

Can Computers be Social?

Bertil Ekdahl

Lund Institute of Technology
School of Engineering
Box 882, SE-251 08 Helsingborg, Sweden
Phone: +4642176346
Fax: +4642176337
E-mail: Bertil.Ekdahl@cs.lth.se

Abstract

Agent based computing is considered a promising and exciting area of Artificial Intelligence in general and particularly of software engineering where it is thought of as a new tool for developing software. Especially the idea that agents can attain socially responsible behavior is of main concern in current research in the multiagent area. This expectation has its origin in the need for agents to interact with one another in a cooperating manner. Such interplay between several agents can be seen as a combinatorial play where the rules are fixed and the actors are supposed to closely analyze the play in order to behave rational. This kind of rationality has successfully being mathematically described. It is when the social behavior is extended beyond rational behavior that mere mathematical analysis is no longer sufficient. For humans, as social agents, the sociality is not inherited (not programmed) but has evolved together with both a language and a cognitive apparatus. Language is decisive for being social and without language the necessary attributes for developing sociality cannot exist. These attributes are not outside the language in which they are communicated but part of it as a wholeness and cannot be separately analyzed. Since language is a holistic entity they cannot even be defined in the language in which we talk about them.

In this paper it is shown that software agents, that is algorithms, do not have a language in the sense that meaning can be conveyed and consequently they lack all the necessary properties to be made social. The attempts to postulate mental properties to programs are deeply flawed and based on the lack of true understanding of language and especially the relation between formal system and its semantics.

Keywords: Agent, multiagent systems, social agents, social behavior, beliefs.

1 Introduction

In the agent community, agent based computing is considered a new and exciting area of Artificial Intelligence in general and of particular interest for software engineering where it is thought of as a new promising tool for developing software. It is further believed that agent technology has the ability to improve many areas in computer science as modeling, designing, and in implementing complex systems (Jennings, 1999). More than that, its influence is supposed to be so great that it is even acclaimed to be a “new revolution in software”. (Ovum, 1994, quoted in Kalenka and Jennings, 1997).

Thus, there are a lot of expectations about the benefit of agents but these expectations are deeply founded on the belief that computers can be ascribed the following qualities that, by Kalenka and Jennings (1997), are pointed out as necessary for agenthood:

- (i) autonomy
- (ii) responsiveness
- (iii) proactiveness
- (iv) social ability

These properties, which are typically human, show that there is a strong tendency to compare the behavior of interacting computer programs with human group behavior. This is a consequence of the tacit assumption that all relevant aspects of human agenthood are algorithmically describable. This conception is based on the idea that the Turing machine model is a model of more than mere calculation, or at least that other human capabilities could be shown to be on par with the calculating capability. So far this belief has no true support. As a matter of fact, not even the concept of software agent is well defined (Ekdahl, 2000A, 2001) and no serious attempts to ask for describability of the concept have been made.

In the history of logic there are several examples of intuitive concepts that have been investigated in order to be made precise. One well-know example is Gödel’s analysis of the concept of provability, which, despite being a basic tool in mathematics, was an unclear concept. Another, may be even more known example, is Turing’s analysis of our intuitive concept of mechanical procedure, which is remarkable. He was able to give it a precise description and was also able to convincingly explain why it agreed with our intuition. Turing’s analysis of computability is an especially good illustration of a thoroughly analysis of an intuitive concept.

Despite the vague characterization of agent, it is never questioned in the agent community whether properties as for example autonomy and social ability really are possible to describe in computable terms. No analysis, comparable to Turing’s, has been undertaken. It is quite simply stated that software agents are capable of having the

attributes enumerated above. It is just taken for granted that human agenthood is representable in a computer program.

In the list above of the assumed necessary conditions (properties) for agenthood, the most crucial is perhaps that of social ability. It also seems to be of key concern in agent research:

Attaining socially responsible behavior is the motivation behind a significant proportion of current research in the multi-agent system arena, indeed Gasser [36] and Hewitt [9] identify it as perhaps the key problem in this area. (Jennings and Campos, 1997)¹

Social ability is a concept that man uses in order to describe the specific human way of living together in groups. Even if we attribute social ability for example to ants, it is still the human conception of being social that is the model. We transfer human concepts to other species because it simplifies the characterization when discussing their behavior. Ants are social because we can trace things in their behavior that, when found in humans, are called social. However, we cannot say exactly which animals are social and which are not because the term *social* is not precisely given. It is characteristic for a natural language that many words, if not most, do not stand for clearly defined concepts. Being social is one of those concepts and there is no attempt in the social sciences to give it a precise definition and not even a need to do it. This is because communication and language play a constitutive role of sociological system as an object. However, our conception of *sociology* is clear.

Therefore, it is astonishing that it is believed in the agent community that social ability is a concept that can be more or less mathematically stated, that is, it is at least partially possible to describe social ability in terms of *necessary* and *sufficient* conditions.

In this paper I will analyze the conditions for computers to be social. It is argued that in order to be social, in the sense of human beings, a communicative language is required with which semantics can be transferred between those who share the language. Computer languages lack this quality.

2 The Background in Game Theory

The origin of the belief that software agents can be made social can be traced back to the at least partially, successful computer simulations of economic behavior. The idea in economic analysis is that the actors behave rationally and this has also been significant for the rational behavior approach in agent technology.

¹ The references are to Gasser (1991) and Hewitt (1991)

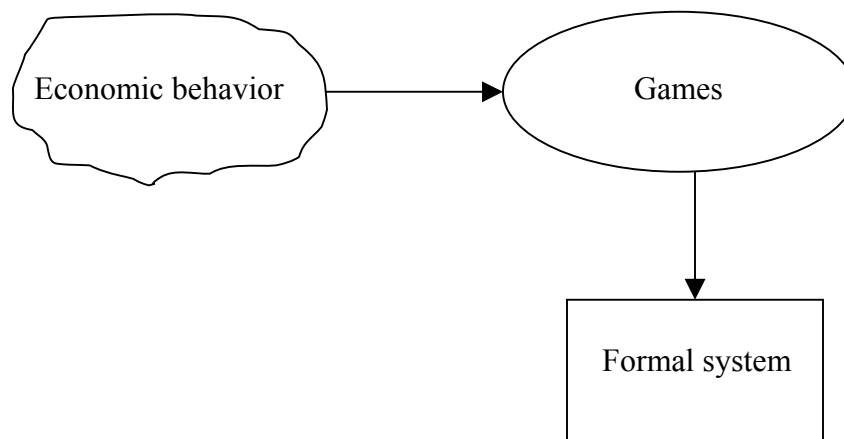


Fig. 2.1: The relation between economics and games

The forerunners in this thinking are von Neumann and Morgenstern who were the first to apply a game theoretical approach to economic behavior. Their work on the theory of games has profoundly permeated economic analysis but has also been applied to fields beyond economics. The game idea has been adopted in areas as diverse as political science and evolutionary biology. To give the background of social agents I will briefly describe von Neumann and Morgenstern's analysis of social economy. To this, I closely follow the work of Leonard (1995). The quotations are from von Neumann and Morgenstern (1944), referred to in the text as *Theory of games*.

von Neumann and Morgenstern regarded economic behavior as social behavior. They saw economic behavior as a game between several people and realized that economic behavior could be analyzed in terms of games. Games, like parlor games and poker, were thought of as rational plays that could be analyzed in strategic terms. The outcome was dependent on the players' rational choice. The idea then was that economic behavior was a game between rational players. From this assumption, the further step was clear: make assumptions about economic behavior that could be described in game theoretical terms and thereby analyzable in mathematical terms. As figure 2.1 suggests, economic behavior was translated to game theory, which became a model of the formal system in which economic was analyzed.

In the construction of n -persons game von Neumann and Morgenstern postulated a dominance relation in order to characterize solutions of the game. Then they connected this mathematical concept of solution to social phenomena. According to von Neumann and Morgenstern, a solution may be correlated with a "standard of behavior", i.e., the particular set of rules, customs or institutions governing social organization at particular

time. In order to explain the analogy between games and social organizations, they advised the readers to temporarily forget the analogy with games and think entirely in terms of social organizations:

Let the physical basis of a social economy be given, — or to take a broader view of the matter, of a society. According to all tradition and experience human beings have a characteristic way of adjusting themselves to such a background. This consists of not setting up one rigid system of apportionment, i.e. of imputation, but rather a variety of alternatives, which will probably express some general principle but nevertheless differ among themselves in many particular respects. This system of imputations describes the “established order of society” or “accepted standard of behavior”. (p.41)

The *theory of games* was intended to constitute a radical rupture with the prevailing opinion. For example, von Neumann completely rejected the earlier approach to economic behavior with its underlying physical metaphor of classical mechanics and the associated mathematics based on the differential calculus. von Neumann’s opinion was that *social phenomena clearly require theoretical categories and mathematical methods of a different kind*:

Our static analysis alone necessitated the creation of a conceptual and formal mechanism which is very different from anything used, for instance, in mathematical physics. Thus the conventional view of a solution as a uniquely defined number or aggregate of numbers was seen to be too narrow for our purpose, in spirit of its success in other fields. (p.45)

In order to understand von Neumann’s axiomatic view some explanation of formalism is necessary. In the beginning of the twentieth century Hilbert created a program of validation and justification of (infinite) classical mathematics. This program became known as formalism. It is the view that mathematics is characterized more by its methods than its subject matter. Its objects are either unspecified or, if specified, are such that their exact nature is irrelevant; it is the form that is relevant. The formalism was a step further from the axiomatic method in order to meet also the infinite. This step was forecasted by Hilbert in 1904, and seriously undertaken by him and his collaborators, especially Bernays, Ackermann and von Neumann since 1920.

It was in this tradition von Neumann was raised and to him it was natural that a rigorous analysis of a concept was necessary in order to give it a mathematical

description. The mathematical definition of the analogy between games and social economic behavior was undertaken in this spirit:

This definition should be viewed primarily in the spirit of the modern axiomatic method. We have even avoided giving names to the mathematical concepts introduced... in order to establish no correlation with any meaning which the verbal associations of names may suggest. In this absolute “purity” these concepts can then be the objects of an exact mathematical investigation ... *The application to intuitively given subjects follows afterwards, when the exact analysis has been completed ...The axiomatic models for intuitive systems are analogous to the mathematical models for (equally intuitive) physical systems.*

von Neumann regarded games as combinatorial plays (von Neumann, 1928). He saw games as a mathematical concept that was completely axiomatizable. This was his Hilbert legacy: asking for an axiomatic way to describe games.

Since von Neumann and Morgenstern, game theory in economics has been developed in new and quite different directions. The work has taken a stronger focus on the exploration of individual rationality than was present in *Theory of Games*. This modern mathematics of social behavior is behind many works on agents but the agent community takes a further step to include mental states that is not in the spirit of Hilbertian mathematics, the model of von Neumann and Morgenstern.

3 Social Agents

In his paper, *Agent Based Computing: Promise and Perils*, Jennings (1999) argues that, “an increasing number of computer systems are being viewed in terms of autonomous agents”. Following Wegner (1998), he considers that “agents are being espoused as a new theoretical model of computation that more closely reflects current computing reality than Turing Machines”. This supposition is however deeply flawed, which has been shown by Ekdahl (1999). Referring to Wooldridge (1997), Jennings states that, “agents are being advocated as the next generation model for engineering complex, distributed systems”.

These are great expectations that seem to change the view of software engineering and also the whole software industry. However, the concept of software agent is very weakly founded and when the agent community argues for the benefit of agent computing its advocates cannot give a satisfactory description that distinguishes agents from non-agents (Petri, 1996). To remedy this shortcoming, there have been several attempts to give the

concept of agent a rigorous definition. However, the best Jennings is referring to is a “definition” of Wooldridge (1997) that has been shown by Ekdahl (2000A, 2001) not fulfilling the most elementary demand for being a definition.

Despite the great vagueness of agent there seems to be a firm conviction that there is an agent view that can be applied when developing software systems.

When adopting an agent-oriented view of the world, it soon becomes apparent that most problems require or involve multiple agents, to represent the decentralised nature of the problem, the multiple loci of control, the multiple perspectives, or the competing interests. Moreover, the agents will need to interact with one another, either to achieve their individual objectives or to manage the dependencies that ensue from being situated in some environment. (Jennings, 1999, p. 1430)

It is in this interaction where the *social* ability is supposed to emerge. Social ability is considered one of the cornerstones on which agents hinge.

To my knowledge, Werner was one of the first to use the idea of *social intentions* for multiagent systems. “The fundamental question we face when dealing with several agents is how to get those agents to achieve a social goal” (Werner, 1988). He used the term *social goal* with which he means a goal not achievable by any one agent. He defines a *social group* as a tuple consisting of a language, a group of agents, a social structure, a distribution of roles on the group and an environment. Here *social role* is the main concept and is defined as an abstract representational state, consisting of the components *information state*, *intentional state*, and *evaluation state*. These components are related to what Werner calls *the state of a conversational participant*. Werner’s social group is defective on the ground that the describability of the social phenomena are not questioned but taken for granted. Social properties are quite simply stated as formally describable without analysis.

Panzarasa and Jennings (2001) have as their aim to “place the study of social influence processes on a more secure and formal footing. To this end, we formalise a model of social influence [...]”. Their approach is to “develop a logic-based architecture for formalising the cognitive agent’s mental state in terms of mental attitudes such as beliefs, goals and intentions.” By *social influence* they mean “the process through which the mere social nature of agents affects, thereby alters, their mental states from what they would otherwise have been, had the agents not engaged in any form of social behavior.” They introduce, what they call, a modal operator *Infl* to formalize the process of social influence in its basic forms. They also use the notions $Bel(a_i, \Phi)(t_i)$ and $Att(a_i, \Phi)(t_i)$ to indicate that agent a_i at time t_i maintains, respectively, a belief that Φ holds and a generic unspecified mental attitude (e.g. a belief, a goal, an intention) towards Φ . They

then use these operators to construct formulas that are claimed to formalize different kinds of *social influence*.

The point of departure is technical but the analysis is completely omitted. It should not come as a surprise that agents are algorithms. An important question should then be whether there is a decision procedure for *belief* and *attitudes*?

Jennings and Campos (1997) argue for a *social level* in accordance with Newell's (1982) *knowledge level*. In their paper they argue that the principle of rationality is insufficient for describing a range of desirable social behavior.

For example, consider the case in which a number of agents *decide to work together*² to search for a lost item [...]. Assuming the item is discovered, one of the agents will have satisfied its original goal. However, it is part of the intuitive notion of working together that the successful agent should inform its fellow searchers that the target has been found, so that they can abandon their search [...]. However, such additional behaviour is not warranted by the principle of rationality, because this informing action does not satisfy one of the finder's goals. In particular for rational agents, the notion of performing actions for the greater good is not permitted: helpful actions not connected to one of the agent's goals contradict the principle of rationality [...].

The principle of rationality is supposed to be a consequence of that the knowledge level is directly applied to social phenomena. In order to overcome the shortcomings of the knowledge level, Jennings and Campos extend it to a *social level* in accordance with the knowledge level but on a level above the rational. They postulate the existence of a social level:

Social level hypothesis: there exists a computer level immediately above the KL³, called the social level, which is concerned with the inherently social aspects of multiple agent systems.

Then, they argue, the social level provides an abstract characterization of those aspects of multi-agent system behavior that are inherently social in nature. Thus it is concerned with representing phenomena such as co-operation, co-ordination, conflicts and competition.

Several things can be said about the proposals. First, in arguing that the principle of rationality is insufficient, Jennings and Campos take as an example "a number of agents

² My emphasis.

³ Knowledge level.

that *decide to work together*". This statement is characteristic in the agent community. It is an example where, algorithms and human behavior are confused. Do really algorithms *decide*, on their own, to work together? The term algorithm is equivalent to a decision procedure, meaning that whatever we put as input to the procedure it will always answer *yes* or *no*. In this sense *all* algorithms are the same: they are exactly decision procedures. In this case the comparison can be done with insects. They are "programmed" to work together; they do not decide it. We may consider the very nature of decision procedure as a description of a (total) recursive function, that is, the procedure gives an answer to every input value.

Secondly, the *social level* (SL) hypothesis is quite incomprehensible. It is an existential claim about computers but Jennings and Campos do not state any reason for assuming it. A hypothesis is a well-founded guess that is supported by observations that give explanations that go beyond existing theories. A hypothesis is a guess about realities in the surroundings that we do not yet understand. As an example we may consider Max Planck's discovery of the quantum of action, which inaugurated a new epoch in the physical science. He had all reasons to state his quantum phenomenon because he had good evidence supporting it (Pais, 1991). So had also Alonzo Church (1935) when he suggested to identify the human cognitive capacity of computability with the recursive functions. All our knowledge of reality is guesses of the same kind as Planck's and Church's. We may never know the reality but may at the most hope for describing it in a consistent way. It is when the theory becomes inconsistent with new observations that we have to change old beliefs. Now, the social level hypothesis is not of the same quality as that of Planck and Church, not even of the same quality as the existential proposition of Zermelo that there is a choice function, leading to the axiom of choice. If we dissect the SL hypothesis, it is a statement about Turing machines saying that in, at least, some programs on a (universal) Turing machine there is a level above the program level that can be regarded as social. Due to the equivalence between Turing machines and formal systems⁴, the statement is equivalent to the statement that formal systems have a social level. We know that all functions, computable on a Turing machine is *representable* in the formal system \mathcal{R} with the following axioms (schemata):

- R1. $\bar{x} + \bar{y} = \bar{z}$, if $x + y = z$
- R2. $\bar{x} \cdot \bar{y} = \bar{z}$, if $x \cdot y = z$
- R3. $-\bar{x} = \bar{y}$, if $x \neq y$
- R4. $\forall v_0 (v_0 \leq \bar{x} \rightarrow v_0 = \bar{0} \vee v_0 = \bar{1} \vee \dots \vee v_0 = \bar{x})$
- R5. $\forall v_0 (v_0 \leq \bar{x} \vee \bar{x} \leq v_0)$

⁴ See paragraph 5 for an explanation.

Every computation is a proof in \mathcal{R} and is also an information reducing process in the sense that every provable sentence says less about the nature of the computable functions, the model of \mathcal{R} , than the axioms from which the conclusion is derived. So, if there is a social level, it does not emerge spontaneously but must be able to be pointed out in the axiom system. Jennings and Campos do not do that. Nor do they give the supporting observations that lead to the hypothesis.

Smit and Verhagen (1995) admit that the intriguing question concerning the emergence of social phenomena as a product of the interaction between autonomous, if dependent, agents is caused confusing answers. They see the discrepancy between the social science and social philosophy on the one hand and the corresponding concepts in the agent community on the other hand, as a big problem and their intention is to bridge over these differences.

We do not pretend to give definitive answers or lead the way to detailed implementations of sophisticated multiagent systems. Instead, we want to share with the audience the embarrassment we have felt and still feel now and again when we look at our own lack of critical thinking and sophistication when we described and named MAS phenomena in the terminology of the social sciences.⁵

As an explanation they continue:

To the defense of MAS researchers and ourselves, we could say that the social sciences lack the formal rigor and level of detail of the computational approach, but this does not license us to start using their concepts without taking in all the problems that have been argued to be attached to their usage.

However, they do not believe that the conceptual problems derive from the *lack of formal specification* or descriptive adequacy, often attributed to the nature of social sciences, but are blamed the complexity of the subject.

The supposition that there is a difference between the formal specification of a concept and its complexity is a misunderstanding that has to be blamed the lack of true understanding of complexity. They regard the complexity of an object as something inherent solely in the object itself and quite independent of our ability to describe it. But how can we argue for such an unspecified complexity? The very nature of complexity is our inability to understand a phenomenon. When we do not understand a thing it means that we cannot explain it, or in other words, we cannot describe it. In fact, the complexity

⁵ MAS: Multi Agent System.

of an object is not anything that in the first place belongs to the object itself but to the language in which we try to describe it. The complexity of a concept starts when we try to describe it and when the language in which we try to make the description does not contain the necessary concepts in order to comprehend the object in question. An object is more complex than another if the first is more difficult to describe in a language than the second. After a time the language may very well be developed so as to also include new ideas that make a thing, that formerly was hard to explain, easier to understand. Then the complexity has decreased. Science is a good example of the fact that complexity decreases according as the physical language has evolved.

The social structure, a result of evolution, has evolved in the way it has and is not complex for those involved in a social group. No human being seems that his or her social ability is complex. The concept of being social is not complex in itself but becomes complex first when we try to describe it since there is no language in which it can be described. Thus, the difficulty, that Smit and Verhagen blame the subject, is in fact exactly equivalent to the lack of formal specification, a connection that they do not appreciate. All measures of complexity can be mirrored in logic and it may be that we will never be able to describe social ability in classical logic.

Despite the lack of formal specification, as Smit and Verhagen point out, a pervading characteristic in many approaches to give agents social properties, is that a mathematical language is used but without the mathematical rigor.

Common to all approaches is that language is not considered central for social ability. Contrary to this opinion, I will show that language is decisive for being social.

4 Language and Social behavior

Despite being a concept of significance in sociology, there is no definition of *social behavior* or of *being social*. Many species are called social because they live in flocks with hierarchies, are able to make coalitions, etc., but there is no clear definition of which animals that are social. However, one necessary condition for being social seems to be the possibility of communicating in order to cooperate since cooperation is the hallmark of social groups. The only reason to work together in a group was from the beginning to gain advantages that was otherwise impossible to achieve, for example, reducing the risk of predation. A group has more eyes to detect stalking predators but in order to get the full benefit of that the individuals have to be able to communicate.

The social behavior for a species has evolved as a result of the animals' own need to adapt to changing circumstances. For all animals, with human beings as a possible exception, there is no conscious adaptation to their social behavior. They just react in an appropriate way which they are not able to change immediately if the surroundings would go through a fast change. Their social behavior is gradually changed due to the evolution, not as an implication of a conscious acting. This is partly also true for human beings

since also our social behavior is a product of the evolution even if we to some extent can affect the outcome in a changing environment

In order to keep the group together and get all the group members to strive for a common good, communication is necessary but a language is not. All communication has a purpose but for most animals this aim is innate and part of the phenotype and not explicitly known for the individual.

In the fifties, the ethologist Karl von Frisch, demonstrated that that honey bees communicate about the location of good sources of nectar when they return to the hive after a foraging trip. A returning forager will often execute a rather stereotype figure-of-eight dance on the vertical surface of the honeycomb. The speed of the dance indicates the distance of the nectar source from the hive while the angle of the bar of the eight to the vertical indicates the compass direction relative the sun.

Without doubts the bees are communicating but apparently it should not really be counted as a language in the human sense. It was very stylized and could communicate only a limited number of facts about an extremely restricted range of topics. It is also instinctive; the bees do not “know” what they are saying. Hence, the bees do not convey knowledge in the sense that meaning is transmitted⁶.

It is when meaning is going to be conveyed that language comes into play. Mere communication is not enough to convey meanings. Language is a social good and its purpose is to communicate social meaning. Due to Gärdenfors (1993), social meaning is the departure for an explanation of the conventional nature of language. From the honeybee example it is clear that communication can be meaningful without conveying an explicit meaning even in social groups but the social behavior in those cases is highly restricted. It is when the concepts in the surroundings are the object of communication as a language is necessary; the main purpose of every language is to admit communication of concepts. People communicate ideas by transmitting descriptions of the concepts of which the ideas are made up. Such descriptions are with necessity finite otherwise there would be no way to convey the whole message. Despite that, the interpretation of the description may very well be about infinity. This fact is not limited to human language but is also true of formal languages as used e.g., in logic. As an example we can take *the axiom of infinity* that states that the set of natural numbers are infinite. The axiom is of course a finite description, otherwise we would not be able to finish the interpretation, but it yields infinity when interpreted set theoretically.

Knowledge of the world is much a consequence of our existential perceptions. We do not know the world; we have to create it in terms of existential objects. Our perceptive mechanism is connected to the cognitive structure. Meaning comes out of the cerebral description mechanism and is consequently perceptually grounded.

With a language it is possible to influence the mind of another individual. Language allows us to communicate ideas, which implies the possibility of having a social

⁶ The description of honeybees is taken from Dunbar, 1996.

meaning. Without language, we each live in our own separate mental world. With language, we can share the worlds inhabited by others.

Considering this discussion of language as a conveyor of cognitively created concepts, social behavior can be better understood and due to Berger and Luckmann (1966), the central question of sociological theory can be put as follows: How is it possible that subjective meanings become objective facilities?

The same question is also raised by Gärdenfors (1993, p. 290):

But, if everybody can mandate his own cognitive meaning, how can we then talk about *the* meaning of an expression? And how can somebody be *wrong* about the meaning?

Gärdenfors as well as Berger and Luckmann maintain that knowledge is socially distributed while their approaches are different. Berger and Luckmann conclude that the sociology of knowledge must therefore concern itself with the social construction of reality. Reality is socially constructed and it is the aim of, what they call, *the sociology of knowledge* to analyze the process in which this occurs.

[...] our conception of the sociology of knowledge implies a specific conception in general. It does *not* imply that sociology is not a science, that its methods should be other than empirical, or that it cannot be “value-free”. It *does* imply that sociology takes its place in the company of the sciences that deal with man *as* a man; that it is, in that specific sense, a humanistic discipline. An important consequence of this conception is that sociology must be carried out in a continuous conversation with both history and philosophy or lose its proper object of inquiry. This object is society as part of a human world, made by men, inhabited by men, and, in turn, making men, in an ongoing historical process. (Berger and Luckmann, 1966, p.189)

In Berger and Luckmann’s view the sociology of knowledge and the language in which the knowledge is conveyed should be holistically treated which is an implication of their claim that the sociology of knowledge presupposes a sociology of language. Also Gärdenfors (1993) claims that the meanings are individually constructed but the social meaning of a locution is not determined by the mental conceptual structure of a single individual. Instead the jointly (social) meaning, together with a semantic power structure, determines the social meaning.

From these connections between sociality and language we may conclude that one meaning of being social is to be aware of how other people feel and think but also to be

able to understand how other people feel and think. Psychologists rather confusingly refer to this as a *theory of mind* (ToM). Due to Dunbar (1996),

[H]aving a ToM means being able to ascribe beliefs, desires, fears and hopes to someone else, and to believe that they really do experience these feelings as mental states. We can conceive of a kind of natural hierarchy: you can have a mental state (a belief about something) and I can have a mental state about your mental state (a belief about a belief). If your mental state is a belief of my mental state, then we can say that 'I believe that you believe that I believe something to be the case'. These are now usually referred to as order of 'intensionality'. Thinking about mental states in this way yields the following rough hierarchy.

Machines such as computers have zero-order intensionality: they are not aware of their own mental states. Presumably, we also have zero-order intensionality when we are in coma, and most insects and other invertebrates are also zero-order intensional beings.

This is in sharp contrast to what is believed in the agent community. Here computing agents are ascribed intensional capabilities as having beliefs and, even more diffuse, as having attitudes.

Most social animals, possibly with chimpanzees as an exception, has no ToM and are regarded social, from a human perspective only, because they live in organizations where they exhibit a behavior that is best described as being social. This way to talk about animals is in fact anthropomorphic, that is, ascribing human attributes to animals and natural phenomena.

As is clear, a necessary condition for having a ToM is to share a language. You have to be told that others can have beliefs that differ from yours. We are not born with the knowledge that other people can think differently than we and can have other beliefs. This is proved by the fact that not all people have a ToM.

Autistic people are characterized by two key deficits. One is a consistent failure to pass the false belief tests. The other is an apparent inability to engage in pretended play. (Dunbar, 1996, p.88)

"The false belief test" asks the question if a child is aware that someone else can hold a false belief (or at least a belief that the child supposes to be false).

Language is the glue of the social behavior of human beings. It allows us to exchange knowledge amongst ourselves so that the whole community becomes wrapped up in the same set of beliefs. Without language, this "same set of beliefs" has to be part of the

species phenotype and built in by the evolution. However, a language is superior to the built in case both for the obvious reasons that the built in set with necessity is restricted to a small part of the environment and that a language makes it possible for the individuals themselves to change the social meaning. The built in cases can be regarded as “algorithmically” steered in the sense that the behavior follows a strict algorithmic pattern.

5 Formalization of Beliefs

What is strikingly with all attempts to characterize social agents in the agent community is that language is left out of account and not considered important for being social. As shown above, the reverse is true; language is of paramount importance for being social. Despite not taking language in account, it is asserted that software agents may have beliefs and moreover can share beliefs.

As already pointed out, the concepts *belief* and *attitude* in agent technology is not analyzed, it is just postulated. The question never seems to be raised whether these concepts really are representable in a formal system. Since, as we will see, there is a close connection between belief and truth, and accordingly provability, we may compare with the pioneering work of Gödel (1934) when he was about to prove his incompleteness theorems. In order to do so he was forced to formalize the concept of *proof*. He could not just state that there is a formula that says of a sentence that it is provable. Showing that *provability* is representable in a formal system was a heavy task that required Gödel to develop forty-six recursive relations (functions). Even the seminal work of Turing (1936) is a good example that concepts cannot just be stated to be representable in a formal system. According to Wang (1995), Turing’s analysis is a good illustration of what is meant by an analysis of an intuitive concept. Like computability, *social influence*, *having beliefs*, and *having attitudes* are really intuitive concepts.

What does it mean to have a belief? Perception and conception are ingredients in beliefs. By our cognitive apparatus we existentially perceive things in the reality. When coming to universal propositions about the world, this is an outcome of an inductive inference, a *guessing* statement, that we do not perceive existentially but whose general properties we believe; the stated belief is reasonable since it gives a reasonable explanation of reality⁷. This way to create universal statements (beliefs) is significant in all sciences and is normally called *hypotheses*. That means that a belief is in fact a *model* of reality since we will never know the reality. It is namely what we believe about the world that becomes our model, i.e., the model is our belief. This has been unfolded at great length in [Ekdahl, 1997] and I will just give a short justification of the proposition starting with an example taken from physics.

⁷ For an explanation of inductive inference, see Ekdahl 2000B.

Copernicus held the belief that the sun and the planets orbit around the earth. This became his model from which he built a theory. It was when the predicted value from the theory did not agree with observations that the model was questioned and finally abandoned to the benefit of another belief. The whole of physics rests on beliefs and all physical theories are theories of beliefs. In fact *every empirical science relies on beliefs*, which forms the models. Thus there is no such relation as *belief* and accordingly there could neither be a theory of beliefs.

Due to Gödel, formal system is nothing but a mechanical procedure for producing theorems and Turing machines yield an exact equivalent concept of formal system. In a postscriptum (3 June 1964) to his 1934 paper Gödel makes the following comment:

In consequence of later advances, in particular of the fact that, due to A. M. Turing's work, a precise and unquestionable adequate definition of the general concept of formal system can now be given [...] Turing's work gives an analysis of the concept of "mechanical procedure" (alias "algorithm" or "computational procedure" or "finite combinatorial procedure"). A formal system can simply be defined to be any mechanical procedure for producing formulas, called provable formulas.

Now, if we consider a computer an interpreter of a description (program sentences) we can regard computers as linguistic systems opposed to the conception of computers as machines. Then the model of a program is a model of a formal system in the meaning of logic; the program is the formal system.

In a seminal work, Tarski (1935) showed that arithmetic truth is not representable (describable) in arithmetic. It has the implication that semantics is not describable in any formal system comprising arithmetic. An immediate consequence of that is that generally, the model of a formal system is not describable in the system itself. It is to say that we cannot describe a language in the language itself. In general, a language can be described only in a metalanguage. Thus, if we regard computers as linguistic systems, we are forced to describe the language in a metalanguage on conditions that such a language is provided. The metalanguage cannot coincide with the object language or be translated into it but must be of a higher order.

Tarski (1935, p. 273) has formulated the relation between object language and metalanguage in the following way.

A. For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, expressions of the language itself, and of terms from the morphology

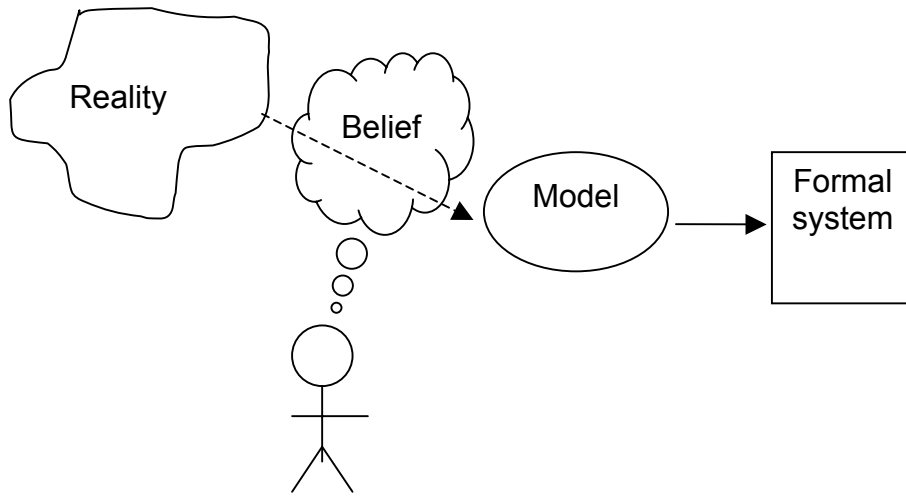


Fig. 5.1: The relation between belief and formal system

of language – but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation.

B. If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed.

Tarski (1935, p. 408) also concludes that, “for any given deductive theory it is possible to indicate concepts which cannot be defined in this theory, although in their content they belong to the theory, and become definable if the theory is enriched by the introduction of higher types”. As is clear from A and B above, higher types are not always possible to describe in the language itself.

There is a close connection between belief and truth. When we have a model of a formal system, the model is our belief that we have formulated as a mathematical structure. We may again compare with physics, which is the science of “guesses about nature”. The description of this model, the axioms in the formal system, is interpreted in the model as true as are all deduced sentences. Consequently, if we cannot describe truth in a model we are of course to an even less extent able to describe the model. However, the contrary is exactly what is proposed in the agent community. An agent, that *is* a formal system, in which the truth of its own sentences cannot be formalized in the system

itself, that is, is not part of the agent, is supposed to contain a belief (truth) concerning a model. The relations between beliefs and computer programs (formal system) can be illustrated as in figure 5.1.

Returning to the discussion above of social meaning, we may now express it as a model that is shared by the individuals in the group. A common model (belief) has emerged out of different models. (Human beings do of course not reason in formal systems but in models.) There is a kind of social power that steer the individuals' different beliefs to a common belief in order to maintain the society; we have to agree upon a model in order to be social. This is in fact a trivial fact also for animals with "built in semantics. In these cases it is solely the evolution that is responsible for their common model (phenotype).

6 Conclusion

The main motivation for developing multiagent systems is the belief that software agents, that is, computer programs, can attain socially responsible behavior. Behind this idea is the belief that social behavior is recursively describable and thereby representable in a computer language. If this was the case, then sociology would be a well defined topic but this, as every sociologist known, is not true.

The only social behavior that has been proven recursively describable is rational behavior as it was analyzed by von Neumann and Morgenstern in the forties, and by their followers. Rational behavior means to behave in a way that is reasonable from a logical point of view. That is, the behavior is completely described as a combinatorial play. Games, as von Neumann saw it, were, as mathematical concepts, completely axiomatizable. The social aspect of game was introduced when social economy was considered a game. In von Neumann's and Morgenstern's time economic behavior was perhaps more conducted by rationality than psychology, which seems to be the case to day.

Even if they saw game as a social behavior, their view of game was in fact not in need of a sociality. In principle, we can play the game together with one person sitting on the Moon, another on Mars, a third on Neptune and so on. The knowledge of the rules was the most important thing together with the assumption that all the players were following the rules in order to get the most out of the play. There was no need for knowing other persons belief. This situation is similar to solitary animals that also know the "rules" but have no need for sociality; the other animals are just other players.

Thus, the situations that von Neumann and Morgenstern analyzed are special situations, which are not of main concern in sociology. The only *social* in the situation was that the game, economic or others, was played by people. This is of course a shortcoming that is recognized by the agent people and in order to keep the idea of agent

alive, the social situations must be broadening to also include such things as beliefs and attitudes, both individually and collectively. This is important for cooperation.

In order to expand the ideas of sociality from economic behavior, a language is necessary in which meanings and ideas can be conveyed. It is the individuals' own concepts that should be possible to transfer to other individuals in order to get a common understanding of the reality in which the social group lives. Ideas evolve as time goes on and a presupposition for being able to change ideas is partly to have a perceptual and cognitive apparatus in which existential concepts are created, partly a language in which those changes can be made explicit and communicated to other people. Such a language is not available for computers since no formal system can be furnished with the requested properties. Not even can the semantics of a formal system be described in the system itself. The lack of a language implies also that a computer cannot have a *theory of mind*, which is necessary in order to be able to know that other individuals may have different views of reality, that is, that my *beliefs* not necessarily are shared by other individuals. Consequently, computers do not have beliefs or other cognitive capacities. The result can be formulated in the following simplified chain of implications:

No language, no theory-of-mind, no beliefs, no social behavior!

When we expand the mathematical analysis to social situations outside games, mathematical methods fell short. This is because very few properties in the social science are recursively describable. The whole idea with social agents that have beliefs and other mental properties is caused by the misconception that the Turing machine is a model also of other cognitive capacities than mere calculation. This is a flaw that is caused by the confusion between theory and model; between the formal system and its semantic structure. We are so imbued with our own way of viewing the world that we easily transpose it into the worlds of Turing machines.

References

Berger, Luckmann, 1966, "The Social Construction of Reality: A treatise in the sociology of knowledge", Anchor Books, Doubleday.

Church, Alonzo, 1935, "Abstract of Church, 1936", *Bulletin of the American Mathematical Society*, May, abstract 205, pp. 332-333.

Dunbar, Robin, 1996, "Grooming, gossip, and the evolution of language", Harvard University press, Cambridge, Massachusetts.

Ekdahl, Bertil, 1997, "Computerized Agents from a Linguistic Perspective", Ph.D. thesis, Lund University, Lund, Sweden, October 10, 1997.

Ekdahl, Bertil, 1999, "Interactive computing does not supersede Church thesis", *The Association of Management and the International Association of Management, 17th Annual International Conference*, San Diego, California USA, August 6-8, 1999, Proceedings Computer Science, Vol. 17, Number 2, Part B, pp. 261-265.

Ekdahl, Bertil, 2000A, "Agent as Anticipatory Systems", *4th World Multiconference on Systemic, Cybernetics and Informatics (SCI 2000) and 6th International Conference on Information Systems Analysis and Synthesis (ISAS 2000)*, July 23-26, 2000, Orlando, Florida, USA.

Ekdahl, Bertil, 2000B, "Anticipation, Induction and Learning", in Daniel M. Dubois (ed.), *International Journal of Computing Systems (IJCAS)*, Published by CHAOS.

Ekdahl, Bertil, 2001, "How Autonomous is an Autonomous Agent?", *5th World Multiconference on Systemic, Cybernetics and Informatics (SCI 2001) and 7th International Conference on Information Systems Analysis and Synthesis (ISAS 2001)*, July 22-25, 2001, Orlando, Florida, USA.

Gasser, L., 1991, "Social conceptions of knowledge and action: DAI foundations and open system semantics", *Artificial Intelligence* 1991, 47, pp. 107-138.

Gärdenfors, Peter, 1993, "The Emergence of Meaning", *Linguistics and Philosophy*, 16, pp. 285-309.

Gödel, Kurt, 1934, "On Undecidable Propositions of Formal Mathematical Systems", in Davis, Martin (ed.) 1965, *The Undecidable*, New York, Raven press, pp. 41-73

Hewitt, C. E., 1991, "Open information systems semantics for distributed artificial intelligence", *Artificial Intelligence*, 1991, 47, pp. 79-106.

Jennings, Nicholas R., 1999, "Agent-based computing: Promise and perils", Proc. *16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, Stockholm, Sweden, (Computers and Thought award invited paper), pp. 1429-1436.

Jennings, N. R., and Campos, J. R., 1997, "Towards a Social Level Characterisation of Socially Responsible Agents", IEE Proceedings online 19971021, *IEE Proceedings on Software Engineering*, **144**, No. 1, February 1997, pp. 11-25

Kalenka, S., and Jennings, N.R., 1997, "Socially Responsible Decision Making by Autonomous Agents", Proc. *Fifth International Colloquium on Cognitive Science*, (ICCS-97), May 1997, San Sebastian, Spain.

Leonard, Robert J., 1995, "From Parlor Games to Social Science: von Neumann, Morgenstern, and the Creation of Game Theory 1928-1944", *Journal of Economic Literature*, Vol. XXXIII (June 1995), pp. 730-761.

Neumann, John von, 1928, "Zur Theorie der Gesellschaftsspiele", *Mathematische Annalen* 1928, **100**, pp. 295-320; translated by S. Bargmann as "On the Theory of Games of Strategy", in *Contributions to the theory of games*, vol. 4, (eds.), Albert Tucker and R. Duncan Luce, Princeton: Princeton Univ. Press, 1959, pp. 13-42.

Neumann, John von, and Morgenstern, Oskar, 1944, "Theory of Games and Economic Behavior", Princeton Univ. Press.

Newell, Alan, 1982, "The Knowledge Level", *Artificial Intelligence*, **18**, pp. 87-127.

Ovum Report, 1994, "Intelligent agents: the new revolution in software".

Pais, Abraham, 1991, "Niel Bohr's Times, In Physics, Philosophy, and Politics", Clarendon Press.

Panzarasa, Pietro, and Jennings, Nicholas R., 2001, "Social Influence and the Generation of Joint Mental Attitudes in Multi-Agent Systems", Proc. *Eurosim Workshop on Simulating Organisational Processes*, Delft, The Netherlands.

Petri, Charles, J., 1996, "Agent-Based Engineering, the Web and Intelligence", *IEEE Expert*, December 1996.

Smit, Ruud and Verhagen, Harko, 1995, "On being social: degrees of sociality and models of rationality in relation to multiagent systems", Proc. *AAAI-95 Fall Symposium Series Rational Agency: Concepts, Theories, Models and Applications*.

Tarski, Alfred, 1935. "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica.*, **I**, 261-405, (English translation 1956 in *Logic, Semantics, Metamathematics*, Oxford, (Second Edition, second printing 1990, J Corcoran (ed.), Hacklett)

Turing, Alan M., 1936, "On computable numbers with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, ser. 2. vol. 42, 1936-7, pp. 230-265; corrections, *ibid*, vol. 43, 1937, pp.544-546.

Wang, Hao, 1995, "Reflections on Kurt Gödel", A Bradford Book, The MIT Press, Fourth printing.

Wegner, Peter, 1998, "Interactive foundations of computing", *Theoretical Computer Science* 192, pp. 315-351.

Werner, Eric, 1988, "Social Intentions", *European Conference on Artificial Intelligence (ECAI-88)*, pp. 719-723.

Wooldridge, Michael, 1997, "Agent-based software engineering", *IEE Proceedings on Software Engineering*, **144**:26-37.