Identifying affordances for modelling second order emergent phenomena with the $\mathcal{WIT}$ framework

Pablo Noriega[1], Jordi Sabater-Mir[1], Harko Verhagen[2], Julian Padget[3], and Mark d'Inverno[4]

[1] IIIA-CSIC, Barcelona, Spain
pablo,jsabater@iiia.csic.es
[2] Stockholm University, Stockholm, Sweden
verhagen@dsv.su.se
[3] Department of Computer Science, University of Bath, Bath, UK
j.a.padget@bath.ac.uk
[4] Goldsmiths, University of London, London, UK
dinverno@gold.ac.uk

Abstract. We explore a means to understand second order emergent social phenomena, that is, phenomena that involve groups of agents who reason and decide, specifically, about actions —their or others'— that may affect the social environment where they interact with other agents. We propose to model such phenomena as socio-cognitive technical systems that involve, on one hand, agents that are imbued with social rationality (thus socio-cognitive) and, on the other hand, a social space where they interact. For that modelling we rely on the WIT framework that defines such socio-cognitive technical systems as a trinity of aspects (the social phenomenon, the simulation model and the implementation of that model). In this paper we centre our attention on the use of *affordances* as a useful construct to model socio-cognitive technical systems. We use the example of reputation emergence to illustrate our proposal.

1 Introduction

There is a rich discussion within the COIN community about the properties and uses of open regulated multiagent systems that may be brought to bear upon the modelling of second order emergent phenomena. Second order emergent social phenomena involve agents that not only decide about their own actions but also about the actions of others and on the effect those actions have in the social environment where they interact. Although some second order emergent phenomena have been explained as complex systems it has been argued that agent-based simulation modelling may prove useful not only for explaining emergent features but also to understand motivational, strategic and organisational features that are ascribed to the individuals involved in these phenomena and the outcomes of their activity within a given social environment.

The $\mathcal{WIT}$ framework is one way to describe those multiagent systems. The $\mathcal{WIT}$ framework postulates that coordination artifacts for open regulated MAS are the amalgam of three aspects: (i) $\mathcal{W}$ a socio-technical system the constitutes actual coordination of a particular collective activity in the real *world*. (ii) $\mathcal{I}$ an abstract or *institutional* specification of the conventions that articulate the interactions in that system; and (iii) $\mathcal{T}$ the *technological* elements that implement the institutional conventions and enable the use of the system in practice. The $\mathcal{WIT}$ framework postulates also the type of relationships that should exist between those three aspects and how to characterise classes of socio-cognitive technical systems by link $\mathcal{I}$ with $\mathcal{T}$ through the correspondence of metamodels for agents and social spaces and platforms that implement those metamodels.

We claim that the use of the $\mathcal{WIT}$ framework provides the relevant foundations to deal successfully with the problem of modelling second order emergent phenomena. In this paper we use a specific example of the emergence of reputation to make a first step in this direction. Namely, when rumours about the behaviour of an individual circulate within a group, the reputation of that individual may change. Often, when members of the group perceive that change, they take actions —endorse messages, add pejorative or benign opinions, spread new rumours— that are intended to influence the formation of that reputation. Therefore, as in other second order emergent phenomena, the perceived signals influence the behaviour of individuals, which in turn influences how that reputation evolves.

Based on the $\mathcal{WIT}$ framework, here we focus our attention on the abstract features that are needed to model both this sort of agents and their social space. In particular we use the $\mathcal{WIT}$ framework (section 2) to elucidate the *affordances* required for modelling second order emergent phenomena. We approach this goal by working at three levels, each level being more specific than the previous one. In the first level we propose a tentative list of *affordances* required for a generic second order emergent phenomena (section 4). In the second level we chose a second order emergent phenomenon, reputation, and taking as a reference the initial list of *affordances* and the characteristics of the social phenomenon we build a second, more specific, list of *affordances* (section 5). Finally, in a third iteration, we focus on a specific scenario based on the social phenomenon analysed in the second level. Again, using the first and second list of *affordances* we build a new tentative list that considers the particularities of the scenario (section 6). We end up with a brief discussion of future work (section 7).

2 Socio-cognitive Technical systems. The $\mathcal{WIT}$ framework.

A socio-cognitive technical system (SCTS) is an open regulated multiagent system where agents —that may be human or software— interact on a shared virtual (online) space. We distinguish SCTS from other MAS by making explicit some assumptions about the agents that participate and the form that participation takes. This may be more precise in the following definition (from [10]):

Notion 1 (STSC) *A* Socio-cognitive technical system (SCTS) *is a multiagent system that satisfies the following assumptions:*

A.1 ***System*** *A socio-cognitive technical system is composed by two ("first class") entities: a* social space *and the* agents *who act within that space. The system exists in the real world and there is a boundary that determines what is inside the system and what is out.*

A.2 ***Agents*** *Agents are entities who are capable of acting within the social space. They exhibit the following characteristics:*

  A.2.1 ***Socio-cognitive*** *Agents are presumed to base their actions on some internal decision model. The decision-making behaviour of agents, in principle, takes into account social aspects because the actions of agents may be affected by the social space or other agents and may affect other agents and the space itself [3].*

  A.2.2 ***Opaque*** *The system, in principle, has no access to the decision-making models, or internal states of participating agents.*

  A.2.3 ***Hybrid*** *Agents may be human or software entities (we shall call them all "agents" or "participants" where it is not necessary to distinguish).*

  A.2.4 ***Heterogeneous*** *Agents may have different decision models, different motivations and respond to different principals.*

  A.2.5 ***Autonomous*** *Agents are not necessarily competent or benevolent, hence they may fail to act as expected or demanded of them.*

A.3 ***Persistence*** *The social space may change either as effect of the actions of the participants, or as effect of events that are caused (or admitted) by the system.*

A.4 ***Perceivable*** *All interactions within the shared social space are mediated by technological artefacts — that is, as far as the system is concerned only those actions that are mediated by a technological artefact that is part of the system may have effects in the system. Note that although such actions might be described in terms of the five senses, they can collectively be considered percepts.*

A.5 ***Openness*** *Agents may enter and leave the social space and* a priori*, it is not known (by the system or other agents) which agents may be active at a given time, nor whether new agents will join at some point or not.*

A.6 ***Constrained*** *In order to coordinate actions, the space includes (and governs) regulations, obligations, norms or conventions that agents are in principle supposed to follow.*¶

SCTS abound, these are rather typical examples: (i) classical hybrid online social systems like *Facebook* [11], (ii) socio-cognitive technical systems like public procurement online system and electronic institutions (see [1,6]), (iii) massive on-line role playing games [17] and (iv) agent based simulation systems [17] in particular the ones we discuss in this paper.

A key feature of all STSC, that is common to these examples, is that they are state-based systems, in the following sense:

Notion 2 (State of the social space) *A SCTS involves autonomous entities that interact in a common restricted environment that we call the social space, so that.*

- *At any point in time the social space is in a "state" that consists of all the facts that hold in the social space at that point in time. Such state is unique and, therefore, common to all participants.*
- *The state of the social space changes either through the actions of individuals that comply with the conventions that regulate the SCTS, or through events that are acknowledged by the STSC conventions.*¶

In order to better characterise SCTS and develop guidelines for their design, we proposed an abstract framework —the $\mathcal{WIT}$ *framework*— whose distinctive contribution is the realisation that every SCTS can be understood as a composition of three "aspects": actual functioning system in the real world, the institutional description of the system and the technological artifacts that support the operation of the system. This realisation provides a separation of concerns for each aspect that is convenient for description and design of SCTS. Moreover, the tripartite representation makes explicit features of the co-dependence of the three aspects that become profitable again for design and deployment of SCTS.

These intuitions are made a bit more precise in the next definitions. First, the three views may be characterised by their core ontologies, a *compatibility* relationship and their particular notion of state:

Notion 3 ($\mathcal{WIT}$ views) *The* $\mathcal{WIT}$ framework characterisation *of a SCTS* $\mathcal{S}$ *is the triad:* $< \mathcal{W}, \mathcal{I}, \mathcal{T} >$, *where*

1. $\mathcal{W}$=$< W, \succ>$, *the view of* $\mathcal{S}$ *as a running system situated in the (real) world. It involves*
   (i) a domain ontology W, *that grasps the intuition that only certain facts and actions that happen in the physical world are relevant for the system;*
   (ii) the $\succ$ *notion that grasp the intuition that only certain actions or events that happen in the world are relevant for the system* $\mathcal{S}$ *and if the proper conditions hold, they are "feasible" in* $\mathcal{W}$*;*
   (iii) the state of $\mathcal{W}$ *at time* t, *is the set of all facts that hold in* $\mathcal{W}$ *at time* t:
   $\mathcal{S}_{\mathcal{W}t} = \{\alpha \mid W \succ \alpha\}$
2. $\mathcal{I}$=$< I, \propto>$, *the institutional view of* $\mathcal{S}$ *is the abstract representation of the system and the conventions that govern the actions that may take place in* $\mathcal{W}$ *and its effects. It includes:*
   (i) an institutional ontology I *that grasps the intuition that institutional representation of* $\mathcal{S}$ *involves an ontology that corresponds to the relevant entities in* W *plus other entities that are needed to represent the conventions that regulate the behaviour of agents in* $\mathcal{S}$*;*
   (ii) the $\propto$ *notion that grasp the intuition that, according to the conventions stipulated in* $\mathcal{I}$, *only some attempted actions are "admissible" and hence have any effects in* $\mathcal{S}$.
   (iii) The state of $\mathcal{I}$ *at time* t, *is the set of all expressions that are admitted ("hold") in* $\mathcal{I}$ *at time* t:
   $\mathcal{S}_{\mathcal{I}t} = \{\psi \mid W \propto \psi\}$
3. $\mathcal{T}$=$< T, \bowtie>$, *the technological view of* $\mathcal{S}$ *is the implementation of the system according to* $\mathcal{I}$ *that receives inputs from and produces outputs in* $\mathcal{W}$. *It includes:*
   (i) a collection of variables whose values can be changed through the execution that implement $\mathcal{S}$.
   (ii) The $\bowtie$ *notion that grasps the intuition that the values of some variables change when the system receives an input.*
   (iii) The state of $\mathcal{T}$ *at time* t, *is the set of values of the relevant variables in* $\mathcal{T}$ *at time* t:
   $\mathcal{S}_{\mathcal{T}t} = \{\phi \mid W \bowtie \phi\}$ ¶

An important feature of the $\mathcal{WIT}$ characterisation is that one would like to express that only those actions that are compatible with the conventions of the system can change the state of the system. For that purpose we need to establish some sort of correspondence between actions in $\mathcal{W}$, $\mathcal{I}$ and $\mathcal{T}$and use the compatibility relationships of each view to indicate that the corresponding state changes only if the attempted action is compatible. In particular, we postulate that if an SCTS is properly specified or deployed, the three $\mathcal{WIT}$views are "coherent" in the sense that their corresponding states evolve as intended. In other words, when an action is attempted, in $\mathcal{W}$—which is expressed as an attempted input in $\mathcal{T}$— its effects in $\mathcal{W}$ should be the ones prescribed in $\mathcal{I}$, which ought to be the ones that are computed in $\mathcal{T}$ and are output back int $\mathcal{W}$. The following notion approximates that "isomorphism":

Notion 4 (Coherence) *Let f_{wi}, f_{it} and f_{wt} be three "bijections" between the $\mathcal{WIT}$ views of a SCTS $\mathcal{S}$; and let α, ψ and ϕ be actions in $\mathcal{W}$, $\mathcal{I}$ and $\mathcal{T}$, respectively, such that $\psi = f_{wi}(\alpha)$ and $\phi = f_{it}(\psi)$ and $\psi = f_{wt}(\alpha)$.*
The $\mathcal{WIT}$views are coherent *iff for every time t,*
$(\mathcal{S}_{\mathcal{W}t} \succ \alpha) \Leftrightarrow (\mathcal{S}_{\mathcal{I}t} \propto \psi) \Leftrightarrow (\mathcal{S}_{\mathcal{W}t} \bowtie \phi)$. ¶

In Notion 4 we invoked "bijections" between the three views. This is an elusive concept in the sense that unless one has a precise specification of each view it is impossible to define such functions. Moreover, the relationships between views is no only a bijection that supports the notion of coherence, it needs to account for other parallelisms between views and it should serve to anchor design and methodological concerns (as suggested in [11]. We will not go into these matters here but we we will merely illustrate the type of properties that the interrelationship between views should have.

Fig.1 labels these relationships to indicate that:

1. We call the $\mathcal{I}$ view *institutional* following the usage of Searle [15]. Thus we expect to have an bottom-up correspondence relationship from $\mathcal{W}$ to $\mathcal{I}$ that serves to create the "institutional reality. This is usually achieved through "constitutive norms" that transform crude actions into illocutions but it entails also a top-down correspondence that legitimises the abstract outcomes in $\mathcal{I}$ so that they have an effect in $\mathcal{W}$.
2. As suggested in Notion 3 $\mathcal{I}$ is intended to be a *representation* of the relevant part of $\mathcal{W}$, thus there must be a mapping of entities of the domain ontology: "crude" facts and actions (the relevant part of reality) that provide pragmatic and semantic meaning to terms of the modelling language of $\mathcal{I}$ (as part of the upgoing arrow) but there is also a prescriptive relationship (from $\mathcal{I}$ to $\mathcal{W}$).
3. $\mathcal{I}$ may be understood as a *prescription* of the intended behaviour of $\mathcal{S}$, that is used to specify the software that implements it. But $\mathcal{I}$ may also be seen as a symbolic or formal or abstract *model* of a social phenomenon or a social activity that will be (ideally) *correctly* implemented in $\mathcal{T}$. But there is also the possibility of a bottom-up understanding of $\mathcal{I}$ as a description of the behaviour of the technological artefacts involved in $\mathcal{S}$.
4. $\mathcal{T}$ enables $\mathcal{W}$ because we postulate (in Notion 1) that $\mathcal{S}$is always an online system. Thus there is an input-output connection between $\mathcal{W}$ and $\mathcal{T}$. But notice that those relationships presume that information is not lost or corrupted, that interfaces are

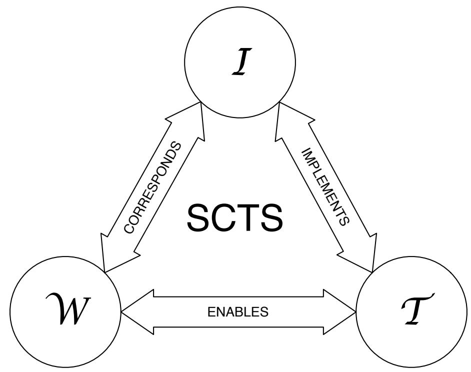

Fig. 1. The WIT trinity: The ideal system, $\mathcal{I}$; the technological artefacts that implement it, $\mathcal{T}$, and the actual world where the system is used, $\mathcal{W}$.

ergonomic and correct so that transfer of information is made according to the conventions stipulated in $\mathcal{I}$.

From a design perspective, the flow from $\mathcal{W}$ to $\mathcal{I}$ and from $\mathcal{I}$ ro $\mathcal{T}$ may be achieved in an ad-hoc manner where using standard software engineering techniques. Another possibility that we discussed in [10] is to rely on a "metamodel" (to specify models in $\mathcal{I}$) and a corresponding "platform" that implements such models. We will comment on this possibility in the next section of this paper but confine our discussion to what pertains to the use of the $\mathcal{WIT}$ framework in the design of agent-based simulation systems.

3 Simulation of second order emergence with the $\mathcal{WIT}$ framework

In broad terms, we want to build simulation systems to study second order emergence social phenomena. As discussed in the next section, these phenomena involve individuals who decide what to do in view of their own motivations and preferences but also taking into account what others may or may not do and the effects of their own actions and the actions of others. Thus, we need socio-cognitive agents. Moreover, these agents do not act in a void but in a social environment that provides them with ques, opportunities and means to interact with other agents and also affect the social space itself. Thus all the assumptions we postulated in Notion 1 apply. Thus we may try to use the $\mathcal{WIT}$ framework to characterise these systems.

Indeed, once we commit to a SCTS representation of the social phenomenon, the rough $\mathcal{WIT}$ characterisation is straightforward: the $\mathcal{I}$ view would be the abstract model of a social phenomenon and $\mathcal{T}$ is the corresponding working model. $\mathcal{W}$is the simulated environment where we observe the social phenomenon and where agents actually interact (if there are humans involved we would have the participatory simulation taking place in $\mathcal{W}$).

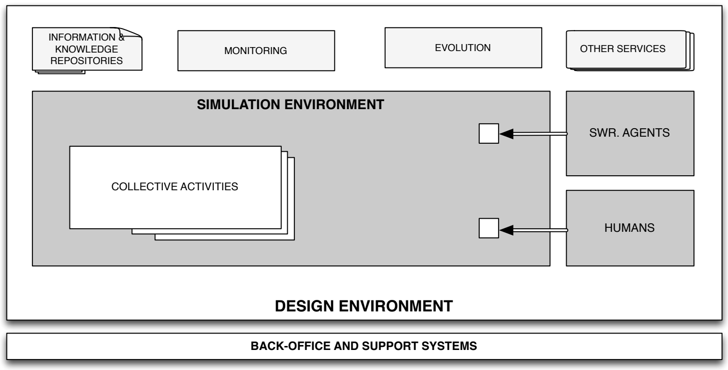

Fig. 2. A simulation platform within a simulation environment

However the $\mathcal{WIT}$ framework becomes more useful when we start traversing from $\mathcal{W}$ to $\mathcal{I}$and postulate the advantage of having a metamodel for second order emergence phenomena and the corresponding platform. Because if we have a metamodel for social spaces and a metamodel for socio-cognitive agents at hand, then we would only need to instantiate these metamodels to the particular traits we want to study. Naturally, once we have good enough models we would then implement them and it would be truly convenient to hava a platform that is mirrors the metamodels and, consequently, is again merely instantiated in order to have the models impelemented.

In [10] we postulated that a metamodel consists of a collection of languages, data structures and operations that serve to represent the agents and the social space of a given SCTS with appropriate level of detail and accuracy. The model, hence, would be a *representation* of a phenomenon through a particular abstract or symbolic notation. A notation that, consequently, will be useful as long as it has the expressiveness needed to capture the relevant features of the phenomenon with the appropriate level of detail.

We do not yet have a metamodel for second order emergence modelling but we know from experience that it is advantageous to have a specific enough metamodel so that instantiation may be natural and easy to debug.We are embarked in that process but we decided to take a bottom-up approach a proceed towards the metamodel starting from a simple scenario and produce progressively more expressive abstractions.

Thus rather than trying to adapt an available metamodel like electronic institutions [6] we proceed from an elicitation of concrete affordances that are specific to the modelling of a particular phenomenon. From these affordances we would choose or develop formalisms or specification languages that make those affordances operational.Similarly, we will start from an ad-hoc implementation of the affordances towards a platform that is closely linked to the resulting metamodels.

The next two sections report on this effort.

Eventually, we will be running experiments in an environment that facilitates the integration of the second order simulation platform with tools and resources for doing experimental research (Fig. 2).

4 Second order emergence

At the core of the old debate on micro foundations (individualism) versus macro properties (structuralism) of societal systems, also known as the micro-macro link problem, we find the notion of *emergence* and how the micro and macro levels interact. Specifically we have to differentiate between two different approaches to the *emergence* of social phenomena.

Based on a *generativist paradigm* [7] we can approach the emergence of social phenomena as a process that goes from micro to macro, from the individuals and their local behaviour to the macro structures that *emerge* as a result of the local interactions. In this approach "the only action takes place at the level of individual actors, and the 'system level' exists solely as emergent properties characterising the system of action as a whole" [4].

This is known as *first order emergence* and is the main approach followed in current state-of-the-art social simulations: "Given some macroscopic explanandum -a regularity to be explained- the *canonical agent-based experiment* [5] is as follows: Situate an initial population of autonomous heterogeneous agents in a relevant spatial environment; allow them to interact according to simple local rules, and thereby generate -or 'grow'- the macroscopic regularity from the bottom up" [7].

This, however, is only half the story. To which extend macro-level properties exercise some kind of causal influence on the micro-level individuals' behaviour? [5] In many cases, in a real human society, many of the macro structures that start to appear as a result of the individual's local behaviour have an effect on macro-level attributes (for example, the creation of ghettos may imply the increase of the crime rates and as a consequence the devaluation of the cost of the houses in that area). The modification of those macro-level attributes, at the same time, has an effect in the individual's local behaviour modifying it (what is known as a *'downward causation'* [5]). This change in the individual's behaviour influences again how the macro structures emerge; how the emergence of the new macro structures modify the macro-attributes; and so on.

The scenario is even more complex if we consider that the individuals may recognise that the phenomenon is emerging and, as a consequence, this phenomenon (and the emergence process itself) can be intentionally supported, maintained, changed or contrasted by the same agents. This is what is known as *second-order emergence*. Many important social phenomena are characterised by *second order emergence*. Examples of these phenomena go from social movements like the african-american civil rights movement, the Arab spring or the 15-M movement in Spain to relevant social constructs like *reputation*.

[5] When we talk about social simulation we have to talk invariably about *agent-based social simulation* (ABSS). The main characteristic of a social simulation is that the simulated *individuals* are not entities whose aggregated behaviour can be properly described using mathematical equations. Every individual is unique and interacts with the other individuals and the environment in an autonomous way. This particularity is what makes the multiagent systems paradigm the predominant approach in social simulation nowadays. From now on, we will use the terms social simulation and agent-based social simulation interchangeably.

Affordances for second order emergence

What are the generic affordances that allow (are necessary for) a *second order emergence* scenario? As we have said, the main characteristic of *second order emergence* is the capacity of the individuals at the micro level to detect that the social phenomenon that will show up in the macro level is starting to emerge. This means that the individuals (or at least some of them) know about the existence of that phenomenon and, more importantly, know about the signals that identify its emergence in a given society. On the one hand, the *social space* makes more or less explicit these signals to the individuals. On the other hand, the individuals need to have the capacity of perceiving them and, more importantly, of interpreting them as indicators of the emergence of the social phenomenon. Invariably this goes through the capacity to anticipate what the other individuals will do in the future, in other words, the individual have to operate with a theory of mind. Theory of Mind is "the ability to understand others as intentional agents, and to interpret their minds in terms of intentional concepts such as beliefs and desires" [8]. Having a theory of mind has been recognised by several authors as a fundamental requirement of an architecture of the social mind [2,16].

The detection of the emergence of a social phenomenon is only the first part of a *second order emergence*. Once the individuals at the micro level become aware of the emergence process, they should have the capacity to influence it. This implies some kind of capacity for action embedded in the individual that at the same time is facilitated by the *social space*.

Said that, a tentative list of generic affordances necessary for a *second order emergence* scenario can be summarised as follows:

Individual affordances

- Cognitive capabilities to understand the emergent social phenomenon.
- Theory of mind. Anticipate what others intend to do, how they will do it and what are they motivations.
- Sensor capabilities to detect the signals that the *social space* makes available associated to the emergence of the social phenomenon.
- Cognitive capabilities to interpret the signals as indicators of the emergent process.
- Actuator capabilities to influence the emergent process.

Social space affordances

- A shared ontology of objects, agents, actions and events.
- Some sort of social model to represent roles, groups, organisations and their relationships.
- Some sort of governance or coordination support.
- Perception channels adapted to the sensor capabilities of the individuals.
- Actuation channels adapted to the actuator capabilities of the individuals.

5 Reputation

Once identified the affordances for *second order emergence*, we will focus on a specific second order phenomenon. In our case we will use a well known social construct: reputation.

Reputation can be defined as what a *social entity says* about a target regarding his/her/its behaviour and characteristics. A social entity is a group which is irreducible to the sum of its individual members, and so must be studied as a phenomenon in its own right [14]. The definition postulates that whomever is saying something about the target is not an individual, but a social entity. An individual is just a messenger of what is supposed to be the opinion of the social entity (in fact, the messenger does not even have to be a member of that social entity to spread a reputation). This is a key aspect because it allows reputation to be an efficient mechanism to spread social evaluations by reducing fear of retaliation [12].

The next important element in the definition above is the action of "saying". Reputation exists because an evaluation circulates. Without communication, reputation cannot exist. You can have the members of a community sharing a belief. This belief however is not a reputation until it starts to circulate. In fact, communication is so important for reputation that there is a specific type of communication specialised for building reputation values: *gossip*.

When messages start circulating and people realises that a reputation on a target is starting to form, many times they will start performing actions (in the form of new rumours, support messages, shame messages, etc.) that are intended to influence the formation of that reputation. Therefore, as in any second order emergent phenomenon, the perceived signals that a reputation is emerging influence the behaviour of the individuals, that at the same time influence how that reputation emerges.

Affordances for reputation

First of all, the individual needs to have a reputation model. This model has to go beyond the traditional computational models of reputation [13] that focus only on how reputation is evaluated. The individual has to be able to influence reputation so it has to know how it spreads (how gossip works), how it is evaluated and what are the elements that lead to the emergence of reputation or its undermining. Notice that this level of knowledge about reputation requires a theory of mind (when the other individuals will spread a reputation value?, who will be receptive to a specific reputation value?). It is also important that the individual knows about the utility of reputation: what is it good for? How reputation can favour/limit the achievement of my goals?

From the previous definition, it is clear that the notion of social group is essential for reputation. It has to be present at both levels, individual and *social space*. An individual has to be able to detect social groups and determine the membership to those groups. At the same time, the *social space* can make more or less explicit this membership to the rest of members of the society. Linked to this capacity and as part of the reputation model, the individual has to be able to understand social relations and how they influence reputation and its spreading.

Finally, as we already said, reputation depends on communication so the individual has to be able to communicate with other individuals and the *social space* has to make possible and favour this communication.

Our proposed tentative list of affordances at this level of abstraction is the following:

<u>Individual affordances</u>

– A [complete] model of reputation (including a “reputation oriented” theory of mind).
– Notion of group. Capacity to detect groups. Understanding of social relations.
– Capacity to communicate with other individuals (receive and send messages).

Social space affordances

– Support for group formation and identification.
– Communication channels.
– Messages of different types.

6 Reputation scenario

The third level of concretion in our exercise towards a metamodel for simulation of second order emergence consists in the identification of the affordances needed for the specification of a particular scenario. The scenario that we will use is an idealised environment to study the spread of rumours and the formation of reputation.

The individuals in our scenario are directed by motivations. Each individual has a set of basic needs that tries to satisfy. The set of needs that are relevant for a specific agent determine its personality and the kind of actions the individual is motivated to perform in the world. In our scenario the kind of actions that an individual can perform are actions that influence the reputation of others.

The world where the agents evolve is divided in what we call *social contexts*. A *social context* is a physical space where individuals perform a social activity. For example, your home is a *social context* where you interact with the individuals that belong to your family in domestic activities, the gym is a *social context* where you interact with people that, like you, enjoy practising sport. Each *social context* has different characteristics that facilitates or restricts social interaction.

In our scenario, every turn the individuals are randomly assigned to a *social context*. Once in a *social context*, an individual can move in order to approach or avoid other individuals present in that *social context*. The movement stage is performed in two steps. First, all the individuals express their intention to approach or avoid other members of the group. Second, using these intentions the system calculates the final position of each agent in the *social context* using the following rules:

Given a pair of agents (A, B):

– If one of the two agents has explicitly expressed its intention to avoid the other, the system will put the agents separated in the *social context*.
– If A wants to approach B and B also wants to approach A or B has not expressed any movement related with A, the system will put the agents together.
– If neither A nor B have expressed any movement intention related with the other, the system will randomly decide to put them together.

The set of individuals that after the movement stage are put together, form what we call a *communication group*. Individuals in a *communication group* can exchange messages (*rumours*) and can listen the messages exchanged by the other individuals in that group. As a result of a received or listened *rumour*, an individual can react and send a *support* message (reinforcing what the original *rumour* says) or a *shame* message (expressing his/her disapproval of the original *rumour*). The cycle message-reaction is repeated until all the individuals in that *communication group* have had the opportunity to send a message.

Specifically, the cycle that follows each turn is the following:

1. Each agent is assigned randomly to a social context.
2. For each social context:
   (a) The environment communicates to each agent the list of other agents that will share with it this social context in this turn.
   (b) Each agent evaluates which other agents wants to approach and avoid.
   (c) The environment collects from each agent the list of agents to approach and to avoid.
   (d) With the previous information, the environment allocates each pair of agents in the physical space according to the following rules:
      i. ([approach], [approach|neutral]) : distance 0
      ii. ([approach|neutral|avoid], [avoid]) : distance 1
      iii. ([neutral], [neutral]) : distance random(0,1)
   (e) For each group of agents (agents at distance 0 among them in the physical space):
      i. The environment communicates to each agent the agents that belong to that group (that is, the agents that can send, receive and listen messages to/from it at that moment).
      ii. The environment chooses randomly one agent from those that want to send a *rumour*.
      iii. The other agents send reactions to that *rumour* till no one has anything to say.
      iv. Repeat from 2(e)ii till no agent wants to send a new message.
   (f) Repeat n times from 2b.
3. Repeat from 1

Affordances for the reputation scenario

Our proposed list of affordances at this level of abstraction follows he guidelines established in the previous level (section 5) taking into account the specific scenario described before:

<u>Individual affordances</u>

– Agent architecture directed by motivations with a "reputation oriented" theory of mind.
– Capability to decide which individuals to avoid or to approach (according to the individual's internal motivations and the personality of individuals in the communication group).
– Reasoning mechanisms to decide when to send a {*rumour*||*support*||*shame*} message (according to the individual's internal motivations an the personality of individuals in the communication group).
– Capability to send a {*rumour*||*support*||*shame*} message.

Social space affordances

- Creation of *social contexts*.
- Creation of *communication groups*.
- Make explicit to each agent in a *social context* which are the other members of the society in the same *social context*.
- Make explicit to each agent in a *communication group* the other members of the society in the same *communication group*.
- Enable movement of individuals according to the scenario rules.
- Communication channel between agents that belong to the same communication group.
- Enforce the communication protocol in a *communication group*.

7 Closing remarks

In this paper we abstract a collection of features that are needed for modelling a particular form of reputation emergence. These features provide ways if conceptualising some affordances that the modelling of most second order emergent phenomena would require. We took these affordances and chose to represent them in a way that is reusable in the modelling of other second order emergent phenomena. This way we are taking the first steps towards a conceptual framework —what we call a "metamodel"— that contains a collection of languages, data structures and operations which are readily usable to specify models of second order emergent phenomena.

The experience we acquired in the developing of electronic institutions —from the earliest conceptual versions to its current metamodels and implementation platforms and numerous applications [9]— shows us that the $\mathcal{WIT}$ approach is adequate for the task and also that we should aim to a second order emergent phenomena metamodel that is powerful (capture a large class of second order emergent phenomena), intuitive (so non-experts can use it to simulate second order emergent phenomena) and easy to use (so the process of instantiating and debugging a specification is convenient).

Mirroring our own experience with the use of auctions as an inspiration for electronic institutions, we chose to start with reputation. First, because it is a second order emergent phenomena that is well-known to social scientists and also one with which we already have experience. Second, because, as was the case with auctions, we believe that it contains archetypal second order emergent phenomena features. thus we expect will bring us to a conceptual framework that is generic enough to be applied for modelling of a wide class of second order emergent phenomena, as well as specific enough for second order emergent phenomena, that it is practical for implementing second order emergent phenomena although it might be quite impractical for modelling other social coordination artifacts.

Acknowledgements

This research has been supported by project MILESS (Ministerio de economía y competitividad - TIN2013-45039-P - financed by FEDER) and SCAR project (Ministerio de economía y competitividad - TIN2015-70819-ERC). We also thank the Generalitat de Catalunya (Grant: 2014 SGR 118).

References

1. Huib Aldewereld, Olivier Boissier, Virginia Dignum, Pablo Noriega, and Julian Padget. *Social Coordination Frameworks for Social Technical Systems*. Number 30 in Law, Governance and Technology Series. Springer International Publishing, 2016.
2. Cristiano Castelfranchi. Cognitive architecture and contents for social structures and interactions. In Ron Sun, editor, *Cognition and Multi-Agent Interaction*, pages 355–390. Cambridge University Press, 2006.
3. Cristiano Castelfranchi. *InMind and OutMind; Societal Order Cognition and Self-Organization: The role of MAS.* http://www.slideshare.net/sleeplessgreenideas/castelfranchi-aamas13-v2, May 2013. Invited talk for the IFAAMAS "Influential Paper Award". AAMAS 2013. Saint Paul, Minn. US.
4. J. S. Coleman. *Foundations of Social Theory*. Belknap Press, 1990.
5. Rosaria Conte, Giulia Andrighetto, Marco Campennì, and Mario Paolucci. Emergent and immergent effects in complex social systems. In *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence*, pages 8–11, 2007.
6. Mark d'Inverno, Michael Luck, Pablo Noriega, Juan A. Rodriguez-Aguilar, and Carles Sierra. Communicating open systems. *Artificial Intelligence*, 186(0):38 – 94, 2012.
7. Joshua M. Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, 2006.
8. Maaike Harbers, Karel van den Bosch, and John-Jules Ch. Meyer. Modeling agents with a theory of mind: Theory-theory versus simulation theory. *Web Intelligence and Agent Systems*, 10(3):331–343, 2012.
9. Pablo Noriega and Dave de Jonge. Electronic institutions: The ei/eide framework. In *Social coordination frameworks for social technical systems*, pages 47–76. Springer, 2016.
10. Pablo Noriega, Julian Padget, Harko Verhagen, and Mark d'Inverno. Towards a framework for socio-cognitive technical systems. In *Coordination, Organizations, Institutions, and Norms in Agent Systems X*, volume 9372 of *Lecture Notes in Computer Science*, pages 164–181. Springer International Publishing, Berlin / Heidelberg, 2015.
11. Pablo Noriega, Harko Verhagen, Mark d'Inverno, and Julian Padget. A manifesto for conscientious design of hybrid online social systems. In S. Cranefield, S. Mahmoud, J. Padget, and A.P. Rocha, editors, *Coordination, Organizations, Institutions and Norms in Agent Systems XII*, Lecture Notes in Computer Science. Springer, In press.
12. Isaac Pinyol, Mario Paolucci, Jordi Sabater-Mir, and Rosaria Conte. Beyond accuracy. reputation for partner selection with lies and retaliation. In *Multi-Agent-Based Simulation VIII*, volume 5003, pages 128–140. Springer, 2008.
13. Isaac Pinyol and Jordi Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.
14. DH. Ruben. The existence of social entities. *Philosophical Quarterly*, 32:295–310, 1982.
15. John R. Searle. What is an institution? *Journal of Institutional Economics*, 1(01):1–22, 2005.
16. Ron Sun. Desiderata for cognitive architectures. *Philosophical psychology*, 17(3):341–373, 2004.
17. Harko Verhagen, Pablo Noriega, and Mark d'Inverno. Towards a design framework for controlled hybrid social games. In Harko Verhagen, Pablo Noriega, Tina Balke, and Marina de Vos, editors, *Social Coordination: Principles, Artefacts and Theories (SOCIAL.PATH). AISB 2013 Convention Proceedings*, pages 83–87, 2013.