Social Intelligence as Norm Adaptation[1]

Magnus Boman & Harko Verhagen

The DECIDE Research Group
Department of Computer and Systems SciencesStockholm University and the Royal Institute of Technology
Electrum 230, SE-164 40 Kista, SWEDEN
Phone: +46 8 16 1678
Fax: +46 8 703 9025
e-mail: {mab,verhagen}@dsv.su.se
URL: www.dsv.su.se/DECIDE[2]

Key words: Learning, Machine guess, Norm, Constraint, Pronouncer, Social space.

Abstract. Machine learning is the core of artificial intelligence. Letting machine learning be driven not only by individual goals, but by goals consistent with those of a coalition of agents, is very difficult. Successful inductive rules for agent behaviour are typically based on machine guesses, the quality of which are measured in terms of precise real numbers representing utility. Random guesses are of low quality, guesses based on heuristics are of the same quality as the heuristics themselves, and guesses based on entire theories are of the quality of the theories together with the quality of the rules for action and their relation to the theory. Such underlying theories have in machine learning thus far not been theories for social action, but theories for individual action. Analogously, the effects of actions on other agents have been studied mainly as feedback to the agent at hand, ignoring the modelling of the utility assessments of those other agents. We propose that intelligent agent action be studied with respect to a social space. The latter consists of a number of agents, their assessments, as well as their sets of norms. Norms are here treated technically, as constraints on individual action. The learning of new norms, and the strife of each agent to act in keeping with the norms of the coalitions of which it is a member constitutes social intelligence.

1. Introduction

1.1 Scope

For agents in a multi-agent system (MAS) to achieve social intelligence is a continuous process which calls for social rationality. To in turn achieve social rationality calls for individually rational action patterns to be constrained also by social obligations. We will focus on the adaptation of norms, as we can find little room here for the evolution of group norms from individual norms. Thus, we study the relation between micro and macro levels of constraints on behavior only in one direction. Moreover, we will in this paper view norms as basically being constraints at the level of actions, and ignore the role of norms in the creation and selection of goals. This makes the agents norm-regulated rather than value autonomous [36].

1.2 Background

When an intelligent agent in a MAS has to decide on what action to take, it might ask for advice. The base case is the agent asking itself what to do next. Almost all AI research, as well as most agent research, deals only with the base case. Many of the classical AI problems, such as the frame problem and the knowledge representation problem appear immediately, and must be addressed. The even more difficult case is when the precarious agent asks someone (or something) else. This case can in turn be analysed by considering two sub-cases. Firstly, the agent may ask other agents in its MAS. This situation can be reduced to the base case, since an agent can see the closure of each agent from which it can receive information as mere extensions of its own knowledge base. Resulting inconsistencies and paraconsistencies must be resolved inductively [4], but any such conflicting information could equivalently be represented locally (cf. [3]). In other words, to form coalitions with agents having opinions in conflict with your own, can be just as irrational as being inconsistent.

---

[1] This work was in part sponsored by NUTEK through the project *Agent-Based Systems: Methods and Algorithms*, part of the PROMODIS program.
[2] From this site, most of the here referenced DECIDE work can be downloaded.

Second, the agent may consult an entity outside the MAS that might not be an agent at all. This entity may come in different guises, e.g., a blackboard, a decision module, or an oracle. Rather than analysing different appearances of the entity, we will base our analysis on the assumptions made about its access to data, and the quality (and to some extent the form) of that data. Our chief motivation is that each of the guises just mentioned have too many variations to allow for them to be studied in precise terms: a blackboard, for instance, does not entail the same agent architecture or model to all researchers that claim to use them. This second case is not reducible to the base case since the entity might at times be inaccessible to the querying agent, and the entity data indeed accessible to the querying agent is usually incomprehensible to the agent.

The availability of data runs from full to zero. In the former case, if each agent represents all its known or believed information in a knowledge base, the entity has access to a database containing the union of all such knowledge bases, with each entry typed to the agent in whose knowledge base the entry appeared. From a syntactical viewpoint, any inter-agent inconsistency is a paraconsistency [30]. In the latter case, one must first define zero data availability. In the strictest possible sense, it means the entity accepts no input, since each input consists of data. Hence, it is solipsistic (in the sense of [40]). Recalling that its sole purpose was to give advice, it is also useless. In a slightly less strict variant, zero data availability means that input may be given to the entity, but that this input reveals no information neither about, nor relevant to, any of the agents in the MAS. In this case, the entity may be used as a random procedure, i.e. it can give advice of a quality equal to casting the dice. In most practical decision situations, data availability is naturally neither full, nor zero.

The quality of data basically runs from precise and known to imprecise and believed. The reason for saying "basically" is that in some circumstances, the quality of imprecise data is equal to that of precise data.[3] Decisions with known precise information about utilities of all consequences (outcomes, world states) of all possible actions is the ideal case, modelling deterministic behaviour. The slightly less ideal situation is where data is imprecise, e.g., expressed in terms of intervals. Things get even more complicated if data does not represent known consequences, but merely believed ones. If probabilities are assumed to be equally distributed within each alternative, and utilities can be stated precisely, we have a game. If the agent can summarise what it knows about the world into complete world states, and it is further able to describe transitions between world states in such a way that the probability of the next state is a function of the current state and action, then the world can be modelled as a Markov decision process (see, e.g., [29]). The first decision-theoretic treatment of probability as degrees of belief is due to Ramsey [31], which was followed by a number of alternative and/or rival Bayesian theories. The pivotal Bayesian assumption reads: "A decision maker's beliefs about the states of the world in a given situation can be represented by a unique probability measure defined over the states" ([19], p.4). The less one knows about the probabilities involved, given that one has at least some opinion, the more important the evaluation procedures become. Combined with less than full data availability, even simple evaluation rules, e.g., the principle of maximising the expected utility (PMEU) are hard to implement efficiently (cf., e.g., [10]).

Even an entity with full data availability of precise and known information is not omniscient: there is always an opponent, usually nature [26]. The decision situation faced is traditionally (and somewhat unfortunately) called a *decision under risk* [19]. Given that the agent has access to some form of decision tree, i.e. that a relevant subset of the set of consequences can be handled by the agent, but the agent has no information about the probability of any consequence actually occurring, the decision situation faced is said to be a *decision under uncertainty*. The typical case for artificial as well as human decision makers is that they have at least some (but not complete and definite) opinions about the relevant probabilities.

---

[3] For instance, when solving a system of linear equations representing constraints on agent behaviour in order to find out what an agent should do next, imprecise data generally speaking yields a solution set of a cardinality bounded only by the number of variables in the system, while precise data yields a unique solution. A well-studied example is situations in games giving rise to multiple equilibria: It is the assumptions about the quality of data available to the players that determine whether the game has a unique equilibrium point or not.

1.3 Disposition

While giving the special cases some consideration, mainly in order to fix our terminology (in this section and the next), we will give most attention to realistic cases. In section 3, we report on our experimental findings and current hypotheses in three different domains. We study entities providing advice to agents by calculating (and subsequently suggesting) a rational agent action, given input that reveals the identity of the querying agent. This input might be incomplete, incorrect, and imprecise. The artificial decision maker will face neither a decision under risk, nor a decision under uncertainty. Related research is described in section 4, while the final section offers conclusions and indicates future research.

2 Terminology

2.1 Naming the Entity

Let us first name the entity giving advice *pronouncer*. This is a new and invented term, so it should be motivated. It suggests an extrinsic entity, and also that the advice given is formal and authoritative, giving the entity a normative status. Thus, it should be used with care, but fits our purposes perfectly. There are a number of less appealing alternatives:

• *Oracle*. Suggests something extrinsic. Has well-defined (different) meaning in complexity theory. Implies high quality of the advice given, if not omniscience.

• *Decision module*. The word "module" suggests internality, i.e. that we are studying one module among others, intrinsic to an intelligent agent. Used in connection with planning, but also for bases of heuristics, and for software providing normative advice to an inquirer.

• *Decider*. Nominative form of "decide". Also short for "decision procedure" in recursion theory. The name of at least one commercial product in the area of risk analysis.

• *Decision machine*. Term in bio-computing. Used in [38] to mean procedures for recommending "patterns of activity". (Collections of decision machines were named *decision factories*.)

2.2 Situating the Entity

There are two possibilities for situating the entity. One is to define a decision module, local to each agent. Just as each agent might have its own list of goals, such a decision module is treated as a customised tool for decision support. Hence, the entity is not merely copied into each agent, but is adapted to the agent to which it belongs from the outset, and increasingly so during its lifespan. The alternative is to have a pronouncer that querying agents call upon repeatedly. The entity is then a resource to be shared among the agents. It will amount to a function, the input of which will have to carry all information about the decision situation, and the output of which will be a recommended action. This pronouncer would be centralised in much the same way as a facilitator in a federated architecture [20].

We choose this latter option, in spite of the complexity of the input to the pronouncer. If our sole concern was individually rational agents, and we also relied only on PMEU, the input could be a decision tree, weighted with probabilities and utilities. The pronouncer would then amount to a calculator recommending (one of) the action(s) with the highest expected value. However, we are investigating socially intelligent agents, and must therefore add group constraints, or use similar means to qualifying individually rational behaviour to achieve social intelligence [6]. This cannot be achieved by merely modifying the weights in the decision tree [18]. Instead, such constraints are part of a local information base, with respect to which each evaluation is carried out by the pronouncer. The necessity of such local bases was previously realised in the context of risk constraints [17]: Not all risk attitudes can be modelled using decision trees.

If we were to vary the evaluation rules in the pronouncer itself, e.g., to experiment with using different extensions of PMEU, it would make sense to have customised entities for normative advice, i.e. to use decision modules instead. Such pluralism with respect to decision support is easy to give arguments for in the case of individually rational agents, but it is perhaps less natural to think that individual utility maximisers are to adhere to the same norms even though their rules for evaluation differ. In any given domain, it is easier to fix a pronouncer and then vary the individual beliefs, preferences, and relevant norms, represented in local information bases. En passant, this is in keeping with Hindess' *styles of reasoning* concept [21].

Naturally, one can imagine simple MASs in which each agent has the same responsibility towards a group. Even in such systems non-trivial problems arise, and there it would suffice to store norms globally, as part of the pronouncer. The realistic and most general case, however, is where each agent has unique obligations towards each and every one of the other agents. For instance, a MAS might consist of 200 agents in which a particular agent has obligations towards the entire population (including itself), but also towards two overlapping strict subsets of, say, 20 and 25 agents that constitute coalitions. These coalitions might be dynamically construed, something which will affect the nature of obligations heavily over time.

We end this section by giving just a flavour of formalisation of a pronouncer. Most of the machinery can be copied directly from [18] and some relevant provisos and assumptions are detailed in [6]. We focus here instead on intuition. First, a *pronouncer* is a function that takes as input a decision tree, an information frame, and a set of norms. An *information frame* is a structure $\langle\{C_1,\ldots,C_m\},P,V\rangle$, where each C_i is a finite set of consequences, P is a finite list of linear constraints in the probability variables, and V is a finite list of linear constraints in the value variables. A *set of norms* N_j consists of constraints typed to agent j, where $j \in \langle a,\ldots,x\rangle$, a vector of agent names in the MAS. The range of a pronouncer is the set of names of leaves in the decision tree input. A *social space* is the union between a finite set of information frames (each typed to an agent in the MAS) and a likewise finite set of norm sets (typed to the same agents). A norm set might be empty.

2.3 Socially Intelligent Artificial Decision Makers

In [6], a general model for artificial decision making constrained by norms was presented. Agents adhere to norms via local adaptation of behaviour or via groups exercising their right to disqualify action options. The adaptation of behaviour consists of an internalisation of group norms, or more precisely a synchronisation of the individual norms to those of the group. This learning of norms is the core of socially intelligent behaviour. The assessments in the information frames gradually evolve, in order for the agent to act in accordance with the norms of its group. The group norms, or social constraints, are not merely the union of the local information frames of its members, but rather develop interactively, as do the local information frames.

Norms can be augmented by rules for the introduction and exclusion of members. Exclusion then is the ultimate punishment for pursuing actions considered intolerable by the group. In order to become a member of a group, an agent has to be able to adhere to the group norms. This means that its information frame should be compatible to the group's action constraints, e.g., consistent with the most important group constraints. How groups may pro-actively invite possible new members is another matter, which will not be analysed in the present work.

3 Evaluation

The ideas presented in the previous sections are currently being evaluated experimentally in several domains, the main hypothesis being that artificial decision makers can benefit from pronouncers, even in dynamic real-time environments.

3.1 Agents for Securing Electricity Contracts

The first domain is electronic power markets, with agents acting as assistants for securing contracts for the delivery of electricity. An important assumption is that such contracts will be very short-term in the future, in part as a result of the de-regulation of utilities in many countries. In order to keep costs down for the consumer, time must then repeatedly be spent on searching for good current offers, an activity which is costly in itself. With the help of autonomous agents acting on electronic power markets this cost can be diminished. Such markets today have a limited number of buyers and sellers, chiefly due to the high fees for participation. This makes them interesting social spaces for artificial agents. By contrast, open systems such as the Internet are so large and dynamic that social factors like trust, rumours, and reputation are almost irrelevant. We are currently investigating the importance of norms in these social spaces.

The agents inhabiting such a social space is typically rather primitive: their basic functionality lies in their bidding algorithms and their logical and/or physical mobility between different markets. However, it is not hard to envisage that an artificial decision maker could enhance its performance, measurable for example in the efficiency of its bidding, by augmenting the bidding algorithm with pronouncer calls and norm adaptation. An example of the former would be an agent choosing whether or not to enter a certain spot market. Factors affecting the weights in the decision tree constituting the pronouncer input would include the time left to secure a contract, the number of actors currently on the market, the other markets currently available, and perhaps statistical information about this particular spot market. An example of the latter would be for the agent to learn not to try to outbid competitors belonging to the same company (or government agency, or sports club...). Instead, the agent should appreciate the gain in utility of forming coalitions with such agents. Understanding, in the weakest possible sense, the dynamics of such coalitions is an integral part of learning on the agent's part. The individual information frames consist of believed imprecise information, and the agents' data availability is relatively high.

3.2 Intelligent Buildings

A multi-agent system for energy saving and enhanced customer value in intelligent office buildings has been implemented as a simulation system, and some physical installations (e.g., temperature and light sensors) have also been completed at a test site ([7], [12]). Customer value is measured in terms of the extent of which the preferences of each person working in the building can be met. For instance, a person might want the light and computer in her room to be turned on as she enters the building. An interesting problem is the amalgamation of customer preferences, e.g., the preferences with respect to lighting in a conference room with eight people at the conference table. The negotiation procedures run in the background before such a meeting involves eight so-called personal comfort agents, but also a conference room agent, and several device (e.g., radiator, lamp) agents. The goal of the room agent is to consume as little energy as possible, and moreover the personal comfort agents typically disagree about the lighting and heat values (lux and degrees Celsius) in the room. Thus, for every social gathering such as a meeting, there is at least one other social gathering of agents, invisible to the human agents. The study of norms in this social space is worthy of study, not least because it has measurable effects on the life of people in the building, and so can be studied empirically once the test site is fully operational. The individual information frames again consist of believed imprecise information, and the agents' data availability is moderate. Data availability can be drastically improved, however, by synchronising electronic calendars. Since this imposes a burden on the people in the building, it decreases customer value, and so should be used with care (e.g., by restricting the shared electronic calendar to the booking of rooms only).

3.3 RoboCup

A team named UBU recently competed in RoboCup'98 in Paris [1]. This first version of UBU had no pronouncer calls. The main part of development has thus far been stable basic functionality. A second version of UBU will compete at PRICAI'98 in November. This team will have pronouncer calls: idle agents will ask a pronouncer for advice in order to put themselves to good team use. A third version of UBU will appear in the next World Cup in Stockholm (at IJCAI'99), and in this team pronouncer calls will hopefully be evaluated relative to subjective sets of norms.

In RoboCup, each agent has very limited vision and communication capabilities. It can only see certain portions of the field, and it can only hear messages uttered within a certain distance. Since there are 22 players and one ball, none of which are still for more than a few seconds, most of the data available is old. Data availability is thus very low. The information coded in the information frames is highly imprecise and only very weakly believed. It is therefore unrealistic to expect a pronouncer to improve the performance of a team more than marginally.

3.4 Pronouncers

The first pronouncer implemented was DELTA [9], originally developed for iterative assistance to human decision makers, but gradually re-implemented for MAS applications ([15], [16]), including socially intelligent artificial decision makers [5]. However, since DELTA was never optimized for real-time use, it turned out to have too slow response time for use in RoboCup. Recently, a new pronouncer was developed, based on a refined version of the original theory (cf. [34], [35]), more appropriate for real-time use. This is currently being implemented, for test use in all of the above domains. We are also investigating a vast range of commercial decision analysis tools, in order to determine their viability as real-time pronouncers.

Even though all three domains described above are dynamic real-time domains, and hence very difficult to handle, the time bound on reasoning (and on pronouncer calls, in particular) varies greatly. During a RoboCup game, a pronouncer call must be made within 100 milliseconds. Any pronouncer output not processed after a few seconds is usually useless. These extreme conditions entail that any adaptation, perhaps any involvement of norms in the reasoning process, must be done between games. We are currently investigating whether there is time for using norms at least as action filters in games, i.e. to let the pronouncer disqualify a certain action on the grounds that it is too risky. This is definitely a possibility in the other two domains, where time is less critical. In those domains, the natural way of implementing pronouncer calls is as part of an anytime algorithm representing the reasoning cycle (cf. [5]).

4 Related Research

First steps in the direction of defining rational behaviour as individually rational behaviour extended by collective awareness, i.e. the micro-macro link, were taken within the MAS area by Castelfranchi and his group in a long string of articles (and a book [8]).

The beginnings of a formalisation of norms in agent action was through a logic approach [41]. The most popular logic is deontic logic, usually appearing as revised versions of Meyer's reduction of deontic logic to dynamic logic [27]. Only actions, and not formulas (representing assertions or assessments) can be obliged in Meyer's logic, and this has led to various extensions. These include (roughly in order of sophistication) a coupling to speech acts using illocutionary logic [13], a "logic for action and norms" [14], a first order action logic [24] (currently limited to pairwise obligations, i.e. broadcasting commitments is not possible), and even first steps towards a logic-based social agent development language [1]. Interestingly enough, the hard work put down to augment Meyer's logic is debatable in view of our somewhat controversial description of plans as normative advice for essentially reactive agents. Our view is in line with Meyer's original position that world states need not be explicitly modelled: Modelling actions is sufficient.

Jennings and Campos refine Newell's principle of rationality [28] into a "principle of social rationality" that says that "if a member of a responsible society can perform an action whose joint benefit is greater than its joint loss, then it may select that action" [23]. Since it is modelled on Newell's principle, it inherits some of its weaknesses, e.g., it does not treat the case of several (or no) suitable actions. Moreover, the central concept of joint benefit is defined for the coarsest possible value scale, viz. one of loss and benefit only. Close to our own paper is also a paper by Kalenka and Jennings [25] in which benefits are divided into individual, social, and joint benefits. As in [23], evaluations are based on utility functions representing agent preferences. These two papers are important first steps towards a social level. The natural second step is to introduce beliefs (i.e, probabilities), vagueness (i.e. imprecise utility assessments), and a more expressive language, including risk profiles, security levels, individual constraints towards groups, etc.

That agent sensors are good enough to perceive every change in the world state after an action has been taken, thus removing any uncertainty about the consequences of the action, is an unrealistic assumption. The sensor problems begin already in semi-realistic simulated environments, such as the RoboCup simulation league competition, as we have ourselves learned. In [29], Parr and Russell manage to keep the Markov property in their agents, even though they relax the assumption about full observability. It remains to show that their solution is efficient enough for real-time use, however.

One might have a similar concern for the anytime algorithm of Horsch and Poole for computing policies for decision problems represented as multi-stage influence diagrams [22]. That said, the authors do report on valuable (typically non-optimal) decision policies being found also for large problems, and treatments which take into account the cost of computation are otherwise rare (cf. also [5]).

5 Conclusions and Further Research

We have analysed artificial decision making in social spaces. In particular, we have recommended the use of pronouncers, operating relative to agent-specific sets of norms and domain-specific assessments. Norms have been treated here as constraints on individual agent action. Our analysis hopefully takes a second step towards a social level in agent programming, augmenting or replacing Newell's celebrated knowledge level, the first step having been taken by other authors (as was explained in the previous section).

This is work in progress. In section 1, we narrowed down the scope of this paper to be considerably smaller than our research interests. Naturally, group decisions, negotiations, coalition formation, and several other issues not treated here (see [37]) are relevant to our analysis. As was explained in section 3, evaluations are on-going, and most of the experimental results have not yet been obtained. We have conducted various simulation experiments, only some of which have yet been analysed (e.g., [12]). Aside from our future reports on the UBU RoboCup team, we hope to present a coherent view of our experimental findings and our more philosophical analysis, as exemplified by this paper.

To us, the most burning question is learning. For improving the UBU team, as well as our analyses, we will next study two forms of machine learning: reinforcement learning [33] and anticipatory systems [32]. Both have been studied with similar goals in mind (see [29] and [11], respectively), but with a focus on planning. We would like to de-emphasize macro-level (abstract) planning for our agents, and to a large extent rely on the pronouncer instead. The hypothesis is that macro-level plans are often superfluous (cf. the treatment of goals in [39]), and could be replaced by risk, security, and group constraints.

References

[1] H. Åberg, Å. Åhman, J. Andreasen, M. Boman, M. Danielson, C.-G. Jansson, J. Kummeneje, H. Verhagen & J. Walter: "UBU: Utility-Based Uncertainty Handling in Synthetic Soccer", *Proc RoboCup98*, in press.

[2] M. Barbuceanu, T. Gray & S. Mankowski: "How to Make Agents Fulfil Their Obligations", *Proc PAAM98*: 255-276, Nwana & Ndumu (eds.), PAC, 1998.

[3] M. Boman: A Logical Specification of Federated Information Systems, Ph.D. thesis, DSV, SU, Stockholm, 1993.

[4] M. Boman & L. Ekenberg: "Eliminating Paraconsistencies in 4-Valued Cooperative Deductive Multidatabase Systems with Classical Negation", *Proc CKBS94*: 161-176, Deen (ed.), Keele Univ Press, 1994.

[5] M. Boman: "Norms as Constraints on Real-Time Autonomous Agent Action", *Multi-Agent Rationality (Proc MAAMAW97)*: 36-44, Boman & Van de Velde (eds.), LNAI 1237, Springer-Verlag, 1997.

[6] M. Boman: "Norms in Artificial Decision Making", *AI and Law J*, in press.

[7] M. Boman, P. Davidsson, N. Skarmeas, K. Clark & R. Gustavsson: "Energy Saving and Added Customer Value in Intelligent Buildings", *Proc PAAM98*: 505-515, Nwana & Ndumu (eds.), PAC, 1998.

[8] R. Conte & C. Castelfranchi: Cognitive and Social Action, UCL Press, 1995.

[9] M. Danielson: Computational Decision Analysis, Ph.D. thesis, DSV, KTH, Stockholm, 1997.

[10] M. Danielson & L. Ekenberg: "A Framework for Analysing Decisions Under Risk", *European Journal of Operations Research* **104**: 474-484, 1998.

[11] P. Davidsson, E. Astor & B. Ekdahl: "A Framework for Autonomous Agents Based on the Concept of Anticipatory Systems", *Proc Cybernetics and Systems '94*: 1427-1434, World Scientific, 1994.

[12] P. Davidsson & M. Boman: "Energy Saving and Value Added Services—Controlling Intelligent Buildings Using a Multi-Agent Systems Approach", *Proc DA/DSM98*, forthcoming.

[13] F. Dignum & H. Weigand: "Modelling Communication Between Cooperative Systems", *Proc CAiSE95*: 140-153, Iivari, Lyytinen & Rossi (eds.), LNCS 932, Springer-Verlag, 1995.

[14] F. Dignum, H. Weigand & E. Verharen: “Meeting the Deadline: On the Formal Specification of Temporal Deontic Constraints”, *Proc ISMIS96*: 243-252, Ras & Michalewicz (eds.), LNAI 1079, Springer-Verlag, 1996.

[15] L. Ekenberg, M. Danielson & M. Boman: “From Local Assessments to Global Rationality”, *Intl J of Intelligent Cooperative Information Systems* **5**(2&3): 315-331, 1996.

[16] L. Ekenberg, M. Danielson & M. Boman: “Imposing Security Constraints on Agent-Based Decision Support”, *Decision Support Systems Intl Journal* **20**(1): 3-15, 1997.

[17] L. Ekenberg, M. Boman & J. Linnerooth-Bayer: “Catastrophic Risk Evaluation”, IIASA Report No. IR-97-045. Intl Institute for Applied Systems Analysis, Laxenburg, Austria, 1997.

[18] L. Ekenberg, M. Boman & J. Linnerooth-Bayer: “General Risk Constraints”, submitted.

[19] P. Gärdenfors & N.-E. Sahlin: “Bayesian Decision Theory - Foundations and Problems”, in *Decision, Probability, and Utility*: 1-15, Gärdenfors & Sahlin (eds.) Cambridge Univ Press, 1988.

[20] M. R. Genesereth & S. Ketchpel: “Software Agents”, *Communications of the ACM* **37**(7): 48-53, 1994.

[21] B. Hindess: Choice, Rationality and Social Theory, Unwin Hyman, 1988.

[22] M. C. Horsch & D. Poole: “An Anytime Algorithm for Decision Making under Uncertainty”, *Proc UAI98*, forthcoming, July 1998.

[23] N. R. Jennings & J. R. Campos: “Towards a Social Level Characterisation of Socially Responsible Agents”, *IEE Proc on Software Engineering* **144**(1): 11-25, 1997.

[24] P. Johannesson & P. Wohed: “Modelling Agent Communication in a First Order Logic”, submitted.

[25] S. Kalenka & N. R. Jennings: “Socially Responsible Decision Making by Autonomous Agents”, *Proc 5th Intl Colloq on Cognitive Science*, 1997.

[26] R. D. Luce & H. Raiffa: Games and Decisions, John Wiley & Sons, 1957.

[27] J. J. C. Meyer: “A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic”, *Notre Dame J of Formal Logic* **29**(1): 109-136, 1988.

[28] A. Newell: “The Knowledge Level”, *Artificial Intelligence* **18**: 87-127, 1982.

[29] R. Parr & S. Russell: “Approximating Optimal Policies for Partially Observable Stochastic Domains”, *Proc IJCAI95*: 1088-1094, Mellish (ed.), 1995.

[30] G. Priest, R. Routley & J. Norman (eds.): Paraconsistent Logic—Essays on the Inconsistent, Philosophia Verlag, 1987.

[31] F. P. Ramsey: “Truth and Probability”, in *The Foundations of Mathematics and other Logical Essays*: 156-198, Braithwaite (ed.), Routledge & Kegan Paul, 1931.

[32] R. Rosen: Anticipatory Systems—Philosophical, Mathematical, and Methodological Foundations, Pergamon Press, 1985.

[33] R. S. Sutton & A. G. Barto: Reinforcement Learning, MIT Press/Bradford Books, 1998.

[34] J. Thorbiörnson & L. Ekenberg: “Congruency of Beliefs”, forthcoming notes from *AAAI98 Spring Symp*.

[35] J. Thorbiörnson & L. Ekenberg: “Globalisation of Belief Distributions”, *Proc FLAIRS98*, forthcoming.

[36] H. J. E. Verhagen & R. A. Smit: “Modelling Social Agents in a Multiagent World”, *Position Papers at MAAMAW96*, Technical Report 96-1, AI Lab, Vrije Univ Brussel, 1996.

[37] H. J. E. Verhagen: “ACTS in Action”, *Proc MABS98*, forthcoming, July 1998.

[38] M. P. Wellman: “Rationality in Decision Machines”, unpublished notes from *AAAI95 Fall Symp on Rational Agency*. Available from http://ai.eecs.umich.edu/people/wellman/decision-machine.html.

[39] M. P. Wellman: “Preferential Semantics for Goals”, *Proc AAAI91*: 698-703, 1991.

[40] E. Werner: “Logical Foundations of Distributed Artificial Intelligence”, in *Foundations of Distributed Artificial Intelligence*: 57-117, O’ Hare & Jennings (eds.), John Wiley & Sons, 1996.

[41] G. H. von Wright: An Essay in Modal Logic, North-Holland, 1951.