, MA: MIT Press. chapter 9.

Paulson, L. C. 1991. *ML for the working programmer.* Cambridge: Cambridge University Press.

Pebody, M. 1997. *Autonomy in the Real-World: A Behaviour Based View of Autonomous Systems Control in an Industrial Product Inspection System.* Ph.D. Dissertation, University of London.

Perram, J. W., and Demazeau, Y. 1997. A multi-agent architecture for distributed constrained optimization and control. In Grahne, G., ed., *Proceedings of the Sixth Scandinavian Conference on Artificial Intelligence (SCAI'97), Helsinki, 18–20 August, 1997*, number 40 in Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press.

Plotkin, G. D. 1981. A structural approach to operational semantics. Technical report, Computer Science Department, Aarhus University, Aarhus.

Pollack, M. E. 1992. The uses of plans. *Artificial Intelligence* 57(1):43–68.

Pollack, M. E. 1998. Plan generation, plan management, and the design of computational agents. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 8–9. Washington: IEEE Computer Society Press. Extended Abstract for an Invited Talk.

Populaire, P., Demazeau, Y., Boissier, O., and Sichman, J. 1993. Description et implémentation de protocoles de communication en univers multi-agents. In *1ères Journées Francophones sur l'Intelligence Artificielle Distribuée et les Systèmes Multi-Agents.* Toulouse: AFCET & AFIA, April, 1993.

Potter, B., Sinclair, J., and Till, D. 1996. *An Introduction to Formal Specification and* Z. Hemel Hempstead: Prentice Hall, second edition.

Prietula, M., Carley, K., and Gasser, L., eds. 1998. *Simulating Organizations: Computational Models of Institutions and Groups.* Menlo Park, CA: AAAI Press.

Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1(1):81–106.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann series in machine learning. San Mateo, CA: Morgan Kaufmann.

Rao, A. S., and Georgeff, M. P. 1992. An abstract architecture for rational agents. In Rich, C., Swartout, W. R., and Nebel, B., eds., *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92), 25–29 October, Cambridge, MA*, 439–449. San Mateo, CA: Morgan Kaufmann, October, 1992.

Rao, A. S., and Georgeff, M. P. 1995. BDI agents: From theory to practice. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 312–319. Menlo Park, CA: AAAI Press / MIT Press.

Rao, A. S. 1996a. AgentSpeak(L): BDI agents speak out in a logical computable language. In Van de Velde, W., and Perram, J., eds., *Proceedings of the Seventh Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96), 22–25 January, Eindhoven, The Netherlands*, number 1038 in Lecture Notes in Artificial Intelligence, 42–55. London: Springer-Verlag. URL: `http://www.aaii.oz.au/bios/rao.html`.

Rao, A. S. 1996b. Decision procedures for propositional linear-time belief-desire-intention logics. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 33–48. Berlin: Springer-Verlag.

Reynolds, R. G. 1994. Learning to co-operate using cultural algorithms. In Gilbert, N., and Doran, J., eds., *Simulating Society: The Computer Simulation of Social Phenomena*. London: UCL Press. chapter 10, 223–244.

Ridley, M. 1996. *The Origins of Virtue*. London: Viking Press.

Rizzo, P., Veloso, M., Miceli, M., and Cesta, A. 1997. Personality-driven social behaviors in believable agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents*. Cambridge, Massachusetts, 8–10 November, 1997.

Rosenschein, J., and Genesereth, M. 1985. Deals among rational agents. In Joshi, A., ed., *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*. Los Angeles, CA: AAAI Press / Morgan Kaufmann. 91–99.

Rosenschein, J., and Zlotkin, G. 1994. *Rules of Encounter*. Cambridge, MA: MIT Press.

Rosenschein, S. J. 1987. Formal theories of knowledge in AI and robotics. Report CSLI-87-84, Center for the Study of Language and Information, Stanford, CA.

Sandholm, T. W., and Crites, R. H. 1995. Multiagent reinforcement learning in the iterated prisoner's dilemma. *Biosystems*. URL: `http://dis.cs.umass.edu/publications.html`.

Schank, R. C., and Abelson, R. P. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. The Artificial Intelligence Series. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Schank, R. C. 1975. *Conceptual Information Processing*, volume 3 of *Fundamental Studies in Computer Science*. Amsterdam: North-Holland.

Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Shoham, Y., and Tennenholtz, M. 1992. On the synthesis of useful social laws for artificial agent societies. In *NCAI-92*.

Shoham, Y. 1993. Agent-oriented programming. *Artificial Intelligence* 60:51–92.

Shorey, P. 1930. *Plato in Twelve Volumes: V The Republic—with an English translation by Paul Shorey*, volume I (Books I–V). London: Heinemann. The Loeb Classical Library.

Shorey, P. 1935. *Plato in Twelve Volumes: VI The Republic—with an English translation by Paul Shorey*, volume II (Books VI–X). London: Heinemann. The Loeb Classical Library.

Sichman, J. S., and Demazeau, Y. 1994. A first attempt to use dependence situations as a decision criterion for choosing partners in multi-agent systems. In *ECAI94 WORKSHOP.*

Sichman, J. S., and Demazeau, Y. 1995. Exploiting social reasoning to deal with agency level inconsistency. In Lesser, V., and Gasser, L., eds., *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95), 12–14 June, San Francisco, CA*, 352–359. Menlo Park, CA: AAAI Press / MIT Press.

Sichman, J. S., Demazeau, Y., Conte, R., and Castelfranchi, C. 1994. A social reasoning mechanism based on dependence networks. In Cohn, A., ed., *11^{th} European Conference on Artificial Intelligence (ECAI'94)*, 188–192. Amsterdam: ECCAI & John Wiley & Sons Ltd., August, 1994.

Sichman, J. S., Conte, R., and Gilbert, N., eds. 1998. *Multi-Agent Systems and Agent-Based Simulation, Proceedings of the International Workshop held as part of the Agents' World, Paris, 4–7 July*, number 1534 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. To appear in December 1998.

Sichman, J. S. 1998. DEPINT: Dependence-based coalition formation in an open multi-agent scenario. *Journal of Artificial Societies and Social Simulation* 1(2). <http://www.soc.surrey.ac.uk/JASSS/1/2/3.html>.

Simon, H. A. 1990. A mechanism for social selection and successful altruism. *Science* 250(4988):1665–1668.

Singh, M. P., and Asher, N. M. 1991. Towards a formal theory of intentions. In van Eijck, J., ed., *Logics in AI: Proc. of the European Workshop JELIA'90.* Berlin, Heidelberg: Springer-Verlag. 472–486.

Singh, N. P., Genesereth, M. R., and Syed, M. A. 1995. A distributed and anonymous knowledge sharing approach to software interoperation. *International Journal of Cooperative Information Systems* 4(4):339–367. URL: **`http://logic.stanford.edu/sharing/papers/`**.

Singh, M. P., Huhns, M. N., and Stephens, L. M. 1993. Declarative representations of multiagent systems. *IEEE Transactions on Knowledge and Data Engineering* 5(5):721–739.

Singh, M. P., Rao, A. S., and Georgeff, M. P. 1999. Formal methods in DAI: Logic-based representation and reasoning. In Weiß, G., ed., *Multiagent Systems—A Modern Approach to Distributed Artificial Intelligence.* Cambridge, MA: MIT Press. chapter 8. URL: **`http://www.csc.ncsu.edu/faculty/mpsingh/papers/mas/formal-DAI.ps`**.

Singh, M. P. 1994. *Multiagent Systems—A Theoretic Framework for Intentions, Know-How, and Communications.* Number 799 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Singh, M. P. 1996. Semantical considerations on some primitives for agent specification. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 49–64. Berlin: Springer-Verlag.

Singh, M. P. 1998. Developing formal specification to coordinate heterogeneous autonomous agents. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 261–268. Washington: IEEE Computer Society Press.

Sissons, R., and Stoddart, B. 1984. *Cricket and Empire: The 1932–33 Bodyline Tour of Australia.* London: George Allen & Unwin.

Sloman, A., and Poli, R. 1996. SIM_AGENT: A toolkit for exploring agent designs. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995*, number 1037 in Lecture Notes In Artificial Intelligence, 392–407. Berlin: Springer-Verlag.

Smith, I. A., and Cohen, P. R. 1995. Toward a semantics for a speech act based agent communications language. In *Workshop on Intelligent Information Agents, held as part of the Fourth International Conference on Information and Knowledge Management (CIKM'95).* Baltimore, MD: ACM, 28 November to 2 December, 1995. URL: `http://www.cs.umbc.edu/~cikm/iia/proc.html`.

Smith, I. A., and Cohen, P. R. 1996. Toward a semantics for an agent communications language based on speech-acts. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI96), 4–8 August, Portland, OR*, volume 1, 24–31. Menlo Park, CA: AAAI Press / MIT Press. URL: `http://www.cse.ogi.edu/CHCC/`.

Smith, T., and the Association of Cricket Umpires. 1993. *Cricket Umpiring and Scoring.* London: Weidenfeld and Nicolson, revised edition.

Smith, I. A., Cohen, P. R., Bradshaw, J. M., Greaves, M., and Holmback, H. 1998. Designing conversation policies using joint intention theory. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 269–276. Washington: IEEE Computer Society Press.

Smith, R. G. 1980. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers* C-29(12):1104–1113.

Soltis, J., Boyd, R., and Richerson, P. J. 1995. Can group-functional behaviours evolve by cultural group selection? An empirical test. *Current Anthropology* 36:473–94.

Spivey, J. M. 1992. *The Z Notation: A Reference Manual.* Hemel Hempstead: Prentice Hall, second edition.

Steiner, D. 1998. Industrial applications of multi-agent technology. In Demazeau, Y., ed., *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS'98), Agents' World, 4–7 July, Paris*, 12–13. Washington: IEEE Computer Society Press. Extended Abstract for an Invited Talk.

Sycara, K. P. 1998. Multiagent systems. *AI Magazine* 19(2):79–92.

Tarditi, D. R., and Appel, A. W. 1991. *ML-Yacc User's Manual.* School of Computer Science, Carnegie Mellon University, Pittsburgh. FTP: {princeton.edu/pub/ml, research.att.com/dist/ml}.

The Agent Society. 1998. Reference model for open agent systems. URL: `http://www.agent.org/`.

Thomas, Jr., G. B., and Finney, R. L. 1993. *Calculus and Analytic Geometry.* Reading, MA: Addison-Wesley, eighth edition.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57.

Tyler, S. A., ed. 1969. *Cognitive Anthropology.* New York: Holt, Rinehart and Winston Inc.

Van Raalte, T., ed. 1993. *XView Reference Manual for XView Version 3.2.* X Window System Series. Sebastopol, CA: O'Reilly & Associates.

Vanderveken, D. 1990. On the unification of speech act theory and formal semantics. In Cohen, P. R., Morgan, J., and Pollack, M. E., eds., *Intentions in Communication.* Cambridge, MA: MIT Press. chapter 11, 195–220.

Verharen, E. M. 1997. *A Language-Action Perspective on the Design of Cooperative Information Agents.* Ph.D. Dissertation, Katholieke Universiteit Brabant, Tilburg, The Netherlands.

Vidal, J. M., and Durfee, E. H. 1996a. Using recursive agent models effectively. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995,* number 1037 in Lecture Notes In Artificial Intelligence, 171–186. Berlin: Springer-Verlag.

Vidal, J. M., and Durfee, E. H. 1996b. The impact of nested agent models in an information economy. In Durfee, E., ed., *Proceedings of the Second International Conference on Multi-Agent Systems (ICMAS'96), 11–13 December, Kyoto, Japan,* 377–384. Menlo Park, CA: AAAI Press. URL: `http://ai.eecs.umich.edu/diag/RMM.html`.

Vieira, R., and da Rocha Costa, A. C. 1993. The acceptance relation and the specification of communicating agents. In Schlageter, G., Huhns, M., and Papazoglou, M., eds., *Proceedings of the First International Conference on Intelligent and Cooperative Information Systems – Special Track in Issues on Cooperating Heterogeneous Intelligent Agents,* 247–255. Rotterdam, The Netherlands: IEEE Computer Society Press, May, 1993.

Watt, D. A. 1986. Executable semantic descriptions. *Software—Practice & Experience* 16(1):13–43.

Watt, D. A. 1991. *Programming Language Syntax and Semantics.* International Series in Computer Science. New York: Prentice Hall.

Weiß, G., and Sen, S., eds. 1996. *Adaption and Learning in Multi-Agent Systems.* Number 1042 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Weiß, G., ed. 1997. *Distributed Artificial Intelligence Meets Machine Learning—Learning in Multiagent Environments.* Number 1221 in Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Weiß, G., ed. 1999. *Multiagent Systems—A Modern Approach to Distributed Artificial Intelligence.* Cambridge, MA: MIT Press.

Wellman, M. P. 1993. A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research* 1:1–23.

Wellman, M. P. 1994. A computational market model for distributed configuration design. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI94), 31 July to 4 August, Seattle, WA,* volume 1, 401–407. Menlo Park, CA: AAAI Press.

Wellman, M. P. 1995. Market-oriented programming: Some early lessons. In Clearwater, S., ed., *Market-Based Control: A Paradigm for Distributed Resource Allocation.* World Scientific. chapter 4. URL: `http://ai.eecs.umich.edu/people/wellman/Publications.html`.

Werner, E. 1988. A formal computational semantics and pragmatics of speech acts. In *COLING-88: Proceedings of the 12th International Conference on Computational Linguistics.* Budapest, Hungary.

Werner, E. 1989. Cooperating agents: A unified theory of communication and social structure. In Gasser, L., and Huhns, M. N., eds., *Distributed Artificial Intelligence,* volume 2. London / San Mateo, CA: Pitman / Morgan Kaufmann. 3–36.

Wiederhold, G., and Genesereth, M. 1995. The basis for mediation. In *Proceedings of COOPIS'96 Conference,* 138–155. Vienna, Austria, May, 1995. To appear in IEEE Expert, 1997. Available from {coopis@cs.toronto.edu}. URL: **`http://db.stanford.edu/pub/gio/gio-papers.html`**.

Wiederhold, G. 1994. Interoperation, mediation, and ontologies. In *Proceedings of the International Symposium on Fifth Generation Computer Systems (FGCSOB94), Workshop on Heterogeneous Cooperative Knowledge-Bases,* volume W3, 33–48. Tokyo, Japan: ICOT, December, 1994. URL: **`http://db.stanford.edu/pub/gio/gio-papers.html`**.

Wooldridge, M., and Jennings, N. R. 1995. Intelligent agents: Theory and practice. *The Knowledge Engineering Review* 10(2):115–152. URL: **`http://www.elec.qmw.ac.uk/dai/pubs`**.

Wooldridge, M., Müller, J. P., and Tambe, M. 1996. Agent theories, architectures, and languages: A bibliography. In Wooldridge, M., Müller, J. P., and Tambe, M., eds., *Intelligent Agents II—Proceedings of the Second International Workshop on Agent Theories, Architectures, and Languages (ATAL'95), held as part of IJCAI'95, Montréal, Canada, August 1995,* number 1037 in Lecture Notes In Artificial Intelligence, 408–431. Berlin: Springer-Verlag.

Wooldridge, M. 1997. Issues in agent-based software engineering. In Kandzia, P., and Klusch, M., eds., *Proceedings of the First International Workshop on Cooperative Information Agents (CIA'97),* volume 1202 of *Lecture Notes in Artificial Intelligence,* 1–18. Kiel, Germany: Springer-Verlag, 26–28 February, 1997. Invited paper.

Wooldridge, M. 1998. Agent-based computing. *Interoperable Communication Networks* 1(1):71–97.

Wyeth, G. 1997. An introductory course in mechatronics: Robo-Cricket. In *Proceedings of the International Conference on Mechatronics and Machine Vision in Practice,* 20–25. Toowoomba, Australia: IEEE Computer Society Press, 23–25 September, 1997. URL: **`http://www.elec.uq.edu.au/~ wyeth/pub.html`**.

Zlotkin, G., and Rosenschein, J. S. 1994. Coalition, cryptography, and stability: Mechanism for coalition formation in task oriented domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI94), 31 July to 4 August, Seattle, WA,* volume 1, 432–437. Menlo Park, CA: AAAI Press.

Glossary

Some of the main concepts (and corresponding abbreviations used, if any) that we have introduced are repeated here to facilitate the reader's reference.

Acceptance Relation: Acceptance is an empirical relation between users and sentences of a language, observed by an experimenter at a certain time who asks questions by means of a set of sentences forming a logical theory. 67

Anthropologist Agent (AA): Agents that migrate to societies with the explicit intention of producing a cultural description, so as to facilitate the adaptation of *migrant agents*. 48

Bankrupt-Excluded (BE): A simulation condition where all agents play at each simulation step, except those that are *bankrupt*, i.e., have run too low in points. 151

Co-Intensiveness: The relation between terms that have the same quasi-intension (of any kind). 68

Fair Shares (FS): A society in which the wealth classification thresholds vary according to points earned in the recent past by all agents in the group. 150

Generous Middle Class (GM): A society in which an agent whose wealth was classified as "medium" also plays the MS strategy. 150

Informant Agent (IA): The computational equivalent of an anthropologist's informant, i.e., an agent that provides the anthropologist agent with information about its specific artificial culture. 54

Intensional Ontology of Terms (IOT): A set of terms where each one is associated with the minimal set of predicates that is necessary and sufficient to distinguish (unequivocally) itself from every other term in the universe of discourse of a communicating society of agents. 68

Intersubjective Quasi-Intension: This concept regards groups of language users, rather than individuals, at a certain time. It is the equivalence class of all the *subjective quasi-intensions* of a certain group of users of the language. 67

Intertemporal Quasi-Intensions: Quasi-intensions that are relative to a particular language user at all times. 68

Legitimate Peripheral Participation (LPP): A concept related to situated learning which emphasises that learning is the accomplishment of an individual who engages progressively in the social practices that are characteristic of a community. 60

Long Memory (LM): A society in which all past earnings are taken into account in the calculations for wealth classification (as opposed to having a *history length*). 149

Migrant Agent (MA): Agents that migrate to other societies for the sake of accomplishing their own goals, or the social goals of the target societies, therefore promoting interoperability of MAS. 48

Objective Quasi-Intensions: At the same time intertemporal and intersubjective quasi-intension of expressions. 68

Polarised Society (PS): A society in which the thresholds were set in such a way that the size of the middle class is significantly reduced. 150

Quasi-intension: Linguistic reductions of the mental entities relative to intensions. 67

Societal Quasi-Intension: The quasi-intesion related to a particular group of agents. 68

Some-Play (SP): A simulation condition where only a variable percentage of the agents are selected to play at each time, and they are not awarded points to recover but are allowed to play even with negative balances in their "accounts." 151

Subjective Quasi-Intension: This notion concerns an individual constant (term), and is defined as the properties a language user associates with the term, as expressed in the sentences that the given user *accepts* at a certain time. 67

Index

Numbers typeset in italics refer to the page where the corresponding entry is *described*, the ones underlined to (formal) *definitions*, the rest to the pages where the entry is simply used (the ones in bold refer to pages where special use was made of the entry). A number followed by '*n*' means that the entry is in a footnote in that page.