Understanding Autonomous Interaction

Mark d'Inverno[1] and **Michael Luck**[2]

Abstract. Autonomy is a necessary part of the design of agents flexible enough to function effectively and efficiently in a sophisticated world. Much work, however, has taken a very restricted view of what is entailed by autonomous interaction; in particular, the effects of an interaction have, to some extent, been guaranteed. In this paper, we argue that no facet of interaction can ever be guaranteed, and that if agents are to be autonomous, they must be able to cope with this inherent uncertainty. We propose a model of autonomous interaction in response, which addresses these concerns, and which can be viewed as a process of motivated discovery. This approach has two important aspects: first, modelling the motivations of the agent allows a more adequate model of autonomy to be achieved, and also provides a control strategy for the process of interaction; second, the discovery paradigm provides a suitable framework for effective action and reasoning in an uncertain environment.

1 Introduction

This paper is concerned with the design of autonomous agents: what it means for an agent to be autonomous and what that entails for any adequate model of interaction between such agents. Complex environments admit an inherent uncertainty that must be considered if we are to cope with more than just toy problems. In such an uncertain environment, an agent must be autonomous; an agent cannot know in advance the exact effects of its or others' actions. This is of paramount importance, and an agent must therefore be designed with a flexibility that enables it to cope with this uncertainty by evaluating it and responding to it in adequate ways.

Autonomy allows the design of agents flexible enough to function effectively and efficiently in a sophisticated world [4]. Typically, real autonomy has been neglected in most research. We hear of benevolent, altruistic, trusting, sympathetic or cooperative agents, yet a truly autonomous agent will behave only in a selfish way. Cooperation, for example, should occur only as a consequence of an agent's selfishness. Autonomy allows for no artificially imposed rules of behaviour; all behaviour must be a consequence of the understanding and processing capabilities of that agent. Modelling this fundamental notion of selfish behaviour and the generation of goals by such a selfish autonomous agent is of vital importance in the design of autonomous agents.

Autonomy is independence. It is a state that does not rely on any external entity for purposeful existence.

In this paper, we use an existing agent framework previously used to respecify Social Dependence Networks [7] and the Contract Net Protocol [8] to address the issues that arise in a consideration of autonomous interaction. We begin by considering several problems that are prevalent in existing models of interaction, and which must be addressed in attempting to construct a model of autonomous interaction. Then we introduce a previously developed agent framework on which the remainder of the paper is based. The next sections describe and specify an autonomous social agent that acts in an environment, the way in which it generates its goals, and finally how it interacts with others in its environment. We discuss how this can be viewed as a process of discovery, and what such a view usefully brings to the problem.

[1] School of Computer Science, University of Westminster, 115 New Cavendish Street, London W1M 8JS, UK

[2] Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

ECAI 96. *12th European Conference on Artificial Intelligence*
Edited by W. Wahlster
Published in 1996 by John Wiley & Sons, Ltd.

2 Autonomous Interaction

In multi-agent systems, the interactions between agents are the basis for usefully exploiting the capabilities of others. However, such a pragmatic approach has not been the concern of many researchers who instead often focus on small areas of interaction and communication, and in particular on specialised forms of intention recognition and interpretation.

In many existing models of interaction, agents are not autonomous. In considering these models, we can identify problem-issues in autonomous interaction. Our intention is simply to show why these models are not adequate for autonomous interaction, and so isolate problems which contribute to the non-autonomous nature of these models.

Pre-determined Agenda Problem-solving can be considered to be the task of finding actions that achieve current goals. Typically, goals have been presented to systems without regard to the problem-solving agent so that the process is divorced from the reality of an agent in the world. This is inadequate for models of autonomy which require an understanding of how such goals are generated and adopted. Surprisingly, however, this is an issue which has received very little attention with only a few notable exceptions (e.g. [14]).

Benevolence In traditional models of goal adoption, goals are broadcast by one agent, and adopted by other agents according to their own relevant competence [17]. This assumes that agents are already designed with common or non-conflicting goals that facilitate the possibility of helping each other satisfy additional goals. Negotiation as to how these additional goals are satisfied typically takes the form of mere goal-node allocation. Thus an agent simply has to communicate its goal to another agent for cooperation in the form of joint planning to ensue. The concept of benevolence — that agents will cooperate with other agents whenever and wherever possible — has no place in modelling autonomous agents [6, 9]. 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.

Guaranteed Effects Speech Act Theory (SAT) [3, 16] underlies much existing work in AI [5], typically because as Appelt points out, speech acts are categorizable and can be modelled as action operators in a planning environment [2]. However, this work admits a serious flaw. Although the preconditions of these operators are formulated in terms of the understanding of the planning agent, the post-conditions or effects of these operators do not update the understanding of the planning agent, but of the agent at whom the action is directed [1]. No agent can ever actually *know* with any certainty anything about the effects of an action, whether communicative or otherwise. It is only through an understanding of the *target* agent and through observing the future behaviour of that agent, that the agent can discover the actual effects of the interaction. This uncertainty is inherent in communication between autonomous agents and must be a feature of any model of interaction which hopes to reflect this reality.

Automatic Intention Recognition A related though distinct problem with using SAT in the design of *communication* models involves the notion that the meaning of an utterance is a function of the linguistic content of that utterance. SAT is unable (even when one tries to force rather undistinguished ad-hoc rules [11]) to model any kind of utterance where the linguistic content is not very close to the speaker's intended meaning. That is to say that the operators themselves are context independent, and information about how context affects the interpretation of utterances is not explicitly captured. Communication varies from utterances with a meaning identical to linguistic content through utterances which have a meaning opposite to the linguistic content to utterances where the meaning does not seem to be categorised at all by the linguistic content. In short, Speech Act Theory cannot lead to a model of autonomous interaction. It merely serves to describe a very limiting case of linguistic communication at a suitable level for planning operators. A more flexible account of how intention is recovered from a multitude of different utterances is required.

Multi-Agent Modelling This is another related but more subtle problem. Much work has modelled communicative actions in terms of mutual beliefs about the operator and its known effects [15]. This proposes to show not only how certain mental states lead to speech actions, but how speech actions affect mental states. We argue that any account of autonomous interaction should only model the effects of an action upon the mental state of the agent initiating the interaction (or another single agent).

In summary, there are several important claims here: first, an agent cannot be truly autonomous if its goals are provided by external sources; second, an agent will only adopt a goal and thus engage in an interaction if it is to its advantage to do so; third, the effects of an interaction cannot be guaranteed; fourth, the intentions of others cannot always be recognised; fifth, an agent can only know about itself.

Note that the first claim requires goals to be generated from within. It is this internal goal generation that demands an explicit model of the motivations of the agent. The second claim requires a notion of advantage that can only be determined in relation to the motivations of the agent. The third and fourth claims demand that the uncertain nature of autonomous interaction be explicitly addressed. We argue that viewing autonomous interaction as motivated discovery provides us with a means for doing this. Finally, the fifth claim imposes constraints on the problem we are considering, and provides a strong justification for our concern with constructing a model of autonomous interaction from the perspective of an individual agent.

3 The Agent Framework

We must first provide some basic concepts using the specification language Z [18] on which to base a subsequent analysis. The world is made up of basic *entities* which can be instantiated as objects, agents and autonomous agents [12]. An entity consists of four constituents as follows: a set of attributes, which are perceivable qualities of the entity; a set of actions, which define the basic capabilities of the agent; a set of goals, which are the goals that can be ascribed to the entity which characterise its $agency$; and a set of internal non-derivable motivations which define an entity's $autonomy$.

$$
\begin{array}{|l}
\hline
\textit{Entity} \\
attributes : \mathbb{P}\ Attribute \\
capableof : \mathbb{P}\ Action \\
goals : \mathbb{P}\ Goal \\
motivations : \mathbb{P}\ Motivation \\
\hline
attributes \neq \{\} \\
\hline
\end{array}
$$

An autonomous agent is an entity which can generate its own goals through motivations. It is an entity with non-empty sets of capabilities, goals and motivations.

$$
\begin{array}{|l}
\hline
\textit{AutonomousAgent} \\
Entity \\
\hline
capableof \neq \{\} \wedge goals \neq \{\} \wedge motivations \neq \{\} \\
\hline
\end{array}
$$

A full treatment, with justifications, of the different entities which arise in this view of the world can be found in [12].

4 A Model of Autonomous Interaction

An interaction episode is initiated by an agent in attempting to satisfy a goal demanded by its motivations, generating a prediction specifying an action intended to achieve the desired effect in the current environment. This section examines this in detail, and specifies what is involved in autonomous interaction. It begins by reviewing some necessary schemas given elsewhere [12] and used here to define a social agent, omitting some specific details due to space constraints.

4.1 Constructing a Social Agent

An environment is defined to be a simple set of attributes that describes all of the features of the world. The environment thus represents all possible percepts in a uniform way. It is convenient also to define a $View$ to be the perception of an $Environment$ by an agent in the same way.

$$
\begin{array}{l}
Environment == \mathbb{P}\ Attribute \\
View == \mathbb{P}\ Attribute
\end{array}
$$

An agent in an environment can perceive certain attributes subject to its capabilities and current state but, due to limited resources, may not be able to perceive all attributes. The action of an agent is based on a subset of attributes of the environment, the agent's *actual* percepts.

Thus in the schema below for agent perception, $perceivingactions$ is a subset of agent capabilities. Two functions specify what the agent perceives: $canperceive$ is applied to the current environment and the agent's capabilities to give potential percepts and $willperceive$ is applied to its motivations, goals and the current environment to give attributes actually perceived.

$$
\begin{array}{l}
\underline{\mathit{AutonomousAgentPercepts}} \\
\mathit{AutonomousAgent} \\
\mathit{perceivingactions} : \mathbb{P}\,\mathit{Action} \\
\mathit{canperceive} : \mathit{Environment} \longrightarrow \mathbb{P}\,\mathit{Action} \nrightarrow \\
\qquad\qquad\qquad\qquad \mathit{Environment} \\
\mathit{willperceive} : \mathbb{P}\,\mathit{Motivation} \longrightarrow \mathbb{P}\,\mathit{Goal} \longrightarrow \\
\qquad\qquad\qquad\qquad \mathit{Environment} \longrightarrow \mathit{View}
\end{array}
$$

An *interaction* with the environment occurs as a result of performing actions in it. Its effects on the environment are determined by applying the $\mathit{effectinteract}$ function in the axiom definition below to the current environment and the actions taken. It is always in scope.

$$
\begin{array}{l}
\mathit{effectinteract} : \mathit{Environment} \longrightarrow \\
\qquad\qquad \mathbb{P}\,\mathit{Action} \nrightarrow \mathit{Environment}
\end{array}
$$

To specify the actions of an autonomous agent, the next schema includes the $\mathit{AutonomousAgent}$ schema and then defines an action-selection function that is determined in relation to the motivations, goals, perceived environment and actual environment of the agent. The function gives the set of actions the agent will perform in order to achieve some goal.

$$
\begin{array}{l}
\underline{\mathit{AutonomousAgentAct}} \\
\mathit{AutonomousAgent} \\
\mathit{autoactions} : \mathbb{P}\,\mathit{Motivation} \longrightarrow \mathbb{P}\,\mathit{Goal} \longrightarrow \\
\qquad \mathit{View} \longrightarrow \mathit{Environment} \longrightarrow \mathbb{P}\,\mathit{Action}
\end{array}
$$

We also need to define the state of an agent as follows by including the previous schemas and the current environment, and introducing variables for possible percepts, $\mathit{posspcpts}$, actual percepts, $\mathit{actualpcpts}$, and the actions selected, willdo.

$$
\begin{array}{l}
\underline{\mathit{AutonomousAgentState}} \\
\mathit{AutonomousAgentPercepts} \\
\mathit{AutonomousAgentAct} \\
\mathit{env} : \mathit{Environment} \\
\mathit{posspcpts}, \mathit{actualpcpts} : \mathit{View} \\
\mathit{willdo} : \mathbb{P}\,\mathit{Action} \\
\hline
\mathit{actualpcpts} \subseteq \mathit{posspcpts} \\
\mathit{posspcpts} = \mathit{canperceive}\ \mathit{env}\ \mathit{perceivingactions} \\
\mathit{actualpcpts} = \mathit{willperceive}\ \mathit{motivations}\ \mathit{goals}\ \mathit{posspcpts} \\
\mathit{willdo} = \mathit{autoactions}\ \mathit{motivations}\ \mathit{goals}\ \mathit{actualpcpts}\ \mathit{env}
\end{array}
$$

All this defines autonomous agents, but we require more if they are to engage in interaction or communication episodes. Specifically, in order for effective interaction and communication, an agent must be able to group the attributes that make up the environment into entity-describing models so that it can identify the other individuals in the world. A *social agent* is an agent that is aware of other agents and their role and function through these models. The schema below includes models of other autonomous agents available to the social agent as $\mathit{autoagts}$. Models of agents and objects are also possible and are excluded only due to space constraints.

$$
\begin{array}{l}
\underline{\mathit{SocialAgent}} \\
\mathit{AutonomousAgentState} \\
\mathit{autoagts} : \mathbb{P}\,\mathit{AutonomousAgent} \\
\hline
\mathit{autoagts} \neq \{\}
\end{array}
$$

4.2 Goal Generation

As stated above, in order for an agent to be autonomous, it must generate *goals* from *motivations*. The initial point in any interaction is when this goal generation process occurs. In this section, we describe how an autonomous agent, *defined* in terms of its somewhat abstract *motivations*, can construct goals or concrete states of affairs to be achieved in the environment. Our model requires a repository of known goals which capture knowledge of limited and well-defined aspects of the world. These goals describe particular *states* or *substates* of the world with each autonomous agent having its own such repository. An agent tries to find a way to mitigate motivations by selecting an action to achieve an existing goal or by retrieving a goal from a repository of known goals, as considered below.

In order to retrieve goals to mitigate motivations, an autonomous agent must have some way of assessing the effects of competing or alternative goals. Clearly, the goals which make the greatest positive contribution to the motivations of the agent should be selected. The $\mathit{GenerateGoal}$ schema below describes at a high level an autonomous agent monitoring its motivations for goal generation. First, the social agent changes indicated by $\Delta\mathit{SocialAgent}$, and a new variable representing the repository of available known goals, $\mathit{goalbase}$, is declared. Then, the motivational effect on an autonomous agent of satisfying a set of new goals is given. The $\mathit{motiveffect}$ function returns a numeric value representing the motivational effect of satisfying a set of goals with a particular configuration of motivations and a set of existing goals. The predicate part specifies that all goals currently being pursued must be known goals that already exist in the goalbase. Finally, there is a set of goals in the goalbase that has a greater motivational effect than any other set of goals, and the current goals of the agent are updated to include the new goals.

$$
\begin{array}{l}
\underline{\mathit{GenerateGoal}} \\
\Delta\mathit{SocialAgent} \\
\mathit{goalbase} : \mathbb{P}\,\mathit{Goal} \\
\mathit{motiveffect} : \mathbb{P}\,\mathit{Motivation} \longrightarrow \mathbb{P}\,\mathit{Goal} \longrightarrow \mathbb{P}\,\mathit{Goal} \longrightarrow \mathbb{Z} \\
\hline
\mathit{goals} \subseteq \mathit{goalbase} \ \wedge\ \mathit{goals}' \subseteq \mathit{goalbase} \\
\exists\, \mathit{gs} : \mathbb{P}\,\mathit{Goal} \mid \mathit{gs} \subseteq \mathit{goalbase} \bullet \\
\quad (\forall\, \mathit{os} : \mathbb{P}\,\mathit{Goal} \mid \mathit{os} \in (\mathbb{P}\,\mathit{goalbase}) \bullet \\
\qquad (\mathit{motiveffect}\ \mathit{motivations}\ \mathit{goals}\ \mathit{gs} \geq \\
\qquad\quad \mathit{motiveffect}\ \mathit{motivations}\ \mathit{goals}\ \mathit{os}) \\
\qquad\qquad \wedge\ \mathit{goals}' = \mathit{goals} \cup \mathit{gs})
\end{array}
$$

4.3 Agent Interaction

Once the goals defining the purpose of the interaction are generated, the agent can continue in its attempt to achieve those goals.

Many traditional models of interaction have assumed an ideal world in which unfounded assumptions have given rise to inadequate characterisations of interaction amongst autonomous agents. If we consider autonomous interaction to be a process of uncertain outcome (which it must be), then we can characterise it in a more general way as a process of discovery in terms of the effects of actions. This allows us to deal effectively with the inherent uncertainty in interaction. In the following discussion, we will begin to introduce the language of discovery to make the relationships clear.

In order to make sense of our environment and to function effectively in it, we continually anticipate the effects of our actions and utterances — we make predictions (or expectations) about what will happen next. The action-selection function, $\mathit{autoactions}$, of the $\mathit{AutonomousAgentAct}$ schema encompasses the deliberation of the agent. The action that is selected is intended to satisfy the goals of the agent through its resulting effects and consequent changes to the environment. In the case of an interaction episode involving two agents, the initiating agent selects an action that is intended to cause the desired response in the responding agent. The uncertainty inherent in such interaction means that the effects cannot be known in advance, but can only be discovered after the event has taken place, or action performed. We describe this by specifying the *predicted* effects of actions selected in the $AutonomousAgentAct$ schema by applying the $socialeffectinteract$ function to the current view of the environment and those actions. The agent thus predicts that these actions will change the environment to achieve the desired results. Remember that the environment includes all of the entities in it, so that a change to an agent in the environment will in turn cause a change to the environment itself. We also introduce a variable to store an agent's actual percepts prior to an operation, $oldpcpts$.

$$
\begin{array}{l}
\underline{\quad SocialAgentPredict \quad} \\
SocialAgent \\
socialeffectinteract : View \longrightarrow \mathbb{P}\ Action \nrightarrow View \\
oldpcpts, prediction : View \\
\hline
prediction = socialeffectinteract\ actualpcpts\ willdo \\
prediction \cap \bigcup goals \neq \{\}
\end{array}
$$

In order to achieve the desired result, the relevant actions must be performed. Effectively, this acts as an experiment, testing whether the predictions generated are consistent with the resulting effects. In this sense, experimentation is central to this model, for such interaction with the environment is the only way in which an agent's understanding of its capabilities and its environment can be assessed to bring to light inadequacies, inconsistencies and errors. When an action is performed, it affects the models of other agents which, after the change, are derived from the previous agent models and the view of the environment through the $updateagts$ function. These models of agents are critical in determining if the action was successful.

$$
\begin{array}{l}
\underline{\quad SocialAgentInteract \quad} \\
SocialAgentPredict \\
updateagts : \mathbb{P}\ AutonomousAgent \longrightarrow View \longrightarrow \\
\qquad\qquad\qquad\qquad \mathbb{P}\ AutonomousAgent
\end{array}
$$

The action also has an effect on the environment which changes accordingly, and a similar effect on the agent itself whose percepts also change. For example, in the case of an action which issues a request to another agent to tell the current time, the resulting model will either encode the fact that the agent is telling the time, or not. By inspecting this model and its attributes, the requesting agent can determine if its action has been successful. Note that the new value of $oldpcpts$ takes the previous value of $actualpcpts$ for later use.

$$
\begin{array}{l}
\underline{\quad SocialEnv \quad} \\
\Delta SocialAgentPredict \\
SocialAgentInteract \\
\hline
env' = effectinteract\ env\ willdo \\
posspcpts' = canperceive\ env'\ perceivingactions \\
actualpcpts' = willperceive\ motivations\ goals\ posspcpts' \\
willdo' = autoactions\ motivations\ goals \\
\qquad\qquad\qquad\qquad actualpcpts'\ env' \\
autoagts' = updateagts\ autoagts\ actualpcpts' \\
oldpcpts' = actualpcpts \\
prediction' = prediction
\end{array}
$$

Evaluating the results of the actions appears simple. At the most basic level, it involves the comparison of predictions with observations. Thus if the intended effects of the actions and the actual effects match, then the actions have achieved the desired result and the episode is successful. If they are anomalous, then it reveals an erroneous understanding of the environment and the agents within it, or an inadequate capability for perception of the results. The important point here is that there is no guarantee of success, and failure can be due to any number of reasons.

This analysis assumes that the evidence is perfect, however, which may not always be appropriate. In any real environment this is not so, and error can be introduced into evidence in a variety of ways, reducing the quality of the observed evidence accordingly. Not only may there be inaccuracy due to the inherent uncertainty in both performing the actions and perception of the results (experimentation and observation respectively), but also, if the actions taken by the agent are communicative actions intended to elicit a response from another autonomous agent, then there may be inaccuracy due to malicious intent on the part of the responding agent by providing misleading information, for example [13]. Thus the response may itself be the vessel for the error.

In addition to assessing the fit of observations with predictions, therefore, the quality of the observations themselves must also be assessed in order to ascertain whether they are acceptable to be used in the comparison at all. Simple tolerance levels for assessing the acceptability of perceived evidence are inadequate, for they do not consider the need for the interaction episode, and the importance of achieving the desired result. The quality demanded of the observations can thus only be assessed in relation to the motivations of the agent which provide a measure of the importance of the situation, and take into account the implications of success and failure. In medical domains, for example, where the agents are highly motivated, even a small degree of error in interaction concerning relevant patient details may be unacceptable if it would lead to the loss of a patient's life, while neighbourly discussion of the weather with low motivations and little importance may allow a far greater error tolerance.

The schemas below describe evaluation with two functions. First, $accept$ is applied to the capabilities of the agent, its perceived environment before and after the actions were performed and the agent models, and returns a boolean value indicating whether to accept the evidence or not. The capabilities of the agent capture the uncertainty information that arises from the agent itself, while the perceived environment and agent models include details of difficulties arising through the environment, or other agents. The $consider$ function compares predictions and observations once evidence is accepted. Note that the potentially difficult question of when observations match predictions is bound up in the function itself which may be interpreted either as a simple equality test or as something more sophisticated.

The $Decide$ schema also states at the beginning that though the agent changes as a result of this evaluation ($\Delta SocialAgent$), the state of the agent remains the same ($\Xi AutonomousAgentState$). Finally, if the evidence is accepted, and the observations do not match the predictions, then the agent models must be revised in an appropriate way as specified by $revisemodels$ which is not detailed here.

$$bool ::= True \mid False$$

$$
\begin{array}{l}
\underline{\quad SocialAgentEvaluate \quad} \\
accept : \mathbb{P}\ Action \longrightarrow View \longrightarrow \\
\qquad\qquad View \longrightarrow \mathbb{P}\ AutonomousAgent \longrightarrow bool \\
consider : View \longrightarrow View \longrightarrow bool
\end{array}
$$

$$
\begin{array}{l}
\underline{\quad Decide \quad} \\
\Delta SocialAgent \\
\Xi AutonomousAgentState \\
SocialAgentPredict \\
SocialAgentEvaluate \\
revisemodels : View \longrightarrow \\
\qquad \mathbb{P}\ AutonomousAgent \longrightarrow \mathbb{P}\ AutonomousAgent \\
\hline
(accept\ capableof\ actualpcpts\ oldpcpts\ autoagts = True) \\
\wedge\ (consider\ prediction\ actualpcpts = False) \Rightarrow \\
\quad autoagts' = revisemodels\ actualpcpts\ autoagts
\end{array}
$$

4.4 Summary

This model provides a unifying basis for much related work, with a clear distinction of problem tasks. For example, Galliers' work on belief revision [10] incorporates just such evaluation, and revision: the question of *whether* to revise by use of endorsements is clearly subsumed by evaluation, and the belief revision itself is covered by the revision that arises in the case of anomalous evidence. Our model, however, makes these distinctions more explicit, and extends to cover the preceding stages of prediction, experimentation and observation.

By taking autonomous interaction to be a process of discovery, we can avoid the problems identified earlier of *guaranteed effects* and *automatic intention recognition*. In discovery, no effects are known for certain in advance, but instead, (tentative) predictions or expectations of future states of the world can be generated. It is only possible to be certain about effects once the actions have been carried out. This can lead to a re-evaluation of existing models.

Additionally, we assert that the process of autonomous communication must be motivated, and consequently a motivated agent does not have a *pre-determined agenda*, nor is it *benevolent*. Motivations provide a means by which an agent can set its own agenda, or set its own goals and determine which actions to perform in achieving them. The effects of benevolent behaviour are possible, but only through self-serving motivations. Moreover, because effects are not guaranteed, failure is always possible, but the combination of discovery and motivations allow effective exploitation of these failures and also recovery from them whenever possible.

5 Further Work and Conclusion

Our aim in constructing the model for autonomous interaction is ambitious. We are attempting to provide a common unifying framework within which different levels of abstraction of reasoning, behavioural and interaction tasks can be related and considered. This work seeks to draw on the foundations established previously both for providing a language with which to precisely discuss agent-based systems, and for providing a base upon which to develop systems and models such as that presented here. We have necessarily concentrated on a high-level specification so that the key principles can be explicated, but without sacrificing the need for preciseness through formality.

This paper has identified several important issues in autonomous interaction which are relevant to the design and construction of autonomous agents. We have described, and formally specified at a high-level, a model for autonomous interaction based on a previously developed framework, and have attempted to set it in the paradigm of discovery. The combination of the two approaches, by explicitly introducing motivated reasoning as part of the agent framework, and the capacity for effectively dealing with dynamic worlds through discovery, provides a way in which the inadequacies in existing models may be addressed.

REFERENCES

[1] J. Allen, 'A plan-based approach to speech act recognition', Technical Report 131, Department of Computer Science, University of Toronto, (1979).

[2] D. E. Appelt, *Planning English Sentences*, Cambridge University Press, 1985.

[3] J. L. Austin, *How to do Things with Words*, Oxford University Press, 1962.

[4] O. Boissier and Y. Demazeau, 'A distributed artificial intelligence view on general purpose vision systems', in *Decentralized AI 3: Proceedings of the Third European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, eds., E. Werner and Y. Demazeau, pp. 311–330. Elsevier, (1992).

[5] J. A. Campbell and M. d'Inverno, 'Knowledge interchange protocols', in *Decentralized AI: Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, eds., Y. Demazeau and J.-P. Müller, 63–80, Elsevier, (1990).

[6] C. Castelfranchi, 'Social power', in *Decentralized AI: Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, eds., Y. Demazeau and J.-P. Müller, pp. 49–62. Elsevier, (1990).

[7] M. d'Inverno and M. Luck, 'A formal view of social dependence networks', in *Proceedings of the First Australian DAI Workshop, Lecture Notes in Artificial Intelligence, 1087*, pp. 115–129. Springer Verlag, (1996).

[8] M. d'Inverno and M. Luck, 'Formalising the contract net as a goal directed system', in *Agents Breaking Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World, LNAI 1038*, eds., W. Van de Velde and J. W. Perram, pp. 72–85. Springer-Verlag, (1996).

[9] J. R. Galliers, 'The positive role of conflicts in cooperative multi-agent systems', in *Decentralized AI: Proceedings of the First European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, eds., Y. Demazeau and J.-P. Müller, Elsevier, (1990).

[10] J. R. Galliers, 'Modelling autonomous belief revision in dialogue', in *Decentralized AI 2: Proceedings of the Second European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, eds., Y. Demazeau and J.-P. Müller, 231–245, Elsevier, (1991).

[11] H. P. Grice, 'Logic and conversation', in *Syntax and Semantics, Volume 3: Speech Acts*, eds., P. Cole and J. L. Morgan, 41–58, Academic Press, (1975).

[12] M. Luck and M. d'Inverno, 'A formal framework for agency and autonomy', in *Proceedings of the First International Conference on Multi-Agent Systems*, pp. 254–260. AAAI Press / MIT Press, (1995).

[13] S. Marsh, 'Trust in distributed artificial intelligence', in *Artificial Social Systems: Selected Papers from the Fourth European Workshop on Modelling Autonomous Agents in a Multi-Agent World, LNAI 830*, eds., C. Castelfranchi and E. Werner, pp. 94–114. Springer-Verlag: Heidelberg, Germany, (1994).

[14] T. J. Norman and D. Long, 'Goal creation in motivated agents', in *Intelligent Agents: Theories, Architectures, and Languages, LNAI 890*, eds., M. Wooldridge and N. R. Jennings, pp. 277–290. Springer-Verlag: Heidelberg, Germany, (1995).

[15] C. R. Perrault, 'An application of default logic to speech act theory', in *Intentions in Communication*, eds., P. R. Cohen, J. Morgan, and M. E. Pollack, 161–186, MIT Press, (1990).

[16] J. R. Searle, *Speech Acts*, Cambridge University Press, 1969.

[17] R. G. Smith, 'The contract net protocol: High-level communication and control in a distributed problem solver', *IEEE Transactions on Computers*, **29**(12), 1104–1113, (1980).

[18] J. M. Spivey, *The Z Notation*, Prentice Hall, Hemel Hempstead, 2nd edn., 1992.