# From Autistic to Social Agents

Frank Dignum
Utrecht University
P.O.Box 80089, 3508 TB
Utrecht, The Netherlands
F.P.M.Dignum@uu.nl

Gert Jan Hofstede
Wageningen University
P.O.Box 8130, 6700 EW
Wageningen, The Netherlands
gertjan.hofstede@wur.nl

Rui Prada
INESC-ID and Instituto
Superior Técnico
Technical University of Lisbon
Lisbon, Portugal
rui.prada@tecnico.ulisboa.pt

## ABSTRACT

The theory, models and architectures of intelligent agents are based loosely on the theory of intentions from Bratman resulting in the so-called BDI agents. Although this functions well for single agents it has been long recognized that this approach falls short for multi-agent systems. It lacks appropriate social aspects to make natural interaction possible. The original concept for intelligent agents was based on a (simple) idea of how people reason about actions. We propose that we go back to the foundation and acknowledge that people are in the core social beings. I.e. we don't function as rational agents with the addition of some "sociality" modules to make us aware of other people. Rather we are social at the base and this sociality pervades all our reasoning, motivation, and any other aspect of our behavior. In this paper we propose a new set of core cognitive elements to replace the BDI approach and discuss the paradigm of a social landscape. However, although we aim for a radical change in the way the community creates social agents we also believe that the new approach incorporates previous work, such as BDI. Our claim is that deliberation about actions and BDI are certainly a part of how agents cope with a dynamic world, but are not the core part of *social* agents that are part of a *social* world interacting with other agents and humans in a natural way.

## Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Intelligent Agents

## General Terms

Theory, Design

## Keywords

Vision, Social Intelligence, Multi-agent systems

## 1. INTRODUCTION

In recent years we have seen a growing demand of more realistic social behavior of (multi-)agent systems. Last year Kaminka ([7]) has appealed to cure robots from their autism. We also saw a (renewed) interest in the human-agent interaction (although sometimes forgetting the work done in other communities, such as, the intelligent virtual agents and the human-robot interaction communities), because agents are more and more seen as partners for humans rather than tools. In (serious) gaming the use of agents to implement characters seems intuitive, but also requires a believable social behavior from these characters both in their mutual interaction as well as in their interaction with the player. In social simulation the use of agents also becomes common place as a way of modeling individuals and generating emerging behavioral patterns.

However, in all of these application areas we see a rapid increase and then a decline of interest in agent technology. For a large part this seems to be due to the fact that current agent technology is not geared towards implementing truly, realistic social behavior. For most applications existing technology is then extended in order to cover the missing elements. Often this leads to complex architectures that work well for specific applications but whose validity is hard to assess. See [11] for an example in agents for games and [2, 3] for social robotics.

As said before, just stating that agents are social because they have a communication language or can be programmed to work in a team does not make the agents *social.* The agent theory will have to support the design and implementation of social behavior in a better and more fundamental way. There are (at least) two issues that need to be investigated to accomplish this:

1. allow for social motivations. i.e. motivations to reach a social rather than a practical goal.
2. recognize that all actions have both practical and social effects that have to be modeled and accounted for.

The above two issues play an important role when trying to model social interactions. I.e. interactions other than fixed protocols. What can/should agents expect from each other? How do they reason about the interaction? If (as nowadays becomes more common) the interaction is described using norms, when do agents follow the norm or violate it? E.g. A reviewer has to deliver her review before November 25. Why do reviewers comply with this rule? Do they calculate the utilities and comply if the benefits outweigh the costs of complying? In that case many reviewers should be late, because the time investment is large and the explicit benefits of complying (against punishment when being late) are minimal.
In general though it seems that reviewers will follow the

**Appears in:** *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), Lomuscio, Scerri, Bazzan, Huhns (eds.), May, 5–9, 2014, Paris, France.*

norm, *because* it is a norm and as long as most reviewers follow it. They follow it because they want to be a respectable member of the PC. In special situations they will *override* this with situation dependent arguments often with a social character to violate the norm. Those reasons can be e.g. that a partner had to go to hospital in the reviewing period. However, an excuse of having to go to a party is usually not accepted.

It might become clear that we need to model and take into account a whole range of social aspects and motives to model social interactions in a realistic way. This in turn would be needed in order to show realistic social behavior in e.g. robots, games and social simulations.

In the rest of this paper we will expand upon the vision of creating truly social agents and the consequences and benefits of embarking upon this road.

## 2. INGREDIENTS

**Social Motives**

As we argued that in order to create truly social agents we need to start at the basis of the agent motivation again and build sociality from the core it makes sense to look at theories of human motivations (again). One of the most influential theories on basic human motivations is that of McClelland [10]. He argues that there are a number of basic natural incentives that give rise to some motives. These motives can be considered as being the core of "energizing" subsequent action. Besides the biological (homeostatic) motives such as hunger and need for sleep (which are, in fact, not very salient in most of the social situations), McClelland distinguishes four motives: (1) *achievement*, (2) *power*, (3) *affiliation* and (4) *avoidance*.

Most of the agents based on the BDI paradigm are implicitly using only the first type of motive (achievement). Based on their beliefs of the current situation they try to create a plan (and execute it) to *achieve* a goal state. However, the achievement motive also includes less practical, and more social, aspects that are not considered in agents, such as, to try to achieve a certain position or reputation in a social group (e.g. to be the most popular or most helpful). In other words, to achieve goals in the state of the social world.

The power motive is about trying to have an impact on the world and reach a sense of control. For people it leads to behavior that tries to change the physical world just in order to see that one has the capability to do that, but is also used to impact the social context in which people live. Thus, it leads to attempts to influence other people and engaging in status and power manoeuvres with others. It is a very social motive, even though one might not think of persons using (abusing) their power as being "social".

The affiliation motive drives people to seek the company of other people. However, it is not just the company of other people that is needed, but rather to establish and maintain positive interactions (relation) with those people. Therefore, one wants positive interactions that give emotional rewards to all parties involved and lead to further interactions. So, both the quality as well as the quantity of the interactions influence the satisfaction of the need for affiliation.

Finally, the avoidance motive drives people to avoid conflicting and/or bad situations. Thus, if interactions with another person are not pleasant, e.g. leading to high levels of anxiety and discomfort, one will withdraw and avoid future interactions with that person. The motive also is active in a broader sense that it tries to avoid situations in which there is a large difference between the perceived and expected situation. That is, situations with a large cognitive dissonance. It leads to self preservation, seeking certainty, and emotional regulation, which fosters the categorization and simplification of behavior so that it becomes more standardized (and thus predictable).

Taking these four basic human motives as starting point has a number of important consequences. Although we do not have the space here to explain all details of the relations between the motives one can easily see that they constitute a balanced system with approach and avoidance mechanisms. E.g. where the power motive can lead one to seek dominance over other people, the affiliation motive makes sure that this is not done at all costs and is kept within "socially acceptable" bounds. In the same way the achievement motive leads to people to explore new ways to achieve goals, but the avoidance motive takes care that we avoid too much deviation from known situations. Thus, having a system of basic motives like this supports a flexible and situational guidance of agents in a dynamic social context.

**Identity**

Making the agent social has implications on how it perceives itself and the world. People position themselves, and others, in terms of membership of social groups (i.e. reference groups) and social goals are often based on comparison with others. For example, if you want to be an influential AAMAS researcher this means that you identify yourself (at least partly) as an "AAMAS researcher" and you need to know the position and activities of some (prototypical/ideal) AAMAS researchers such that you can ascertain what kind of action is needed to become respected in that group. People will usually identify themselves as part of several reference groups. Some of these groups are quite stable, such as family and profession while others are more volatile, such as the group of people in a shop or at a party. People have different emotional attachments to each of the social groups, which elicits social goals to maintain and pursue certain identities.

It may seem that maintaining an agent identity creates quite some overhead, but this is not actually true. The reference groups come with roles structure, values, norms and prototypical behaviors that can be used to make quick decisions and take action in situations where the group is salient.

**Skills**

In the above we have emphasized the core *social* aspects that should be part of models for social agents. It does not mean that social agents do not also have more individual attributes. One such attribute is the set of skills (or capabilities) that an agent has to perform actions. We assume that different agents have different (although possibly overlapping) sets of skills. The complementarity of the skills leads to dependencies and the need for cooperation to achieve goals. The skills also influence the reference groups that an agent will consider to be part of. If its skills are in high demand in a reference group it will get more recognition (and status) from that reference group. The agent may use a skill just to get the emotional reward of receiving status. On the other hand, the agent may have the skill but not the motivation to use it at its maximum potential.

**Values**

Another core aspect of individual agents is the set of values and their priorities. Many definitions of values exist and many research communities use them in different ways. We see them as criteria with which pairs of situations can be ordered. E.g. the value "environmental friendly" can be used to compare two situations on the basis of how well the nature is preserved in each of them. It can very well be that another value, such as, "comfort" will sort the two situations exactly the other way around. What the person will judge to be the most preferred situation in that case depends on the priority he gives to the values, which can be determined by meta-norms and culture. The ordered set of values of an agent will thus determine its basic preferences for types of situations.

The set of values of an agent is also used to reconcile the set of values of different reference groups the agent belongs to. This is needed for an agent to behave in a consistent (and expected) way even though it functions as member of different groups. We argue that the very abstract level of values is needed to make this comparison, rather than behaviors or even norms, because the groups usually function in different contexts which makes comparing behaviors and norms quite difficult, while the values can be used across wider contexts.

**Social reality**
Agents will still have their model (knowledge and beliefs) of the world. As already argued by Kaminka [7], social intelligence needs recognition of other agents and fundamental understanding of those agents. This is only possible if an agent maintains a model of the others and the history of interaction with them, including the attitudes they present toward others (e.g. status conferrals [8]). However, we do not believe that an agent needs to have a fully detailed model of the mind of each agent it encounters. In many situations it suffices to have a simple model of the role the other agent plays in the reference group and context in which the agents meet. E.g. when we interact with a shopkeeper we (usually) use a model of a generic shopkeeper from which we extract that she will have information about the products in the shop and will be able to assist in buying a product. It also should be enough to store only the most relevant interaction events (e.g. those that generate more emotional arousal).

Another important aspect to take into account is that not all details of social reality are salient at the same time. For example, we may use the identity of AAMAS researcher at the AAMAS conference room, but later, at dinner with the very same people we may use a reference group based on friendship and drop the researcher "attitudes".

## 3. SOCIAL LANDSCAPE

In order for agents to plan for social goals it will be necessary to create some structure in the social reality that the agent perceives and tries to navigate in. One way to start doing this is through the creation of a social landscape as a social counterpart of the physical landscape. It seems rather trivial to talk about the social world as a landscape. However, we often refer to social relations in physical terms. E.g. "me and my brother are really *close*", "we don't see eye to eye" or "she is of a *higher* class". Thus, we seem to think about our social relations in terms of a physical space. This idea is taken a step further by some work from one of the founders of social psychology Kurt Lewin, who talks about principles of topological psychology [9] and by psychotherapists that use physical positions of patients and their relations to get insights into their problems [6].

In a social landscape we position an agent in an environment that indicates the social value of different situations and agents. The landscape is subjective and thus different for each individual agent. Situations that are socially very desirable for the agent might be thought of as lying upwards, while undesirable situations lay more down. Closeness in this social landscape indicates social similarity. Thus, agents that are socially similar are in close proximity. Social interactions denote movements in the landscape. Successful (or positive) interactions can tighten the social ties between agents and move them closer together. Note that getting closer to an agent does not always mean that one gets in an overall more preferred social situation. Because agents that you want to be close to, of course, also have interactions with other agents they might be a moving target. Therefore, one should not just have positive interactions with such an agent but also frequent interactions in order to update the relation and keep in the neighborhood. This idea is confirmed by reality where not only the character of the interactions but also the frequency matters for keeping a good relation.

The above paragraph only briefly sketches some of the possibilities of using a social landscape metaphor. One of the prime reasons to use such a metaphor is that it enforces in a very natural way some basic social consistencies. It also facilitates the reuse of algorithms that have been used to move and plan objects and agents in a physical world. However, we have presented social interactions in this section as if they only have an effect in the social landscape. Reality is more complex. We only have interactions that have both a physical and social effect. Here, with "physical effect" we refer to functional effects in the agent environment. This can be, for example, a bid in an auction or an update of a database or choosing the distance, facing and speed while approaching another agent. Every social interaction needs to be performed through such physical interactions and, thus, social interactions also have physical effects (and vice-versa). E.g. small talk in a pub is meant to maintain social relations. However, it means that the person has to go to the pub, possibly spend some money and time there. So, the social interaction also has physical repercussions. We see this connection between physical and social reality as one of the challenges to create social agents.

## 4. COMPUTATIONAL MODELS

Creating a social landscape to structure the social reality for an agent is one way to operationalize some of the social aspects an agent has to monitor and navigate. Given all the (social) aspects that *social* agents have to consider we claim that it is no longer effective to use a deliberation cycle based on that of BDI type agents. E.g. it is not the case that we can start with the basic motives of an agent, choose the most salient one for the current moment and then plan some actions based on values and goals. Basic motives influence the choices of the agent at many different points in the deliberation. In the other hand, values are certainly not always explicitly considered when constructing plans. They are often implicit in determining the set of possible plans under consideration.

The above considerations indicate that a traditional fixed deliberation loop through all modules is not effective nor leads to realistic social behavior. So, what alternative can be used? Based on psychological literature [10, 1, 5] it seems a kind of subsumption architecture [4] should be used. In most social situations people do not engage in deep cognitive thinking. This means that different levels of deliberation should take place depending on the situation. E.g. when getting hungry and it is around 12:30 an agent goes to the canteen when at work. Basically, this can be seen as a simple trigger-action pair (or heuristic). It is enforced by habit and the fact that (most) other agents also go to the canteen at that time (thus, driven by the affiliation motive). The structure of social reality provides means to support short-cuts for fast adequate decision making (e.g. by means of norms), but, in addition, we need a good mechanism to evaluate the relevance and salience of a situation that implies a more careful deliberation (e.g. in some cases it is important to violate a norm to support a value that is important in the group).

In general, motives are primary drivers that are always considered when a trigger arrives from the environment. Values are cognitive components that are considered when a cognitive choice has to be made about the course of action to follow. In each case there might be goals involved if several courses of action can be chosen that are equally good with respect to the motive and/or value. Thus we have the traditional goals and plans if needed. Values are often used off-line to connect or prioritize certain types of actions and plans for goals. They will more often be explicitly used when choices have to be made about actions that have long term (or possible irreversible) consequences. E.g. buying a house, choosing a career, etc.

In our opinion current agent models lack important *ingredients*. However, a careful reader will have noticed that our proposed ingredients do incorporate the traditional BDI components, but placed in richer contexts rather than having added some extra (social) modules to the BDI model. However, it seems clear that the complexity of adding social reality to agents requires a new type of layered deliberation cycle for the agents rather than the traditional BDI deliberation cycle.

## 5. LET'S GET SOCIAL

We have sketched a grand vision of social agents. The vision is based on some fundamental social aspects from human psychology. Most of these aspects have not been considered yet in a principled way by agent research. The main message is that we cannot just add a few "social" modules to the existing BDI architectures to create social agents. We need to start with a fresh look and build social agents based on social modules. In order to create these social modules we will need to investigate more closely how the social aspects function and how they relate together. Then we can build a theory that can be used to create architectures and implementations of these social agents.

Considering the amount of work involved in the above steps, one might wonder whether the results will ever warrant the investment. We really believe they will! Of course, not all agent systems need all the social aspects that we have outlined in this vision. However, having a grounded theory for social agents it is also possible to show how to select those parts that are needed for situations in which a number of social aspects don't play a role. This will facilitate the compatibility of these agent systems with agent systems that do contain social agents because one can explicitly see which parts should be taken for granted.

It also facilitates later extensions of agent systems with new social aspects. A good example from the past is adding norms to agent systems. Norms are essential social constructs. Because most agent systems lack the required social aspects to implement all elements of norms, the norms have been implemented in many different (ad-hoc) ways. As a consequence it is unclear (and sometimes counter intuitive) how the resulting normative agent system functions.

Finally, a good theory and architecture for social agents might be a basis for connecting different agent communities again. E.g., research performed on intelligent virtual agents seems remote from research done on game theoretic interactions or agent based social simulations. By creating a broader theoretical framework it is possible to show where interesting connections exist and that disparate research results arise from concentrating on different aspects of the social agent. This will open up opportunities for new applications of agents and collaborations between different communities.

We know that the challenge of creating real social agents is a large one. But we are convinced that the rewards will be equally large. Therefore, we already embarked on this adventure and hope this paper convinced you to join us!

## 6. ACKNOWLEDGMENTS

This work was supported by a Grant from The Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS) and by FCT under project PEst-OE/EEI/LA0021/2013. It also benefited from discussions with Yoshihisa Kashima, Virginia Dignum and Catholijn Jonker.

## 7. REFERENCES

[1] M. Argyle. *Social Interactions.* Transaction Publishers, USA, 2009.

[2] R. Aylett, G. Castellano, B. Raducanu, A. Paiva, and M. Hanheide. Long-term socially perceptive and interactive robot companions: challenges and future perspectives. In *Proceedings of the 13th ICMI*, pages 323–326. ACM, November 2011.

[3] C. Breazeal. *Designing Sociable Robots.* MIT Press, Cambridge, USA, 2004.

[4] R. Brooks. Intelligence without representation. *Artifial Intelligence*, 47:139–159, 1991.

[5] G. Gigerenzer, R. Hertwig, and T. Pachur. *Heuristics.* Oxford University Press, New York, 2011.

[6] B. Hellinger. *Love's hidden symmetry.* Zeig, Tucker and Co., USA, 1998.

[7] G. Kaminka. Curing robot autism: A challenge. In *AAMAS 2013*, pages 801–804, May 2013.

[8] T. Kemper. *Status, Power and Ritual Interaction.* Ashgate, Farnham,UK, 201.

[9] K. Lewin. *Principles Of Topological Psychology.* McGraw-Hill, USA, 1936.

[10] D. McClelland. *Human Motivation.* Cambridge University Press, New York, 1987.

[11] B. Silverman, D. Pietrocola, B. Nye, N. Weyer, O. Osin, D. Johnson, and R. Weaver. Rich socio-cognitive agents for immersive training environments: case of nonkin village. *Journal of Autonomous Agents and Multi-Agent Systems*, 24(2):312–343, March 2012.