The AAAI 2017 Spring Symposium on
Artificial Intelligence for the Social Good
Technical Report SS-17-01

Societal Challenges need Social Agents

Virginia Dignum
Delft University of Technology
The Netherlands
m.v.dignum@tudelft.nl

Frank Dignum
Utrecht University
The Netherlands
f.p.m.dignum@uu.nl

Abstract

Big societal challenges involve many interdependent parties at different levels of society that are all interacting and reacting to each other. Although agent based simulations can help to support the analysis of these problems, we argue that we need to develop a fundamental social framework for the agents in these simulations in order to get a grip on the underlying mechanisms at work. We show some scenarios where we applied some of the ideas and sketch the sociological foundation of the agents and an agent architecture that can be used for building agent based simulations for societal challenges.

Introduction

Many of the big societal challenges are difficult to solve because they involve several interrelated complex systems. Specific theories or tools only model one aspect of these situations and thus only solve a problem from that specific point of view. E.g. a political solution for the refugee crisis in Europe was a treaty with Turkey that prevents the refugees to get into the European Union, and in compensation provides Turkey and its citizens easier travel and trade activities with the EU. Although this led to a diminishing number of refugees in Europe it does not solve the refugee problem, nor the collaboration between EU and Turkey. In the same way, a logistics solution to the refugee problem, will focus on optimization of the spread of refugees over an area or the time they spend in temporary camps. In fact, taken independently, none of the solutions is sufficient to mitigate or understand the refugee crisis.

Increasingly, Artificial Intelligence is employed to model these type of policy problems. In particular, Agent-Based Modeling and Simulation (ABMS) frameworks to analyze the possible effects of a policy, and intelligent Data Analytics that support both automated decision-making, where policy decisions are taken instantly by predictive algorithms fed by huge amounts of real-time data, and opinion dynamics, by providing means to collect and analyze public opinions in social media.

Given the diverse interests, backgrounds and knowledge of the parties in complex societal challenges, it seems that agent based models might support the analysis of these situations. However, these agent models should be fundamentally social. In situations where the normal daily way of life is threatened or disrupted, the more fundamental social aspects are needed to explain the behavior and possibly see how to change this behavior. In the refugee scenario highlighted above, ABMS can help predict the effects of each of the two possible policies, but without really understanding the motives and cultural backgrounds of the different stakeholders, these models will only support superficial analysis. On the other hand, it is unclear that an analysis of social media data will give a correct view of the real feelings of populations, in the case of complex problems for which only partial solutions are being presented, as in the above scenarios. Thus, although we believe that agent based social simulations could support the analysis and solution of the societal challenges we need to develop new types of social agents for these simulations.

In order to fully understand the effect of policy events and context on the actors, we claim that rich social agent architectures are needed, that can link agents' decisions to their values, identities, motives, norms and social practices. Moreover, policy-makers and policy-subjects (the populations affect by the policies) tend to have quite different views on the reasons and solutions to a given problem. This calls for participatory approaches, where different groups can design their own view of the situation as an agent model representing their perspective, which can then interact with agents provided by other stakeholders. That is, a fundamental approach to policy-making, requires tools and theories that analyze the challenge from a more abstract level. This requires formal models to represent actors, contexts and events and their interdependencies in domain independent ways.

In this paper, we propose an agent architecture to represent the socio-cognitive grounds of reasoning, based on our decade-long experiences in several domains. This architecture can account for human decision-making and interactions in diverse circumstances, that take into account different value systems and social identities and that also model the motives, interests, values and desires of individuals of all the groups involved. E.g. refugees are grateful to receive shelter and food in a safe environment, but at the same time have aspirations for their lives and those of their families, and will plan to meet goals that fit these aspirations, even

if these are not as expected by the receiving populations. Thus they want to perform useful work, follow education, find their place in the new society and have some prospects for the longer term. All of these elements will determine their behavior while being a refugee and in particular how they interact with governments, local authorities and local communities.

The paper is organized as follows. In the next section, we discuss the need for social agent models, by describing several scenarios where ABMS is used for social good, supporting better understanding and policy making. In section we describe the main components of a social agent, and in section we give an informal account of the social agent architecture. We conclude the paper with an overal discussion and directions for future work in section .

Societal Challenges and Applications

In this section we discuss some societal challenges we have been working on for the past decade or so. They come from quite different domains, but all demonstrate the importance of including the social aspects of agents behavior and interactions into the simulation models.

Fisheries

Governments and the EU regulate a large part of the fishing practices in European countries. Most of the policies in fishery are informed by extensive ecological models that indicate the consequences of catching certain amounts of fish in specific periods. However, policy makers start realizing that fishery management does not only include the catching of fish (e.g. the conditions of resources, technologies, economy) but also considering the social structure of the fishing communities. Thus, the fishery industry should be modeled as a complex socio-ecological system (CSES) (Heidar and Dignum 2016). In such a system, almost all of the elements are influenced by social phenomena such as culture and social norms (Pastrav and Dignum 2016). E.g. the introduction of fishing quota that can be traded has resulted in fishing communities falling apart and the quota being owned by a few big companies that fish with very big trawlers and take little care of the environment. The final result of the quota (meant to prevent over fishing) can be negative for the environment. Numerous other examples exist in fishery management that show that the social aspects of fishery communities are of great influence on the outcome of policies.

Urban resilience

Increasingly, detailed environmental data is needed to inform the decisions of citizens and governments that are affected by environmental conditions, such as flooding and heavy rain, bush fires, or tornadoes. For example, in the SHINE project[1], we worked on models to provide early warning flooding systems, or to support choice of transportation based on the weather. Given that necessary data is not always available and its collection using sensors is costly and requires the non-flexible deployment of a high-density sensor network, crowdsensing can be leveraged to obtain fine-grained, weather-specific, just-in time information about environmental conditions (King et al. 2015). For example, monitoring the rain levels around a crucial area of the city can be made possible by making ongoing agreements with a large number of citizens to collect and send rain data with their cellphones whenever they are in the area.

However, owners of existing sensor resources, such as rain sensors on bicycles communicating via cellphones, will likely not want them to serve all requests all the time. Owners will want to maintain control over their resource by specifying use policies governing how, when and to whom their resource can serve requests. For example, owners of resources could specify that they do not want to serve requests that prohibit them from turning their device off, or require serving untrusted people. Furthermore, a resource owner can have requirements to be met for the use of their resource, such as payment. Since a request expresses what the requester wants the provider to do and not do, a request can be seen as a set of *norms*. In (King et al. 2015) we introduced *use policy* as the means to govern a resource by specifying what norms ought and ought not be entailed by a request. A use policy is a set of meta-norms since it specifies norms about norms. From a policy-making perspective, the interest is in understanding how to provide incentives for resource owners to set their own use policies in a way that it benefits the overall aims of the data collection task, ensuring that enough data is collected to provide assessments on the weather conditions, and derived traffic and flooding areas information.

Lifestyle coaching

AI systems are increasingly used for tailored advice in several lifestyle or health related situations, including smoking cessation, weight control, physical activity. We present here two systems developed in our labs.

Robin: Coping with Cyber-bullying Children spend long amounts of time online, and are therefore increasingly in risk of cyber-bullying. Cyber-bullying can be defined as ‘any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others’ (Tokunaga 2010). Robin is a virtual character that ‘lives’ on the computer screen of potential victims of cyber-bullying (van der Zwaan, Dignum, and Jonker 2012). When a child feels uncomfortable because of a cyber-bullying incident, it turns to Robin for emotional support and practical advice on how to deal with the situation. Implementing the buddy as a local application on the host computer, also serves as a protection of the child’s privacy, as its data will not be shared with the social web applications the child uses. The short term goal of the buddy is to lower the victim’s negative emotions (coping). On the long(er) term, the buddy aims at teaching the victim how to deal with cyber-bullying. In order to be successful, the buddy must be able to persuade the user to follow its advice. Interaction between the buddy and the user happens in three stages: first the user indicates she is distressed because of a cyber-bullying incident, next the buddy gathers information

[1] http://shine.tudelft.nl/

about the current situation by asking questions and finally the buddy gives practical advice on how to deal with the situation.

Opting for Vegetarianism Changing behavior is often difficult when the behavior is tied to routines and social practices. In these cases giving information that might change the deliberation and intention formation does not succeed. In this scenario we used a new model in which context influences a combination of both habitual and intentional environmental behavior using the concept of social practices. We illustrated the model using meat-eating behavior while dining out and indicated how this approach can lead to new insights into effective policies for changing behavior (Mercuur, Dignum, and Kashima 2015).

Societal patterns

We have been working on all of the above scenarios (and many more), ranging over many different domains, but in all cases, the agent models need to include some fundamental social concepts such as values, norms, social practices, social identity and motives. By working on these scenarios we observed that when social concepts are modeled based on (simple) rules they offer little in terms of lessons and best practices, that one can transfer from one to another domain. Connections and similarities only become apparent when social concepts are explicitly included as reasoning components. We do not advocate to use a very complex agent model for every application, but rather we advocate a rich social agent framework that supports the choice of the right elements, salient for a specific application. When different agent applications will all be based on this social framework, they will have consistent properties for the concepts that they incorporate and thus can be compared, combined and re-used. In the next sections we will describe each of the social concepts and their role in the framework in more depth.

Social Agents

Societal problems are wicked problems, in the sense that the structure of the system affects and is affected by the behavior of the agents. Given any policy, implemented to ensure some desired system level behaviour, agents typically will figure out how to exploit the new policy to their own needs, often precluding the original aim of the policy. Existing approaches for social system modeling, are often based on simple agent models consisting of a few (reactive) rules. These models are not suitable to capture such adaptive and manipulative behaviour. That is, addressing societal problems using AI, requires a new agent architecture.

In order to create truly social intelligent agents, we need to start with the understanding of the socio-cognitive grounds for behaviour (Dignum, Prada, and Hofstede 2014). Human behavior results from a basic need to balance between novelty and control. I.e. in the one hand we seek out new situations, while in the other hand we try to avoid uncertainty by striving for control of our environment. The balance between these two forces, *approach* and *avoidance*, is different between persons, but always present. However, it does not really indicate easily how we can get to concrete actions. For this end, different aspects of human behaviour, as identified in the social sciences, must be considered. Social sciences have provided many different theories, including values, motives, identity, social practices and norms, to explain this basic drive for behavior. Given that each of these theories only provides a partial explanation or is specific to a certain type of context, we propose to consider all as the building blocks for a social agent architecture. Methods are then needed to determine which are the most salient for the specific scenario being modeled. In the remainder of this section, we briefly introduce these social concepts. In the next section, we describe how they are connected into a social agent architecture.

Values

Human Values (e.g. honesty, beauty, respect, environment, self-enhancement) are a main influencer of human decision making. Schwartzs theory identifies ten Basic Human Values that are recognized throughout all major cultures, and further describes the dynamic relations amongst them. (see e.g. (Schwartz 2006)).

Although values seem intuitively clear they do not have an agreed upon precise definition. From the perspective of building a social agent architecture, values can be seen as criteria to measure the difference between two situations (usually before and after an action is performed). This way, a value can be used for comparing *alternatives* for decisions: the alternative that matches the best with values is preferred. As we already stated, values are seen as criteria, but usually values are to abstract to measure directly. E.g. the value of *wealth* is not measured itself, but maybe we measure the amount of money someone owns. Moreover, different persons can have different measurements for the same value and thus make different decisions at any given moment (which is actually very realistic). One person might e.g. just look at the money someone owns and another look at the monetary value of all his possessions or maybe even all the money he has power to decide upon (including e.g. the money maintained by a partner). (Miceli and Castelfranchi 1989) discusses the consequences of this indirect use of values.

Values combine two core properties, which make that values are a useful "*decision heuristic*":

- *Genericity:* values are abstract in the sense that they can then be instantiated in a wide range of concrete situations (e.g. the value of timeliness can influence all decisions implying time management). Moreover, values can be linked along abstract-concrete (e.g. the "no delay" value is more concrete than the "timeliness" value) and multiple values can be related to one another (e.g. "no delay" is an indicator for "timeliness").
- *Support:* values tend to drive decisions towards objectively desirable outcomes (e.g. being timely generally leads to better outcomes).

Motives. McClelland's theory of basic human motivations (McClelland 1987) distinguishes four motives for human behavior, besides the biological (homoeostatic) motives such as hunger and need for sleep: (1) *achievement*, (2) *power*,

(3) *affiliation* and (4) *avoidance*. All humans have these basic needs, but each person will have different characteristics depending on their dominant motivator. This dominant motivator is largely dependent on our culture and life experiences. Motives provide a balance between *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, but is kept within "socially acceptable" bounds.

The *achievement* motive drives persons to create and execute a plan (based on their beliefs of the world) to *achieve* a goal state. The *power* motive is about trying to have an impact on the world and reach a sense of control. It also leads to attempts to influence other people and to engage in status and power manoeuvres with others. The *affiliation* motive drives people to seek the company of others, to establish and maintain positive interactions (relations) with those people. Finally, the *avoidance* motive drives people to avoid conflicting and/or 'bad' situations. It leads to self preservation, seeking certainty, and emotional regulation.

Social Identity Social identity is a person's sense of who they are based on the social groups (family, football team, work, etc.)they belong to. That is, people position themselves, and others, in terms of membership of, possibly many, social groups (aka reference groups) and their social goals are often based on a comparison with others (Tajfel 1974). People have different emotional attachments to each of their social groups, which elicits social goals to maintain and pursue certain identities. E.g. what constitutes a "good citizen of the USA" relates to the set of values and their priorities associated with a reference group. A socially intelligent agent must also be able to perceive itself and (its position in) the social world. This requires the ability to reason about the motives and goals of the others and possibly also about how others would reason about oneself. Such Theory of Mind (Whiten 1991) model can be quite complex and is not necessary in all situations. However, at the least a representation of the social networks one belongs to and one's position in these networks must be part of the agent's reasoning.

Social practices In order to cope with the complexity of combining social and physical aspects of reality and planning in such an environment, we use the concept of *social practice* (Reckwitz 2002). Social practices are conditioned behaviors that result from past interactions. Researchers in social science have identified three broad categories of elements of practices (Holtz 2014):

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

These components are combined by individuals when carrying out a practice. Each individual embeds and evolves (through conditioning) meaning and competence, and adopts material according to its motives, identities, capabilities, emotions, and so forth, such that it implements a practice. 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.

Norms Norms are behavior regulation mechanisms specifying behaviours that promote values. 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". 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. For a recent overview on norms in multi-agent systems we refer to (Andrighetto et al. 2013).

Norms 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. From its social side, 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.

Towards a Social Agent Architecture

In (Dignum and Dignum 2014), we discuss the requirement for social agents to combine the pro-active drive, based on social motives, with reactive behaviour. Social practices can be used as structuring tools to keep the complex deliberation process efficient. Figure 1 shows an abstract architecture that integrates these two types of behaviors based on the social components described in the previous section.

As in traditional BDI reasoning, agents revise their beliefs and goals, based on sensory information about the context. However, in our architecture, **social practices** drive the sensing process in the **context management** module, and can be seen as an early input into the deliberation process. That is, the sensors will also 'fill-in' some elements of a social practice, making that practice salient to be acted upon. For example, when the context is sensed to be a workday morning, the social practice *going-to-work* is triggered. Further observation of the context will determine which types of transport are available, whether working at home is an option, whether the children have to be taken to school

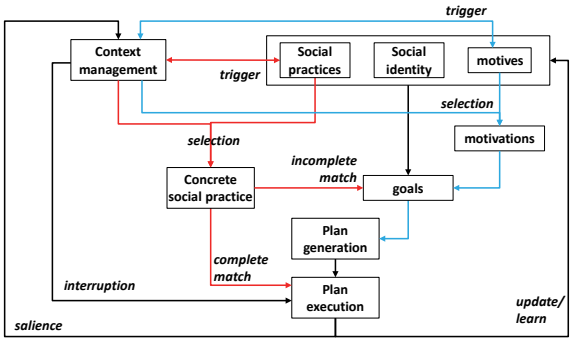

Figure 1: Abstract architecture for social reasoning

by me today, resulting in a more concrete social practice of *going-to-work by-car-and-taking-the-children-to-school*. The agent will then generate a plan for this social practice based on its identity (beliefs, goals and reasoning process). It should be emphasized that the social practices do not replace the traditional deliberation of BDI agents. Rather, they provide a background and behavior patterns that can be used for the plan deliberation. This will increase efficiency and also allows for dividing context checks and preconditions of actions and plans. However, given that sensing is influenced by one's personal characteristics, this module is also fed by the agent's **motives**. E.g. when taking the kids to school a person with a high affiliation motive starts talking with teachers and other parents, while a person with a low affiliation motive might drop the kid and quickly retreat to the car.

The context management module results in a set of **concrete social practices** through a two-pronged process. Firstly, social practices selected as potential fitting with the current situation will drive the search for salient features in the environment that confirm this fit. E.g. a person dropping a kid at school, expects to encounter a *saying-bye* social practice and therefore will look for extended arms of the kid. Obviously, this is not the only feature that is searched for, and several patterns can be searched in parallel. However, their number is limited to patterns that can be *expected* within the current social practice. The second way social practices drive the context management is when a social practice is active. This social practice indicates important states that can be expected and are subsequently searched, and might be used for determining (further) actions. E.g. when a class room is entered and the teacher is present the parent checks for a signal that he has to leave.

If the concrete social practice that is detected to be relevant is specific enough to leave only one course of action open, then this course of action is directly executed. E.g. when a person recognizes that a handshake is used to greet another person she will immediately extend a hand to start the handshake, without deliberation about possible other actions. However, in a meeting between hostile companies the persons might not be willing to shake hands with their opponents. In that case a greeting should still be performed but the course of action is not directly clear from the situation and more deliberation takes place. In the figure this is illustrated by the fact that the concrete social practice is used as input for the **goal module** that will determine possible goals within the context of this social practice. This deliberation can contain a complex process itself, such as, the ones used in Fatima (Dias, Mascarenhas, and Paiva 2011) or BRIDGE (Dignum, Dignum, and Jonker 2009) but here we have limited ourselves for simplicity to goal deliberation. E.g. suppose a CEO is strongly achievement driven. During a management meeting he will try to use a social practice that serves that motivation and maybe directs all discussion towards quick and efficient decision making. However, in order to achieve the ultimate goal of the decisions he needs the strong commitment of his staff. Thus, in order to get quick decisions with a high amount of commitment he will give opportunity for all persons to give their opinions before making a decision that reflects the inputs. He will balance different aspects in choosing a course of action within a social practice.

Thus social practices take a leading role in organizing possible **plans**. 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.

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 decision making and update the priority 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.

Fast and slow reasoning

In this architecture, reasoning happens in two parallel tracks of pro-active and reactive behavior. We have emphasized these tracks by using red arrows for the quick (reactive) deliberation and blue arrows for the deliberation involving the motives and other cognitive concepts leading to the slower deliberation. Whether the motivation actually leads to set-

ting a new goal is influenced by the social practices again. If a parent is dropping of a kid just when the class is about to start, he might leave quickly and the teacher will be happy that he does not interfere with the class. However, when the parent arrives well in time the social practice might dictate that the parent mingles and talks with the teacher and other parents.

This approach, is consistent with psychological findings on how people reason[2] (Kahneman 2011):

- 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 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 subsection we will discuss some essential elements that are needed for the above sketched deliberation process.

Social Instruments

We conclude this section, with a description of essential components of social agents that make social reasoning possible.

Sensing and representing the social landscape
In the previous subsection, we talked about pursuing social goals, having social motives, etc. This somehow presupposes that we can represent *social* states. Of course one can represent states with collections of facts, but it is important to also have formal rules and constraints concerning these facts. Thus we need to analyze and record properties of social relations, such that it is possible to reason about the effects of social interactions in terms of how the social state changes, without having to specify all facts explicitly. Although the metaphor of a social landscape has been used by several authors, there is little systematic research on representing this social landscape (Lewin 1936). In this paper, we only mention it as important and advocate the actual use of a kind of 3D representation of it, because this seems to be close on how people perceive it and allows for the use of many representation and planning techniques that already exist for the physical world.

Link physical to social
Another issue that has to be solved in order to create realistic social behaviour is the connection between physical and social states. The importance of this aspect is already indicated by the example of the introduction. Although the physical and social states are different parts of reality, they cannot be treated completely separate. In fact, we claim that every physical action does have both a physical as well as a social effect. This means that we have to add these social effects to the action descriptions.

Memory
Again, it is obvious that an intelligent system has to keep track of its interactions, e.g. to learn from them. However, the collection of interactions in similar situations (e.g. with the same interaction partner) can be seen as representing the condition of the relation with a partner or situation. E.g. a person will become a "friend" only if there are regular positive interactions with that person. Once a person has become a friend she will only stay a friend if interactions are mainly positive. Thus we need to keep track of some summary of these interactions to gauge the condition of the friendship. This might influence e.g. what kind of favors one could ask from that person.

Conclusions

One possible application of AI to solve societal problems, is to develop agent-based models of social interactions, that provide users and policy makers with the needed insights in the problem. In these models, it will be possible to capture the different values, identities, norms, motives and social practices of the parties involved. Having the social agents interact in the system highlights the reactions and effects of each other's actions and as such gives insights for the parties involved in the societal problem. It is clear from the description of the scenarios that traditional agent models do not suffice for this role and a new type of social agents have to be developed for this purpose.

In this paper, we proposed a new agent architecture, based on social practices that should facilitate socially intelligent behaviour. It puts social context and social motives at the heart of the deliberation rather than use them as additional modules. We have sketched how this architecture provides some structure in the complexity of the deliberation process, facilitates the combination of social and physical aspects of a situation, integrates fast and slow thinking patterns as described in the psychology literature, and balances between pro-active and reactive behaviour. As such it can be seen as combining the features of goal directed (BDI) architectures as exemplified by 2APL (Dastani 2008), situation based reasoning as performed in Case Based Reasoning (Richter and Weber 2013) and workframe based deliberation as done in Brahms (Sierhuis, Clancey, and van Hoof 2009).

Although we intend to formalize this architecture and implement it, we do not expect the whole architecture to be needed for all applications. It will be possible as well to select the most salient features for the application and fill in the rest with default values and processes.

[2]The labels of System 1 and System 2 are widely used in psychology.

References

Andrighetto, G.; Governatori, G.; Noriega, P.; and van der Torre, L. W. N. 2013. *Normative Multi-Agent Systems*. Germany: Dagstuhl Follow Ups.

Dastani, M. 2008. 2apl: a practical agent programming language. *Autonomous agents and multi-agent systems* 16(3):214–248.

Dias, J.; Mascarenhas, S.; and Paiva, A. 2011. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Proceedings of the International Workshop on Standards for Emotion Modeling*.

Dignum, V., and Dignum, F. 2014. Contextualized planning using social practices. In *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, 36–52. Springer.

Dignum, F.; Dignum, V.; and Jonker, C. 2009. Towards agents for policy making. In *MABS IX*. Springer. 141–153.

Dignum, F.; Prada, R.; and Hofstede, G. 2014. From autistic to social agents. In *AAMAS 2014*.

Heidar, S., and Dignum, F. 2016. Exploring socio-ecological models of fishery management using multi-agent social simulation. In *MSEAS (Marine Social-Ecological Systems)*.

Holtz, G. 2014. Generating social practices. *JASSS* 17(1):17.

Kahneman, D. 2011. *Thinking, fast and slow*. Farrar, Straus & Giroux.

King, T. C.; van Riemsdijk, M. B.; Dignum, V.; and Jonker, C. M. 2015. *Supporting Request Acceptance with Use Policies*. Cham: Springer International Publishing. 114–131.

Lewin, K. 1936. *Principles Of Topological Psychology*. USA: McGraw-Hill.

McClelland, D. 1987. *Human Motivation*. Cambridge Univ. Press.

Mercuur, R.; Dignum, F.; and Kashima, Y. 2015. Changing habits using contextualized decision making. In *European Social Simulation Conference*.

Miceli, M., and Castelfranchi, C. 1989. A Cognitive Approach to Values. *Journal for the Theory of Social Behaviour* 19(2):169–193.

Pastrav, C., and Dignum, F. 2016. The role of social institutions in the resolution of transboundary conflicts over shared fishing resources. In *iEMS (8th International Congress on Environmental Modelling and Software)*.

Reckwitz, A. 2002. Toward a theory of social practices. *European Journal of Social Theory* 5(2):243–263.

Richter, M. M., and Weber, R. O. 2013. *Case-based reasoning: a textbook*. Springer Berlin.

Schwartz, S. 2006. An Overview Basic Human Values: Theory, Methods, and Applications Introduction to the Values Theory. *Jerusalem Hebrew University*.

Sierhuis, M.; Clancey, W. J.; and van Hoof, R. J. 2009. Brahms an agent-oriented language for work practice simulation and multi-agent systems development. In El Fallah Seghrouchni, A.; Dix, J.; Dastani, M.; and Bordini, R. H., eds., *Multi-Agent Programming:*. Springer US. 73–117.

Tajfel, H. 1974. Social identity and intergroup behavior. *Social Science Information* 13(2):65–93.

Tokunaga, R. S. 2010. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in human behavior* 26(3):277–287.

van der Zwaan, J. M.; Dignum, V.; and Jonker, C. M. 2012. *A Conversation Model Enabling Intelligent Agents to Give Emotional Support*. Berlin, Heidelberg: Springer Berlin Heidelberg. 47–52.

Whiten, A. 1991. *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. Basil Blackwell Oxford.