

# Situational Deliberation Getting to Social Intelligence

Virginia Dignum<sup>1</sup>, Catholijn Jonker<sup>1</sup>, Rui Prada<sup>2</sup>, and Frank Dignum<sup>3</sup>

<sup>1</sup>Delft University of Technology, The Netherlands, email:  
{M.V.Dignum,C.M.Jonker}@tudelft.nl

<sup>2</sup>INESC-ID and Instituto Superior Tecnico, Universidade de Lisboa, Portugal, email:  
rui.prada@tecnico.ulisboa.pt

<sup>3</sup>Utrecht University, The Netherlands, email: F.P.M.Dignum@uu.nl

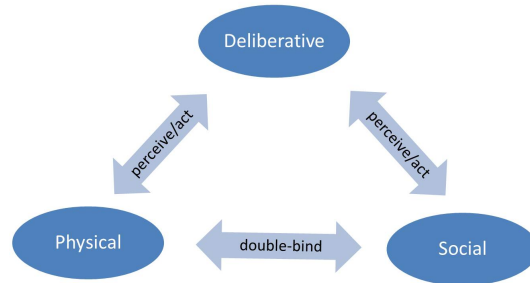
**Abstract.** Socially intelligent systems exhibit, understand, and reason about social behavior, in order to support people in their daily lives. We claim that a fundamental new approach based on social concepts is needed to build these socially intelligent systems. In this paper, we explore how the concepts of social practices and social identities can be used to structure deliberations about actions. We then show the consequences for the architecture and reasoning capabilities of these systems.

## 1 INTRODUCTION

Even though there are enormous advances in Artificial Intelligence, Natural Language Processing, and Vision and Planning, the vision of pioneer AI researchers of truly intelligent systems is still far from reality. Artificial systems must be endowed with forms of social intelligence engrained in the core of the systems reasoning, such as have been developed by humans[3]. Social reality is not given, but socially constructed [13],[2]. Current systems still have very limited understanding of their context, and of their social role. They are not able to reason about their identity and goals in a social context, and therefore cannot be expected to function outside the situations they've been designed for.

The sociability of current robots and virtual characters is engineered into their system, in a situational and context dependent way. Social signals are not appraised as such but implicitly built into their functionality. Therefore, their behaviour is not conceived as social outside that particular context, and they are not able to adapt to significant changes. This implies that reuse in different social contexts or cultures requires a complete re-engineering of the system. A next step forward in AI, is the ability to perceive, reason about and exhibit *social* intelligent behaviour. This will require a framework containing explicit social principles that can be described, represented and manipulated in a symbolic way.

Thus we argue that deliberative, social, and physical principles must be considered first class components of a computational theory of social intelligence (cf. Figure 1). Being socially intelligent requires a keen understanding of the principles of social reality, and the ability to link social interpretations with individual goals into plans and



**Fig. 1.** socially situatedness

vice versa. The double-bind represents the interrelationship between social and physical contexts, where the social context defines the possible social interpretations of the physical reality and limits the set of admissible actions; and the physical context determines and constrains the possible social contexts. For example, a raised hand can mean many things: in a class room: a question, in an auction hall: a bid, on the street: a greeting or a threat.

It is clear that splitting the context in a social and physical one adds quite some complexity. If an agent plans for a social goal (such as gaining acceptance in a group of soccer fans) it needs to plan physical actions to reach such a goal. Thus any plan has both social as well as physical consequences. The agent has to deliberate about all these elements in order to decide upon the best cause of action. It might be clear that agents need some structures in order to limit and facilitate this process. In this paper we describe the first steps towards agents which can be called socially intelligent.

In order to make our arguments more concrete we sketch a use case scenario in section 2 to illustrate our points. In section 3, we identify the key elements of a framework to build socially intelligent systems. We describe how motives can be used to give an agent both social as well as physical direction. Then we describe how social identity, norms, values and social practices can be used as means to preserve both consistency (at individual and society level) as well as simplicity. In section 4, we propose an abstract architecture for social reasoning built on top of the concepts introduced in section 3. In section 5, we sketch how this architecture can be used in order to determine a cause of action in the example scenario, described in section 2, in an efficient way. Finally, we draw some conclusions and indicate future work in section 6.

## 2 SCENARIO

Imagine a situation where a group of people is walking together in the street. They are all fanatic fans of a soccer team A. In the particular situation team A just played against team F. This contextual fact highlights their common social identity of being fans of team A. They feel and act according to that identity which promotes three core values:

unconditionally proud of team A, disrespect for authority and aggressiveness towards rivals. As a result they are loud, bragging about team A and show a rebel attitude.

The situation develops as a group of supporters of team F appears. Team F supporters typically wear a T-shirt of the team. This new context (the presence of another group) reinforces the salience of the social identity in the fans of team A. The two teams have a history of rivalry that is more salient today because the two soccer teams competed earlier that day. For the fans of team A this new situation presents an opportunity to express their value of “aggressiveness towards rivals” which they should do to be worthy members of the social group. Therefore, beating-up members of the other group becomes an appropriate course of action and a fight between the two groups emerges.

During the fight one of the fans of team F gets badly injured, which changes the situation again. A new social identity may become salient supported by this new contextual fact. People from both groups may now use a common social identity of the good citizen that promotes the values of respecting authority and helping others in need. If that social identity prevails they will stop fighting and help the injured person. However, this is not necessarily the case. Members of both groups will have a conflict of identities and the identity that prevails depends on factors, such as, their commitment to the identity of fans of the team and their respect for the status of members that might trigger the identity shift. For example, a leader of a group has more chances of stopping the fight than a low status member.

While a shift to the citizen identity is still in the air, an ambulance arrives with two paramedics. They are easily recognizable by the symbols they wear. This introduces a new social identity in the context, i.e., the one of the paramedic with values of saving people. The presence of the paramedics may highlight the fact that a person is injured promoting the adoption of the citizen identity by the soccer fans. On the other hand, the paramedics can also be seen as an authority group by both groups of soccer fans, which they should disrespect. The groups may join together bound by a common identity of fanatic soccer fan (hooligan) that is independent of the team they support, and together fight the paramedics avoiding the injured person from getting care.

The paramedics when arriving on the scene have their own identity and corresponding social norms and practices. Important to them is their own safety and helping the wounded. In the midst of the fanatic soccer fans they have to weigh the risks: try to help the wounded, take off to protect their own safety, or wait for the police to try to both take care of the wounded and stay safe. The paramedics rush towards the injured person. The hooligans start harassing the paramedics.

### **3 SOCIAL FRAMEWORK**

In order to create truly social intelligent systems, we need to start from an understanding of the motives for behaviour [6]. There seems to be a very basic need to balance between novelty and control in (social) persons. I.e. in the one hand we seek out new situations, while in the other hand we try to avoid too much novelty and strive for control of our environment. The balance between the two forces is different between persons, but always present. However, it does not really indicate easily how we can get to concrete actions. Thus we look at theories of human motivations [10] in order to get more

concrete handles on drivers of behaviour. McClelland argues that there are a number of basic natural incentives that give rise to some motives. 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*.

Goal-oriented systems (such as e.g. BDI agents) are implicitly based only on the achievement motive. Based on their beliefs of the current situation, BDI agents try to create and execute a plan to *achieve* a goal state. This motive clearly serves to satisfy the need for novelty as the agent tries to achieve a different (new) situation. Of course the degree of novelty in a situation depends on how much an agent can predict the outcome of its actions and the consequences of reaching a certain state. Driven by the motive for achievement whenever people enter a new situation, they immediately start exploring the environment. This exploration can take long or can be done in a split second depending on circumstances. The exploration enables a person to react quickly to events as the appropriate social practices in relation to the possible social identities have already received a high probability of execution.

The power motive is about trying to have an impact on the world and reach a sense of control. Thus, this motive clearly serves the need for control. Training skills (such as driving skills) can be driven by the power motive when the mastery of the skill gives a person more control over his environment. If an agent masters a skill it can predict the outcome of its actions better and thus control the amount of uncertainty. However, it also leads to attempts to influence other people and engaging in status and power manoeuvres with others in order to control the people in one's environment. Again, if one can control other people it becomes easier to predict their behaviour and thus control the amount of uncertainty.

The affiliation motive drives people to seek the company of others. However, it is not just the company of other people that is needed, but rather to establish and maintain positive interactions (relations) 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. This motive serves a need for social novelty. Contacts are more interesting if one has to add something to the feelings and knowledge of the other. Thus, one constantly tries to vary the interactions slightly in order to keep interactions going. Whether one interacts with the same persons or different persons depends per individual. If one needs more control it is safer to stick to people one knows, but if one needs more novelty meeting new persons is positive.

Finally, the avoidance motive drives people to avoid conflicting and/or 'bad' situations. That is, 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). So, it is very clearly serving the need for control. It is a main motivator for trying to categorize the en-

vironment in which a person lives. Categorizing and standardizing means that real-time reactions can be generated in most situations and thus more (and new) situations can be sought out. This is of prime importance in a more and more complex and dynamic world. We will discuss this issue in more detail in the next subsection.

As said above, people strive to keep a balance between novelty and control. When we consider environments with many people and social interactions between people (requiring both physical and social planning) it is clear that the need for novelty is easily satisfied by the environment. The agent does not need any special mechanisms for creating novel situations besides the achievement and affiliation drives that it has. Rather there is a clear need to simplify the context of a social agent, satisfying the need for control, in order to be able to deliberate and effectively decide on a course of action in a complex and dynamic environment in which decisions have to be made real-time. In the next subsections we discuss a number of mechanisms that play an important role in this respect for human deliberation and that can be used for social intelligent agents as well. They are a kind of concrete ways to fulfill the power and avoidance motives on a very high level. First of all we describe three constructs that are used to ensure consistency of behavior over time: Identity, Norms and Routines/Habits. Finally we discuss the concept of *Social practices* as a way to describe the combination of social and physical context with respect to a (standard) course of action.

### 3.1 Identity and Values

A social intelligent system must be able to perceive itself and (its position in) the social world. People position themselves, and others, in terms of membership of, possibly many, social groups (i.e. reference groups) and social goals are often based on comparison with others [14] [15]. For example, if you want to be a good parent this means that you identify yourself (at least partly) as a “parent” and you need to know the position and activities of some (prototypical/ideal) good parents such that you can ascertain what kind of action is needed to become respected in that group. Some reference 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.

What constitutes a “good parent” relates to the set of values and their priorities associated with a reference group. 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. Values are used to reconcile the different reference groups the person belongs to, such that her behaviour is consistent (and expected).

From the above it follows that once an agent selects an identity that it wants to assume or that the situation triggers, the identity comes with a set of values and thus prescribes a certain type of behaviour. For example, when a person has the identity of a medic she will every time when she gets to an injured person try to help that person. There is no deliberation about whether she would prefer to go shopping or go on a

date at that moment. Identities are also social, because they give people the sense of belonging to the reference group. The identity is visible for others and expectations can be formed on the basis of knowledge about the reference group. Thus, the identity provides consistency of behavior on an individual level (because an individual with an identity will behave according to that identity) as well as on social level (because all individuals with a certain identity will behave similarly in the same context).

### 3.2 Norms

Norms are the second construct that can be used to categorize and classify behaviours. Norms specify behaviours that promote values. We will not get into all the different kinds of norms and description of their properties at this place (we refer to [1] for a recent overview on norms in multi-agent systems). For the purpose of this paper the function of norms as behavior regulation is the most important. Regulative norms can be described by seven elements: the activation- and termination condition, the normative direction (obligation, permission or prohibition), the action, the violation condition, the punishment and the repair. Norms will indicate whether in a certain context (when the activation condition is true) an action is obliged, permitted or forbidden. E.g. “when a person gets attacked by another person he is permitted to defend himself”. The activation condition of the norm is that a person gets attacked. The termination condition for the norm is when the attacker stops attacking. A norm like “medics are not attacked” is always active and thus the activation and termination conditions are “true”. The violation condition indicates how one can detect the violation of a norm. Thus, what counts as an action that can be seen as attacking a medic? Does that start with verbal abuse? Or should physical violence be involved? The punishment and repair are the actions needed to get from a violation situation to a normal situation again. E.g. the punishment for fighting might be paying a fine or going to prison, while the repair might be to pay for the damages. Although the punishment and repair are important for how norms function in society, the mere fact that the norm exists and is known is enough for most people to follow the norm in normal situations, without deliberating for alternatives.

Norms also have an individual and social side just like identities. When an individual has accepted a norm it means that that individual will act according to that norm (in the appropriate context where the norm is active). Thus the norm ensures individual consistency of behaviour. The norm also has a social side, because norms indicate what is socially acceptable behaviour. Thus, they ensure consistency of behaviour, not just over one individual, but over all persons for which the norm is active. Often the activation conditions of norms refer to reference groups that are also used for identities. It, therefore, seems logical to include the norms pertaining to particular reference groups with the description of that identity. E.g. the “hooligan” identity will come with some norms on how to behave around games and groups of other clubs. It should be kept in mind though that we do not assume a particular order of importance between norms and identities. The particular context will determine the most salient aspect and based on that the other aspects can be connected and become salient as well. E.g. when a hooligan sees a person gets seriously hurt he might directly start assisting that person and abandon his identity as hooligan in favor of that as a good citizen. In the other hand

a hooligan might pick a fight with other people because he sees himself as a hooligan and this is an important norm for hooligans.

### 3.3 Habits

According to [11] habits are psychological dispositions to repeat past behaviour. The definition already indicates why they are important for simplifying the deliberation process. When a certain behaviour has become a habit we do not deliberate about that behaviour anymore, but just repeat it whenever it is triggered by the context. As we will see in the next part on social practices habits are close to practices. However, we see habits as something that can in principle also be completely material driven, while social practices also have a social dimension.

Our daily life is full of habits. According to the psychology literature they are related to goal directed behavior and their origin can often be traced to a deliberate action with a particular goal. However, when the action is repeated often enough in the same condition we can become “conditioned” and perform it without thinking about the goal anymore. E.g. many people have the habit to get up when the alarm sounds. After that they perform a sequence of actions including getting dressed and eating breakfast. All these actions are often performed half-conscious and without deliberation about a “best” course of action. Creating habits frees up our mind to deliberate on more important things and leaving the reaction to many daily recurring situations to our habits.

As many behaviour is performed out of habit, habits can also be used to predict behaviour. This is explicitly done in the Conumat agent model that is used for consumer simulations [8]. It makes the agents simple and the focus can be on the cases where habits are broken. Habits also lead to consistent behaviour. Exactly because they are repetition of the same behaviour in similar conditions.

### 3.4 Social Practices

Having looked at several mechanisms to simplify deliberations about actions while keeping consistency in the previous sections we now turn to the issue on how to manage the three elements of social intelligence mentioned in the introduction. We need to look for structured theories that govern the rules with which these three elements are related. In the social science in recent years the concept of *social practice* seems to target exactly the same problem. Researchers in social science [7, 12] have identified three broad categories of elements of social practices:

- Material: covers all physical aspects of the performance of a practice, including the human body and the actions that can be performed as part of the social practice. This relates to our physical aspects of a situation.
- Meaning: refers to the issues which are considered to be (socially) relevant with respect to that material, i.e. understandings, beliefs and emotions. This relates to our social aspects of a situation.
- Competence: refers to skills and knowledge which are required to perform the practice. This relates to our notion of deliberation about a situation.

One could see a social practice as an elaborate condition-action rule. First a situation is assessed in order to check whether a social practice is relevant. When it is relevant all material elements get a social interpretation (partly) determined by the social practice. Thus, a group of people can be given meaning as being “my friends”, “a rival gang” or “soccer fans”, depending on the social practice that is activated. Within a social practice several behaviours are possible. Depending on the competence that a person has, she will disregard some. Then an action is chosen for which a kind of standard social effect is also expected within this social practice. Of course all elements from the previous sections play a role in social practices as well. People in a social practice are given social meanings through their social identities. The actions available in social practices are often (social) norms and when a social practice is often followed the actions within it become habits. In this way all the concepts of the social framework are directly related.

The components are linked by individuals when carrying out a practice. Each individual embeds and evolves meaning and competence, and adopts material according to his or her motives, identities, capabilities, emotions, and so forth, such that a practice can then be implemented as a composition of components. However, because the social practice is also a shared notion one can expect the other participants in the social practice to share the same interpretation of the situation. There is an (expected) shared situation awareness.

Individuals and societies typically evolve a collection of practices over time that can be adopted in different situations. Social practices are like social norms in that they emerge from individuals, but are not dependent on the individuals any more. They are continuously shaped when they are followed and can differ for individuals with different experiences. E.g. we all share an understanding of the greeting practice, but the exact behaviours and social connotations may differ. Moreover, depending on the situation, the personality and the skills of an individual, carrying out a practice will be a more automatic or a more deliberated process. This corresponds to the different modes of thinking proposed by psychologists<sup>1</sup> [9]:

- System 1, or fast thinking, operates automatically and quickly, with little or no effort and no sense of voluntary control. This includes recognition, perception, and orientation.
- System 2, or slow thinking, allocates attention to activities that demand a high amount of mental effort. Such activities include complex computations, rule following, comparisons, and weighing of options. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration.

When a social practice is experienced very often the interpretation of the situation can be done in a standard way and leads to a quick decision on an optimal behaviour. E.g. when driving a car we hardly ever think about using the shift or the clutch when changing gear. However, when learning to drive we have to first learn which are the salient elements in the environment that trigger the gear change. Thus, we react slower and need more attention for the driving behaviours themselves. This is an important issue

---

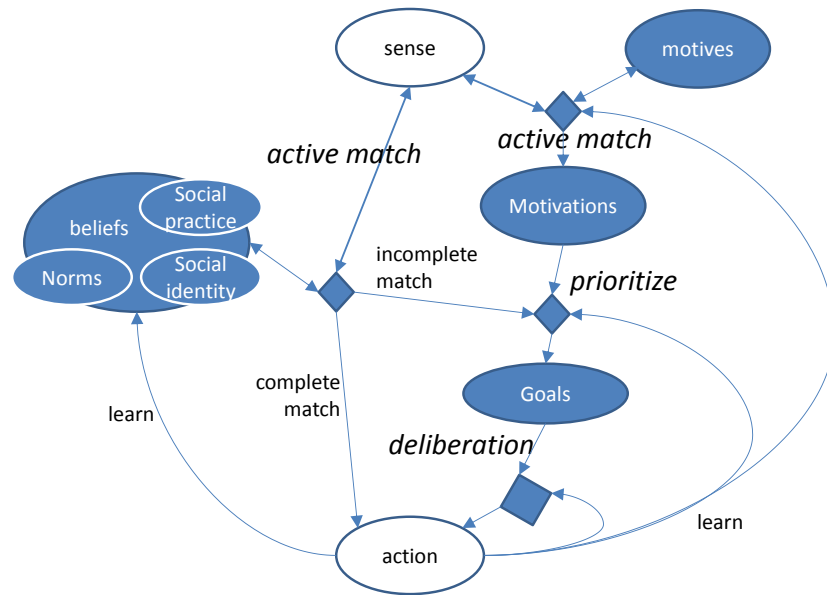
<sup>1</sup> The labels of System 1 and System 2 are widely used in psychology.



for persons in crisis situations, where decisions have to be taken quick. Experienced persons will very quick distinguish the salient elements in the situation and decide which social practice is most salient and act according to it. In the next section we will go more in-depth into the architecture for individual deliberation based on the use of social practices.

## 4 ARCHITECTURE

Social practice theories are, until now, mostly descriptive. Recently, a model was proposed to investigate the emergence of social practices [7]. However, this model does not explain how social practices are used in deliberation and how they are influenced by and influence individual agents. The abstract architecture, depicted in Figure 2, is a first attempt to capture this interaction.



**Fig. 2.** Abstract architecture for social reasoning

The architecture shows the prominent place of social practices as an early input into the deliberation process. Although the sensing still is the start of the deliberation it is actively guided by both the motives as well as the social practices. As soon as social practices are selected as potentially fitting with the current situation they will drive the search for salient features in the environment that fit that social practice as well and might be used for determining (further) actions. E.g. if it is likely that a robot is entering a greeting situation it will start looking for an extended hand of the other person. This does not mean that this is the only feature that is searched for! Several patterns can be

searched in parallel. However, their number is limited to patterns that can be *expected* within a current social practice.

A similar drive to search for patterns comes from the motives. E.g. when entering a room with many people a robot with a high affiliation motive starts looking for (possibly known) persons to interact with, while a robot with a low affiliation motive might scan the room for a place to stand quietly and wait for orders. Thus, we see that the parallel tracks of pro-active and reactive behavior already start with the sensing behavior. Whether the motivation actually leads to setting a new goal actions are influenced by the social practices again. If the robot is entering the room when a meeting is about to start, it might find a quiet place and people will be happy that it does not interfere with their activity. However, when there is a birthday party the social practice might dictate that the robot should take orders and start delivering drinks (even when it would prefer to stand quietly in a corner).

From the architecture one can also distinguish the fast and slow reasoning tracks. The most obvious fast thinking track is the leftmost track from sensing to acting. If a social practice matches the features of a situation to an extent that it dictates a cause of action right away, this leads to a reactive action. Note that we do not need all aspects of a social practice to be completely determined before action is taken (unlike the way traditional frames would work). E.g. when a person gives a robot the order to get a drink, the robot can start getting the drink right away, possibly without first deliberating whether the drink is still available or what kind of relation they have to the person. A second fast track (that is less obvious from the architecture figure) can be taken when a social practice is determined from features in the environment, while the same situation in combination with a motive also leads to a motivation to a certain type of behavior. This motivation plus the social practice can then directly lead to the most appropriate course of action to satisfy that motivation in that situation. E.g. if a robot has a strong achievement motive to have all persons in the room being happy this can be done in many possible ways through different goals. However, if the robot gets an order from a person to get a drink it can right away start satisfying its achievement motive by achieving the goal to get a drink for that person. No more prioritization between goals is needed and the deliberation is no more than retrieving the appropriate plan from the social practice.

If the course of action is not directly clear from the situation more deliberation takes place. In the figure the deliberation is represented by a simple kite symbol. This deliberation can contain a complex process itself, such as, the ones used in Fatima [4] or BRIDGE [5] containing emotions, goals, intentions, beliefs, roles, identity, etc. Suppose that the persons in the room are all happily chatting together or having a drink. The robot does not have an immediate role in that case and can retrieve several social practices that might fit on the situation. If it has an affiliation motive that it wants to satisfy it can start deliberating which goal and type of situation could be used to satisfy this type of motive. Given that the robot might not identify itself as socialite (related with a role of initiating conversations about every day topics, etc.) but does fancy itself a servant it might search for social practices it can use in which it can play the role of servant in the present situation. Getting drinks or food for people are social practices that would fit, but require that the robot gets an order from a person. This deliberation will lead

to retrieve another social practice that has receiving an order as one of its actions. This can be the social practice of asking a person whether he or she wants something to drink or eat. This might include greeting a person, introducing itself and getting some information from the person before asking for an order or just move closer to attract some attention without disturbing, expecting the person to initiate the interaction. The first social practice might have a social interpretation of getting acquainted with a person and, thus, satisfying the affiliation motive while the second satisfies more the avoidance motive.

When discussing the architecture on the basis of the above scenarios the social practices take a leading role in organizing possible courses of action. Note that our architecture does not depend on a fixed set of plans per goal nor that it needs a large set of plans to be searched through. The social practices combine material and social aspects in such a way that one can start from either side and check the appropriateness of the other aspect for the current situation. This avoids having to reason separately about both aspects and combining them afterwards. Having the social practices can also instantiate elements in the deliberation even if they are not totally clear from the initial interpretation of the context, such as the roles and expected goals. For example, the role of others may not be clear until the robot assigns a role to itself based on a social practice.

The final aspect that we included in the architecture is the learning that takes place after the action has been executed. After each action the system should not just check whether the action succeeded or failed, but also whether it can use the result as feedback on the choices it made during the deliberation and whether it should refine or adjust its library of social practices. E.g. it might notice that it expects a handshake in the greeting practice while not everyone is shaking its hand. Thus it can extend this social practice with some alternative ways of greeting like bowing or hand waving. However, it might also learn that it successfully completed the social practice of fighting and update the salience of the plans it executed for this social practice. In this way it can update its memory even without explicitly storing every interaction. Finally, we should remark that where physical effects of actions can usually be measured with sensors, the social effects are often not visible and have to be derived from consequent actions of the partners. Thus, more subtle sensing and interpretation is needed to learn the most efficient social interaction patterns.

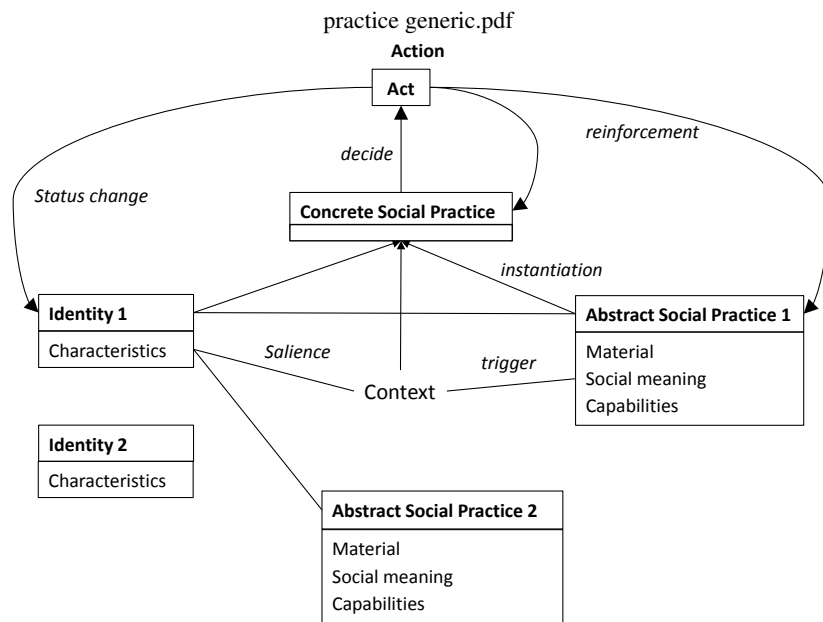
## **5 SOCIAL PRACTICES IN PRACTICE**

The previous section illustrates how our proposed architecture for social intelligent systems could function. This section describes how the elements of identity, norm and social practices can be used in the scenario sketched in section 2 in order to provide efficient and consistent behavior, while still reacting to events in the environment.

People use their identities and motives, and their assessment of the current situation, to determine which action to take and what the meaning of the behavior of someone else is. Roughly, the deliberation process goes as follows (illustrated in Figure 3):

- Determine the most salient Identity for the situation

- Identify abstract Social Practices associated with that Identity
- Match the context to those Social Practices
- Identify a possible concrete instantiation for a matching Social practice
- Decide which action to take from the actions associated with the Social Practice (this choice depends on own goals, past experiences, skills, etc.)
- Take action
- Evaluate the result of actions as feedback to Identity (change of status and commitment) and to Social Practice (reinforcement)



**Fig. 3.** Abstract architecture for social reasoning

We show the use of this process by applying it to the deliberation by actors in our scenario. We assume three identities “Citizen”, “Medic” and “Fan of Team”. All actors have the default “Citizen” identity, and “Medic” actors may also have the “Fan of Team” identity. Figure 4, an actor may decide that its identity of “Fan of Team A” is the most salient in the situation where two groups (of roughly the same size) of different teams encounter each other. The situation triggers the instantiation of the social practice “Fight out-group” to the social practice “Fight F group”. The actor may then decide to take the action “Beat F fans”. The result of this action feeds back into its identity (e.g. losing the fight may lead to a loss of status, which can lead to abandonment of the identity) and into its appreciation of the applicability of social practice.

In order to determine which of the actions associated with the social practice to take, the agent will take into account its goals, norms and habits, as described in section

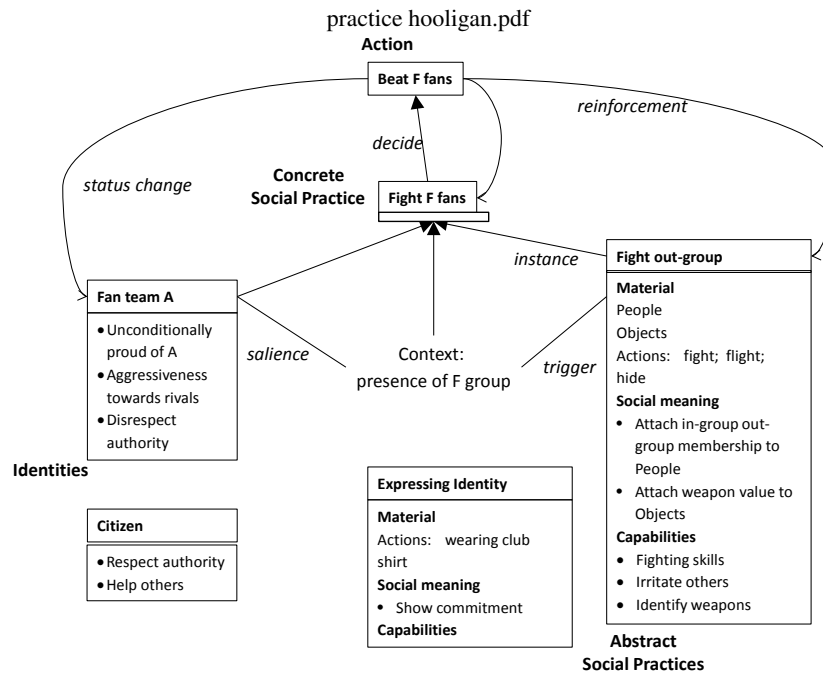


Fig. 4. Social deliberation for a hooligan

4. In fact, the ‘decide’ activity identified in Figure 3 is achieved by the deliberation architecture illustrated in Figure 2.

In common situations, e.g., daily activities, agents will resort to habits, which are practices that require little to no deliberation and yield acceptable results (fast thinking). In such a case, there is a complete match between the information sensed from the context and some social practice the agent knows about (c.f. deliberation architecture in Figure 2). For instance, the seeing a group of people wearing jerseys with the colours of team F, immediately triggers a group of fans from team A to start shouting abuse at the F fans.

In other cases, deliberation is less direct and takes into account goals and motivations of the agent. E.g. a team A fan that has often badly lost fights with F fans, may instinctively duck behind his mates upon an encounter with the other team. This includes the avoidance motive and the goal of preserving one’s health into the deliberation process. However, another fan of the A team may actively seek the fight and even pick on a large opponent if he has the goal of achieving a better status within the group. The reasoning here is that being aggressive towards rivals is a social norm in the group and by upholding to that norm and showing skill (capability) in that activity will improve the view others have on him.

A similar deliberation mechanism for a medic is illustrated in figure 5. Consider a professional medic on duty. The arrival of an emergency call will make his medic identity as the most salient, even if the person may be himself a fan of the A team. Upon

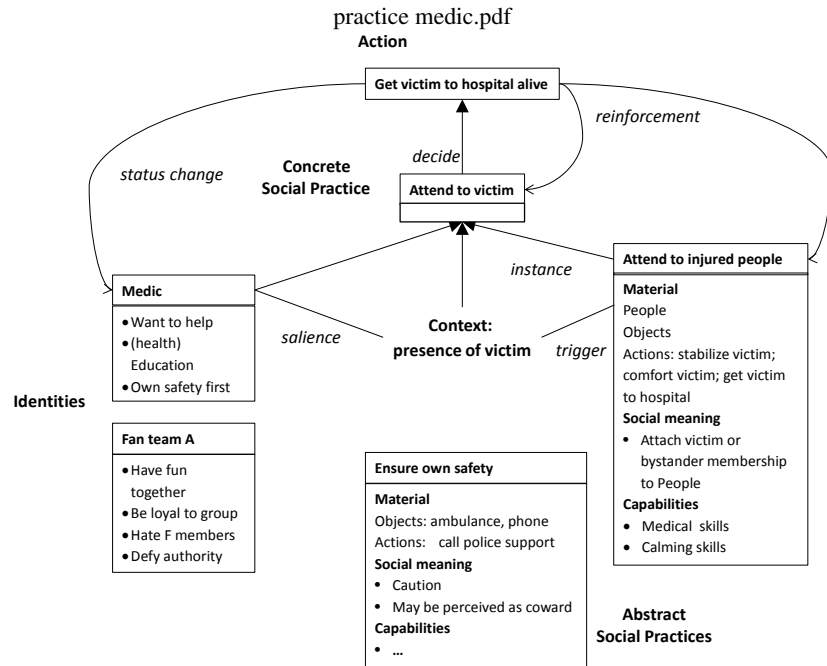


Fig. 5. Social deliberation for a medic

arrival on the fight scene, the medic will tend to automatically, by the force of habit, rush to help the injured without taking into account whether the wounded is a fan of team A or F. That is, the match between sensed input, identity (of health professional) and social practice is direct and complete, and little explicit deliberation is needed. As described in the scenario, this action may cause unexpected reactions from the bystanders, from both teams A and F. In this case the hooligans see the medic as a person of authority, and unite to harass the medic, following their social norm ‘defy authority’. As a result the medic now finds himself in an unsafe situation. By taking into account own goals (e.g. safety) and motives, this changes the context and triggers another social practice for the medic, i.e., “ensure own safety”, which leads to the action of call for police support.

Another example is the case that the medic is himself a fan of team A. In this case, he may take its affiliation motive into account, that makes him wanting to express its belonging to the group, and decide to tend first to a less critically injured A fan, than to the more severe casualty of the other group.

These examples show that many different behaviours can be expressed by the integration of the deliberation architecture described in section 4 and the social reasoning architecture depicted in Figure 3. Even though more extensive evaluations are needed, this scenario illustrates the applicability of the main concepts.

## 6 CONCLUSIONS

We have shown that intelligent systems will need to be more socially aware of their context in order to take the appropriate action. As a result the deliberation of these systems should keep track of both its social and physical context when it deliberates about actions. If one would add the social context as an extra module (or aspect) to traditional intelligent systems this would overload the deliberation cycle and render it inefficient to cope with real time situations.

As a solution we have indicated several elements that can support the social behavior and that lead to a new type of architecture. Social identities, norms and habits are used as mechanisms to prioritize potential behaviours. They also serve to keep individual behaviour consistent over time. Norms and identities also have a strong social component in that they are shared by individuals in a society. Therefore, they can also be used to predict behaviour of others in known situations. E.g. when we see an injured person that is laying on the street and a medic is arriving we assume he will take care of the victim. The use of social practices facilitates the combination of social and physical aspects of a situation. Although these social practices might over constrain the combinations they lead to a good and natural reduction of the complexity. We have shown how the use of social practices in the architecture can lead to the fast and slow thinking patterns that are described in the psychology literature. Using these patterns allows for quick reactions whenever possible while there is a natural entrance for more extensive deliberation as well. Besides facilitating the slow and fast thinking systems, the social practices can also be used to balance between pro-active and reactive behavior. When a current social practice dictates a reactive behavior this will prevail over the pro-active behavior. But when the social practice allows for more freedom of behavior the pro-active drive can influence or even determine the behavior completely as long as it fits within the practice.

Of course we only described a first step on the way towards truly social intelligent systems. We indicated that it is possible to use some mechanisms from social science and psychology to structure the deliberation process in a way that keeps it efficient in the advent of the increased complexity caused by adding the social context to a situation. There are many issues that are interesting for future research. First of all, we have shown how deliberation can take place when a social practice is recognized and triggered. What will happen if more than one social practice fits the situation? Or if no social practice fits (completely)? More work on these aspects will need a more rigorous (formal) definition of social practices and the interpretation processes involved. Another very interesting issue is the emergence and evolution of social practices. We already indicated that there is a feedback from the actions performed in a social practice, but how this will exactly adapt the social practices is an interesting question. We hope that the paper inspired enough to make people willing to embark on this adventure with us rather than just wait for its completion.

### **Acknowledgement**

The work of Frank Dignum and Rui Prada was supported by a Grant from The Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS). Rui Prada was also supported by Portugal's Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013.

## References

1. Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W. N. van der Torre. *Normative Multi-Agent Systems*. Dagstuhl Follow Ups, Germany, 2013.
2. P.L. Berger and T. Luckmann. *The social construction of reality: a treatise in the sociology of knowledge*. Penguin Books, New York, 1996.
3. C.L. Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, USA, 2004.
4. J. Dias, S. Mascarenhas, and A. Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Proceedings of the International Workshop on Standards for Emotion Modeling*, 2011.
5. F. Dignum, V. Dignum, and C.M. Jonker. Towards agents for policy making. In *MABS IX*, pages 141–153. Springer, 2009.
6. F. Dignum, R. Prada, and G.J. Hofstede. From autistic to social agents. In *AAMAS 2014*, May 2014.
7. G. Holtz. Generating social practices. *JASSS*, 17(1):17, 2014.
8. M.A. Janssen and W. Jager. Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics. *Journal of Economic Psychology*, 22:745–772, 2001.
9. D. Kahneman. *Thinking, fast and slow*. Farrar, Straus & Giroux, 2011.
10. D.C. McClelland. *Human Motivation*. Cambridge Univ. Press, 1987.
11. David T Neal, Wendy Wood, Jennifer S Labrecque, and Phillippa Lally. How do habits guide behavior? perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology*, 48(2):492–498, 2012.
12. A. Reckwitz. Toward a theory of social practices. *European Journal of Social Theory*, 5(2):243–263, 2002.
13. J.R. Searle. *The construction of social reality*. Penguin, 1996.
14. H. Tajfel. Social identity and intergroup behavior. *Social Science Information*, 13(2):65–93, 1974.
15. J. Turner, M. Hogg, P. Oakes, S. Reicher, and M. Wetherell. *Rediscovering the Social Group: A Self-Categorization Theory*. New York: Basil Blackwell, 1987.