Matthias Nickles, Michael Rovatsos, Marco Schmitt, Wilfried Brauer, Felix Fischer, Thomas Malsch, Kai Paetow and Gerhard Weiss (2007)

The Empirical Semantics Approach to Communication Structure Learning and Usage: Individualistic Vs. Systemic Views

Journal of Artificial Societies and Social Simulation vol. 10, no. 1
<http://jasss.soc.surrey.ac.uk/10/1/5.html>

For information about citing this article, click here

Received: 20-Jan-2006 Accepted: 05-Aug-2006 Published: 31-Jan-2007

Abstract

In open systems of artificial agents, the meaning of communication in part emerges from ongoing interaction processes. In this paper, we present the empirical semantics approach to inductive derivation of communication semantics that can be used to derive this emergent semantics of communication from observations. The approach comes in two complementary variants: One uses social systems theory, focusing on system expectation structures and global utility maximisation, and the other is based on symbolic interactionism, focusing on the viewpoint and utility maximisation of the individual agent. Both these frameworks make use of the insight that the most general meaning of agent utterances lies in their expectable consequences in terms of observable events, and thus they strongly demarcate themselves from traditional approaches to the semantics and pragmatics of agent communication languages.

Keywords:
Agent Communication, Open Multiagent Systems, Social Systems Theory, Symbolic Interactionism, Pragmatism, Computational Pragmatics

Note: This article is accompanied by a glossary of the most important technical terms used in the context of our approach and Distributed Artificial Intelligence in general. The glossary is available by clicking on the underlined red terms.

Introduction

1.1

Truly autonomous artificial agents whose internal design is more or less opaque to each other cannot share information directly, and, what is more, their behaviour cannot be controlled by peer agents or by a human observer. Rather, the only way to interact with them is by way of symbolic, voluntary *communication*. Thus, a main focus in the development and analysis of *open* multiagent systems (open MASs) composed of autonomous agents that can enter and leave the system at will, is the provision of an adequate agent communication language (ACL) semantics.

1.2

Traditional attempts to modelling the semantics of agent communication languages are mostly based on describing the *mental states* (beliefs, intentions) of communicating agents (e.g., Cohen

and Levesque 1990; Cohen and Levesque 1995; Labrou and Finin 1997; Singh 1993), on publicly visible *social commitments* (e.g.,Guerin and Pitt 2001; Pitt and Mamdani 1999; Singh 2000), or on publicly visible *social attitudes* (Nickles et al 2005b; Gaudou et al 2006; Fischer and Nickles 2006). The theoretical advantages of the first of these approaches are that it is able to make the whole mechanics of utterance generation and utterance understanding transparent, provided that the agents are equipped with the capability to act intelligently in social situations. But it has two obvious shortcomings, which eventually led to the development of the latter "objectivist" approaches based on either social commitments or social attitudes: In open multiagent systems, agents appear more or less as black boxes, which makes it impossible to impose and verify a semantics described in terms of cognitive states, at least in the general case. Furthermore, the description of interaction scenarios in terms of the cognition of individuals tends to become extremely complicated and intractable for large sets of agents, even if one could in fact "look inside the agents' heads". This is not so much caused by the complexity of communication processes themselves, but due to the subjective, limited perspective of the involved individual agents, which makes it hard to construct a concise, comprehensive (and comprehensible) picture of supra-individual processes. Current mentalistic approaches either lack a concept for preventing such complexity at all, or they make simplifying but unrealistic assumptions, for example that all interacting agents are benevolent and sincere.

1.3

Objectivist approaches to semantics, in contrast, yield fully verifiable models, and these approaches achieve a great deal of complexity reduction through limiting the description of a concrete semantics to a relatively small set of normative rules. Therefore, such approaches have marked a big step ahead. However, they tend to oversimplify social processes, and do not offer an adequate concept of meaning dynamics and generalisation. In general, it is questionable whether the predominately normative, static and definite concepts of most of the current agent communication research are really adequate to cope with concepts like agent autonomy, agent opaqueness and the emergence and vagueness of highly complex, dynamic and socially constructed meaning, which communication is in fact about. Therefore, both mentalistic and objectivist views currently fail to recognize that communication semantics evolve during operation of an open multiagent system (MAS), and that they always depend on the view of an observer who is tracking the communicative processes of black-box agents in the system. Yet communication dynamics and observer-dependency are crucial aspects of inter-agent communication, especially in the context of open systems in which a pre-determined semantics cannot be assumed, let alone the compliance of agents' behaviour with it.

1.4

To tackle these issues, the *empirical semantics* approach to the representation (and automated derivation) of agent communication semantics has been developed. This approach comes in two complementary variants, one inspired by *social systems* theory and focusing on supra-individual, so-called *expectation structures* and global utility maximisation (Nickles and Weiss 2003; Nickles et al 2004a; Nickles et al 2005a; Nickles et al 2005c), and one based on *symbolic interactionism*, focusing on the viewpoint and utility maximisation of the individual agent (Rovatsos et al 2003; Rovatsos et al 2004b; Rovatsos 2004).

1.5

In both cases, the empirical semantics approach is based on the assumption that recording and extrapolating observations of message exchange among agents in a multiagent system empirically is the only feasible way to capture the meaning of communication, if no (or little) a priori assumptions about this meaning can be made. Being empirical about meaning naturally implies that the resulting model very much depends on the observer's perspective, and that the semantics would always be the semantics "assigned" to the utterances by that observer, hence this view is inherently constructivist. Since, ultimately, no more can be said about the meaning of a message than that it lies in the expected consequences that this message has, empirical semantics also adopts a "consequentialist" outlook on meaning. Thereby, the "semantics" aspect mentioned above plays a crucial role, because the meaning of agent communication is, from an observers point of view, *entirely* captured by related *expectations* in a system (Lorentzen and Nickles 2002; Nickles et al 2005a). Empirical semantics captures precisely these expectations and how they evolve.

1.6

The three central assumptions underlying our approach are that

- the meaning of communications lies in their *expected* consequences (besides the Luhmannian concept of expectation, this assumption is also related to Peirce's famous *pragmatic maxim*, and Wittgenstein's concept of *language games*, where meaning is associated with practice),

- that these consequences can be learned and anticipated from the observation and as extrapolations of past communication processes (without too much reasoning about what is "inside the agents' heads", which significantly reduces the complexity of the learning task), and that

- the meaning of communications evolves during the interaction processes.

Also a pronouncedly deliberation- and cognition-oriented outlook on agency is adopted, which demarcates itself from approaches that make use of cognitively "light-weight" agents or agents inspired by physics or biology rather than human cognition and high-level sociality.

1.7

Apart from this common ground, the *systemic* and the *individualistic* views of empirical semantics are different in the following aspects:

- While the systemic view focuses on obtaining a "supra-individual picture" of the communication structures within the MAS, obtained by an observer (i.e., the regularities of agent communication processes captured as expectations regarding their continuations (Lorentzen and Nickles 2002)), the individual view is that of a specific agent with preferences regarding different courses of interaction. Although the systemic view can in principle also be endowed with utilities, supra-individual interests likely have a different orientation than those of individual agents, as *inter alia*, a system observer or the system designer might be expected to be interested in normative behaviour or abstract states like equilibria or social welfare, or useful supra-individual mechanisms, rather than in *local* utility maximisation of a single agent.
- The systemic view maintains communication structures of an entire social system of artificial agents, whereby social systems can be so-called *interaction systems* with the agents observing each other as well as large, complex agent societies. The respective data structure used to this end (so-called *expectation networks*) is therefore designed to stochastically model coherences in large-scale systems of communication, rather than local interaction contexts. The obtained empirical semantics can be used e.g. to actively influence the systems in order to maximise global utility (which is not a topic of this paper, cf. Nickles et al 2005a; Nickles and Weiss 2005), or be imparted to the agents in order to provide them a common communication semantics. In contrast, the so-called *interaction frames* used in the individualistic version of our empirical semantics approach are designed for the representation and strategic usage of *local* interaction patterns (Nickles et al 2004b has a detailed comparison of both data structures).
- The systemic view recognizes intentionality, goals and rationality in communication by *imputing* these attitudes to the actors' observed and expected behaviour from an observer's external point of view (the observer could be the MAS designer, or the agent the respective communication was addressed to). In contrast, the individualistic variant of empirical semantics does not make explicit assumptions about rationality as part of the semantics itself. Instead, any reasoning about intentionality and rationality of other agents is part of the cognition of the agent which maintains the empirical semantics. The system perspective does also not make assumptions about the "real" (mental) attitudes or goals of any of the participating agents, but assumes that each communication informs about a certain alleged intention. The idea of communicating intentions is somewhat in line with Grice (1968). But in contrast to his psychologically-inspired approach (based on mentalistic concepts like "intending", "believing", "thinking" etc.), an alleged intention is generated by the process of communication itself in our model, and the alleged rational behaviour of an agent is only imputed by the observer, in the form of communication-generated expectations.

Because this systemic version of empirical semantics recognizes rationality on the level of the supra-individual ACL semantics itself (and not on the level of modeling the mental attitudes of the participating agents), it is called *empirical-rational semantics*. Of course there also exists an empirical semantics that is *not* empirical-rational in our system-view on empirical semantics.

- Whereas our approach to the systemic view proposes a communication model which distinguishes a certain communication (precisely: a certain behaviour which is recognized by an observer as part of a communication) from other kinds of agent action *a priori* by observing initially a difference between "message" (the immediate act of uttering) and (initially possibly unclear and revisable) "information", our individualistic view proposes a communication model in which these two levels are initially indistinguishable, and difference manifests itself only a posteriori in the form of a distinction between anticipated and subsequently observed actual course of interaction.

1.8

In this paper we provide a summary presentation of the empirical semantics approach from both the systemic and the individualistic point of view, and present combined experiments which illustrate how both views can be integrated in implemented systems.

1.9

The remainder of this paper is structured as follows: We begin with an introduction to the sociological foundations that underlie our approach. Section 3 then presents a model for describing empirical communication semantics from the systemic viewpoint. In Section 4, we present the interactionist approach to empirical semantics. Section 5 epitomises on first results from computational experiments conducted with a *combination* of both these frameworks. Section 6 concludes.

Sociological Foundations

2.1

The emergence of social order from intentional or non-intentional results of action or other forms of operations is of vital interest to every social theory. Structure generation, structure preservation, and structural change are among the most central issues in classical and contemporary sociology. In the spirit of the *Socionics* endeavor (Müller et al 1998) which is based on making use of sociological theories in the construction of computer programs two distinguishable social theories have been chosen to use: Niklas Luhmann's theory of social systems (Luhmann 1995) and an elaborate version of symbolic interactionism (with a special focus on the concepts of George Herbert Mead (1934) and Erving Goffman (1974)). Luhmann's approach is assumed to be a macro-perspective, while Mead and Goffman stand for the micro-perspective on sociality. A sociological macro-perspective is typically defined by a focus on societal structures, like the overall distribution of resources or power, long-term processes of social change, the evolution of modes of social differentiation, and large-scale social entities, like social institutions. A sociological micro-perspective, on the other hand, is typically defined by focus on the actions, interpretations and attitudes of individual actors. Both perspectives represent central sociological insights on sociality, from the viewpoints of individuals and from the viewpoint of a social system. The hallmark of our approach can be summarised as taking both perspectives on sociality seriously. So modelling starts from the two directions simultaneously and the question of divergence or convergence is left open for the process to decide. As we pointed out above, the central problem resides in the achievement of social order in systems where multiple autonomous agents interact. Abstractly speaking, there are two strategies to generate social order. The first strategy is to impose rules and the second is to observe regularities. A combination of both strategies seems to be the most promising point to start from and Luhmann's as well as Goffman's and Mead's approaches to social order can be read as such a combination strategy. In summary, the objective of the research described here was *to model and simulate the emergence of social order from two oppositional theoretical perspective through a combination of the mechanisms of rule imposition and pattern observation.*

Communication, expectations, and social systems — the social construction of systems and order according to Luhmann

2.2

According to Niklas Luhmann, social systems are operationally closed on the basis of communication (Luhmann 1995). The term "operationally closed" stems from the concept of *autopoiesis* developed by Humberto R. Maturana (1970) and means that an ongoing flow of communication is the sole constituent of the social system. All patterns, structures, and order are generated out of this flow of communicative events. Starting from this, communication is used as the pivotal concept in Luhmann's sociology. In his view, communication is not about sending messages, nor is it about "doing things with words" (Austin 1962). Actually, he defines a unit of communication as an event that synthesises three meaningful *selections* (in the sense of choices among a set of alternatives): an utterance, an information, and a process of understanding. The unity of these selections is construed *a posteriori* starting from the act of the understanding, which is defined as the process of differentiation between the utterance and the information uttered. To count as a unit of communication, an event has to be observed as the utterance of a piece of information. Furthermore, communication is inherently dynamic, because communication operates in the medium of meaning, and Luhmann defines meaning as the distinction between *actuality* and *potentiality* (i.e. the distinction between what actually occurs and all which might have occurred instead), so that every actualisation ("update") of a specific meaning has to be considered as the actualisation of new potentialities as well (Luhmann 1995). So, social order has to be generated in communication processes, based on unstable units constituted by observations. This kind of order is achieved through expectation formation. Expectations constrain future possibilities according to past actualisation and reduce the overwhelming mass of possible future worlds to some smaller set of expected paths. Expectations only affect the present operations, they do not determine the course of future events and they do not reach into the past, but they direct attention to probable outcomes. Every expectation can be disappointed. Disappointment is a common feature of expectation-building and almost always demands a reaction, may it be learning (changing the expectation according to the disappointment) or stubbornness (holding on to the disappointed expectation). Expectations provide structural ordering and a certain degree of openness to change at the same time, because they connect past and future in an actual communicative operation. In communication, expectations are triggered by marks or hints, leading from names of persons, role names, descriptors of standing procedures, or a whole variety of other abstract terms to sometimes vast networks of past usages of these marks in communication. To conclude, social systems are ordered through expectation formation by communicative observation of the course of communicative events. Social order is generated through further communication that is based on observation of communication. Resting on this assumption, observers are needed to build up a meaningful social structure. These observers may be single actors or agents in communicating their expectations in interaction processes or some "macro-social" entity like a social system with its codes and programs[1]. Luhmann is mostly concerned with the system perspective and how it constitutes an assumed point of observation which explicitly (and quasi-intentionally) includes or excludes certain events it observes, thereby constructing a system-level structure.

Social reasoning, frames, and framing — the social construction of minds and situations according to Mead and Goffman

2.3

Symbolic interactionism is another perspective that bases social order on expectations and meaning, quite similar to Luhmann in that respect. George Herbert Mead relates the emergence of mind, self, and society to forms of communicative behaviour and describes the basic attribution of meaning to gestures as the starting point for this process (Mead 1934) . Essentially, communicative behaviour generates expectations that structure the subjective view on social situations and what others expect from oneself. To imagine the expectations of others is the foundation for every social behaviour, for the construction of the self, and for the emergence of society as an ordered and meaningful interaction space. According to Mead, the social actor consists of several components: the "me" represents the actual imagination of these expectations, while the "self" is constituted from the sum of multiple different "me's" as a more or less consistent identity integrating a lot of expectations for different contexts together. To summarise, the central insight of Mead's approach lies in the observation that the incorporation of the expectations of others is the fundamental base for social cognition. A socially intelligent

agent has to account for the expectations of others and has to learn continually from their reactions to her behaviour. Evaluating this precondition of social intelligence is the first of two promising concepts stemming from interactionism. The second promising concept to build social order from an agents perspective is Erving Goffman's exploration of social frames and processes of framing (Goffman 1974). Basically, a frame is principle of organizing events, a condensation of cognitive patterns of meaning and interrelated expectations. These interrelated expectations can refer to appropriate behaviour, possible courses of events, which roles to take, and the spatial and temporal setting of the social situation. All these expectations were stereotyped and condensed into a single frame with varying degrees of freedom for individual actions. A frame (and most of the interrelated expectations) is triggered by small hints the agent may take from the environmental setting, from the behaviour of other actors or from semantic codes used in the conversation. Framing processes start from these small hints and make use of the degree of performative freedom the frame has to offer. On the one hand, framing means building or activating a frame from all the little indicators a situation confronts the agent with. On the other hand, framing is an activity carried out by agents who actively shape and transform a frame to cope with the arising situation and make sense of it. From the perspective of the single agent, social order is condensed into frames and framing is the activity to build up social order in a single situation and to manipulate that order individually to some degree.

Interactionism vs. social systems theory? — individualistic vs. systemic views

2.4

Systems theory and interactionism represent different sociological paradigms and sometimes they are seen as essentially opposing approaches with totally divergent interpretations of the social world. Interactionism is considered as the the prototypical micro-sociological approach, whereas systems theory, in the line of Talcott Parsons (1951), is widely seen as prototypical macro-social. As we have shown above, the differences between the approaches are exaggerated in this interpretation. Just to the opposite, both theories share some striking resemblances. The most fundamental one is the centrality of the expectation concept. Expectations are the form of social order from an agents perspective as well as from a systems perspective and both perspectives, coming from different directions, can converge on the establishment of social order through nets of interwoven expectations. The two perspectives start from different points of departure but complement each other in their focus on expectations. The social reasoning of agents and the emergent order of communication processes work together in the establishment of social structure and combining this views should facilitate the modelling of social order as well as our understanding of the structuring of social life.

The Systemic View: Agent Interaction Systems and Expectation Networks

3.1

Following Luhmann, our systemic view assumes that social structures (i.e., the structures of a social system) are *expectation structures*, consisting of behavioural *expectations* (informally explained in the previous section). In that, behaviour is seen as an observable simplification of communication events (in the form of utterances) as well as non-symbolic agent acting and other events. Empirical semantics makes use of this view by defining the semantics of agent communication ("semantics" in the computational sense) in terms of expectations which are updated each time new utterances are observed. Such expectations can in principle refer to all events that communication can refer to - most importantly the interactions among agents.

Overview

3.2

In the following, we outline our general model of communication semantics from the systemic perspective, and then, in the following subsection, provide the most important aspects of the technical realization. For lack of space, and since this article is intended as an overview, please refer to Nickles et al 2004a; Nickles et al 2005a for technical details.

Expressiveness of utterances

3.3

Empirical semantics (in the systemic variant) allows the uttering agent to make his utterance in the form of a so-called *projection*, corresponding to the Luhmannian concept of "information". A *projection* constitutes the content of a communication, i.e., what the communication requests, asserts etc. But in contrast to traditional ACL semantics, *projections* are not encoded by means of some logic and ontology. Instead, they *select a desired part of the current empirical semantics*. Utterances in form of *projections* allow agents to state ostensibly desired future states of the communication system, with desired non-symbolic events (like "closing a window") as special cases. A "state" here denotes any situation in which some event happened in a certain context of previous events. E.g., projecting "The window will be closed (by someone)" corresponds to a request to close the window, with the window being closed as the desired state (or set of alternative states in which the window is closed). This schema also works with assertions: To assert that the window is already closed, an agent would project that the addressee acts (communicates) from now on in consistency with the information that the window is closed, i.e. *as if* states in which this information is true had actually been realised. Thus, assertions (with the alleged intent to convince someone) become in our model essentially requests to act (communicate) *as if* the addressee is convinced by the assertion. Analogously, assertions without the intent to convince could in our framework be encoded as requests to act (communicate) in conformance with the information that the uttering agent holds the asserted opinion.

3.4

In all cases, the meaning of the projected states itself is emergent from communication. If, e.g., someone utters "Pay me the money", the meaning of "paying money" is not defined using some static "payment-ontology" as with traditional agent communication languages, and it is also not to be found somewhere in the uttering agent's mind, but is part of the empirical semantics itself. As empirical semantics evolves (i.e., is continually revised with each newly observed communication act), a *projection* refers to the current version of the empirical semantics - in the example the meaning of "pay money" at the time of the respective utterance. But since with this and subsequent utterances the empirical(-rational) semantics is revised, the meaning of the *projection* may change dynamically after the event, in line with the Luhmannian concept of communication which defines itself recursively.

Empirical semantics

3.5

As we have seen, a single utterance can be seen as a request to act towards a specified ("projected") state. The empirical semantics of such a request is defined as the probability distribution of events which are caused by this request. Essentially, we thus define the meaning of utterances in terms of their pragmatics. But our approach differs in certain aspects from the most influential pragmatic theory, namely *speech act theory*. A short comparison of our model with speech act theory can be found here. Note that most traditional ACL semantics are based (or claim to be based) on speech act theory.

3.6

Since social systems theory defines the structures of a social system as expectation structures, the empirical semantics of utterances is part of the social structures of the system. Example: If an agent utters "Close the door" to another agent, the desired state is the door being closed by the addressed agent, and the empirical semantics of this utterance is the way in which subsequent communications are expected to respond to this request: By, e.g., accepting, complying or denying, as well as by the way the speaker himself promotes communicatively the (seemingly) desired effect in further communications, e.g., by means of argumentation or by threatening the addressed agent with sanctions in case of possible non-compliance. Such pragmatics can also include non-symbolic "physical" events and indirect effects like the empirical semantics of further communications. Empirical semantics aims to identify generalised, reusable patterns of such effects, obtained from empirical observations and, optionally, a small set of *a priori* assumptions regarding the rationality of agent communication (see below).

3.7

As another example, assume an agent performs the act of nomination "You are the group leader now" ("nomination" denoting a performative act in terms of speech act theory, which is admittedly not completely adequate in the context of social systems theory). This act demands from the participating agents to act in the future in conformance with the claim that the nominated agent is a group leader from now on. With our semantics in terms of expectable consequences, even the success of such a performative act becomes visible only *a posteriori* (naturally, neither the performer of the act nor an observer can be fully sure about its success at the time of the utterance, given a MAS with truly autonomous agents), but if the nominating agent has been assigned the necessary social power in the past, the success probability can be learned from, e.g., past successes.
The expectations regarding the consequences of utterances in terms of subsequent events is the empirical semantics of these utterances, and is calculated from observed event trajectories, optionally taking into account assumptions that the observed communication process is rational and reliable in a *minimal sense* in order to be functional, irrespective of the hidden mental intentions and beliefs of the agents.

Empirical-rational semantics

3.8

Our approach is not just based on a plain stochastic extrapolation of observed event sequences, but can optionally take into account the ostensibly rational "social image" of artificial agents. If this is taken into consideration, the systemic variant of empirical semantic is called *empirical-rational semantics*.

3.9

Although our approach is strongly influenced by social systems theory, we also introduce rationality and even rational choice into our model - but not at the cognitive level of individual agents. Instead, we use the term rationality to denote a way of reduction of communication contingency (and thus of complexity), assuming that communications within a MAS are interrelated to some degree in a rational way which steer the *acceptance* or *rejection* of communications (respectively the acceptance or rejection of their informational content). Concretely, we assume that i) communications attributed to the same person do not contradict each other within some emergent spheres (the so-called *spheres of communication*), and ii) that communications attributed to the same person support each other's respective *projections*, again within certain emergent spheres of reliability. i) is related to the assumption of agent sincerity and trustability, but allows for the determination and revision of the spheres of reliability at run-time. The assumption behind i) is that although communications do not need to be sincere (let alone trustable), it is rational to adhere even to insincere communicated opinions and desires at least for some time. As for ii), since an utterance has no (significant) direct impact on the physical environment, its physical consequences are achieved socially and indirectly, and, most importantly, an autonomous addressee is in principle free to deny the communicated proposition. Since in our model an utterance is from the viewpoint of an observer *seemingly* generated by a self-interested person in order to influence an addressee who is not already convinced (otherwise the communication would not be necessary), it needs to be accompanied with communicated reasons given to the addressee in order to increase the probability of an acceptance of the communicated content. This can be either done explicitly by previous or subsequent communications (especially *reciprocally*: "If you comply, I'll comply too", or by a threatening with sanctions, or by means of argumentation), or implicitly by means of a generalisation over past events (e.g., trust) or given social structures (like norms which steer the acceptance of requests). The whole of the expectations which are triggered by a communication in the context of the preceding communication process is called the *rational hull* of this communication act[2].

Framework

3.10

The basic requirements in order to put the empirical semantics approach into practice are:

- The presence of a so-called *semantics observer* who derives the communication semantics from empirical observations. This observer, which can be seen as a meta-agent who overhears the multiagent system, does so by recording agent communications and calculating expectation structures. If the observer has a complete view of all communication processes in the MAS, it represents in a certain sense the social system as a whole. But the observer could likewise be one of the interacting agents, overhearing only parts of the message traffic.
  The *semantics observer* learns communication structures and thus the empirical semantics mainly by assuming that past experiences will to some degree repeat themselves in the future (the so-called *stationarity assumption*). If he additionally makes assumptions about the alleged rationality of communications, we speak about an *empirical-rational semantics* instead of a "plain" empirical semantics.
- A data structure for the computational representation of expectation structures (recall that the empirical and empirical-rational semantics is part of these structures). We use so-called *expectation networks* (also referred to as *expectation graphs*) for this purpose. Expectations can also be formalized using mental attitudes of agents (namely beliefs and intentions) (Nickles et al 2005a), allowing for the use of traditional reasoning techniques on empirical semantics.
- A model for the dynamics of expectations structures, i.e., of the way expectations are created, updated and revised during communication processes.

Expectation networks

3.11

Expectation networks (ENs) are the data structure we want to use for the computational representation of social structures and the empirical semantics of communications. Starting with a bootstrapping EN, our EN evolves during the observed course of communications (respectively, communications acts). The formal EN definition presented here is a simplified version of the older definition presented in Nickles et al (2004b). The current definition of "full" ENs (which need, other than the variant presented in this work, not to be trees in general) can be found in Nickles et al (2005a). In addition to the simplified version used in this article, full ENs also allow for the representation of *adaptive social norms* (in form of normative expectations), variables and generalisations like *agent roles*.
A formalization of expectation structures based on the BDI framework (i.e., a logical framework for the modeling of beliefs, desires and intentions of agents) can be found in Nickles et al (2005a).

3.12

The central assumption that is made in ENs is that observed events like agent actions (especially symbolic agent messages) may be categorised as expected continuations of other observed event sequences. Nodes in an EN thus correspond to events (and for simplicity, a node is sometimes referred to as its corresponding event), and the path leading to a node corresponds to a sequence of preceding events, i.e., the *context* of the event. Since utterances can be highly complex, they are decomposed into several events of type *elementary communication act* (ECA).

3.13

An edge leading from event m to event m' is thought to reflect the probability that m' follows m from the observer's point of view.

$$
\begin{aligned}
Agent &\rightarrow agent\_1 \mid \ldots \mid agent\_n \\
PhysicalAction &\rightarrow move\_object \mid touch\_agent \mid \ldots \\
Action &\rightarrow ECA(Agent, Projection) \\
&\quad \mid do(Agent, PhysicalAction) \\
ActionPattern &\rightarrow Action \mid ? \\
Projections &\rightarrow (Conditions, GoalStates) \\
Conditions &\rightarrow SimplePath \\
GoalStates &\rightarrow SimplePath \\
SimplePath &\rightarrow Action{:}SimplePath \mid \varepsilon
\end{aligned}
$$

Table 1. A grammar for event nodes of ENs, generating the language $\mathcal{M}$ (the language of concrete actions, starting with $Action$).

3.14

We use a formal language $\mathcal{M}$ to define the events used for labeling nodes in *expectation networks*. Its syntax is given by the grammar in table 1. Agent actions observed in the system can be either "physical" `(non-symbolic) actions of the format (a, ac) where a is the executing agent, and ac is an domain-dependent symbol used for a physical action, or symbolic *elementary communication act* $ECA(a, c)$ sent from a with content c. We do not talk about "utterances" or "messages" here, because a single utterance might need to be decomposed into multiple ECAs. The symbols used in the $Agent$ and $PhysicalAction$ rules might be domain-dependent symbols the existence of which we take for granted. For convenience, $agent(eca)$ shall retrieve the acting agent of an ECA *eca*.

In addition to normal node labels, we use the symbol $(\triangleright_{EN})$ to denote the root node of an specific EN. The content c of a non-physical action is given by type $Projections$. The meaning of $Projections$ will be described later.

Syntactically, *expectation networks* are here represented as lists of edges (m, p, n) where m and n are actions, and p, $p \in [0; 1]$ is a transition probability (*expectability*) from m to n. C is the set of all edges, V the set of all nodes in the EN. We use functions $in : V \rightarrow 2^C$, $out : V \rightarrow 2^C$, $source : C \rightarrow V$ and $target : C \rightarrow V$ which return the ingoing and outgoing edges of a node and the source and target node of an edge, respectively. $children : V \rightarrow 2^V$ returns the set of children of a node, with $children(v) = \emptyset$ in case v is a leaf. Edges denote correlations in observed communication sequences. Each edge is associated with an expectability (returned by $Expect : C \rightarrow [0; 1]$) which reflects the probability of $target(e)$ occurring after $source(e)$ in the same communicative context (i.e. in spatial proximity, between the same agents, etc.).

Sometimes we denote a path p in an EN leading from $v_0 \in V$ to $v_n \in V$ as concatenations of message labels (corresponding to ECAs) $Label(v_0) \sqcup \ldots \sqcup Label(v_n)$. The $\sqcup$ symbols are sometimes omitted for brevity and we denote the length of a sequence by $|p| := n$.

$Node : SimplePath_{\mathcal{EN}} \rightarrow V$ results in the last node of a certain path given as a string of labels. Nodes or corresponding messages along a path p will be denoted as p_i. $\mathcal{EN}(\mathcal{M})$ is the set of all possible *expectation networks* over $\mathcal{M}$.

Definition 1. An *Expectation Network* is a structure $EN = (V, C, \mathcal{M}, Label, Expect) \in \mathcal{EN}(\mathcal{M})$ where

- V with $|V| > 1$ is the set of nodes,

- $C \subseteq V \times V$ are the edges of EN. (V, C) is a tree called *expectation tree*. (V, C) shall have a unique root node called $\triangleright_{EN} \in V$ which corresponds to the first ever observed action. The following condition should hold: $\forall v \sum_{e \in out(v)} Expect(e) = 1$
- $\mathcal{M}$ is the *action language.* As defined in table 1, actions can be symbolic ($ECA(\ldots)$) or physical actions ($do(\ldots)$). While we take the existence and the meaning of the latter in terms of resulting observer expectations for granted and assume it is domain-dependent, the former will be described in detail later. Physical actions could be assigned an empirical semantics also (being their expected consequences in terms of subsequent events).
- $Label : V \rightarrow \mathcal{M}$ is the *action label* function for nodes, with

  $\forall v \in V : \forall e, f \in children(v) : \neg unify(Label(e), Label(f))$

  (where $unify$ shall be $true$ iff its arguments are syntactically unifiable. (Cf. Nickles et al 2005 for the use of variables in ENs),
- $Expect : C \rightarrow [0; 1]$ returns the edges' expectabilities. For convenience, we define $Expect(label|path) = Expect(in(v))Node(path \sqcup label) = v$.

Paths starting with $\triangleright$ are called *states* (of the communication process)[3].

Modeling social interaction

3.15

Based on the definition of ENs, we can now define *social interaction structures* as a special kind of expectation structures. Social interaction structures capture the regularities of communication processes in an interaction system. They also include other observable events, like non-communicative activity, as long as these events are observable for all participating agents.

3.16

Social interaction structures capture the insights that i) agent sociality emerges from agent communication, and that ii) communication events/actions form a so-called *social system* which is closed in the sense that, to some degree, communication regularities come into being from communications themselves (in the context of their environment) (Luhmann 1995), such that the *semantics observer* does not need to have to "look inside the agents' heads" to derive these structures. Because of that, communication structures can meaningfully be learned from observations. Nevertheless, this learning process needs to be continuously repeated to adapt the EN to new perceptions (since open systems with truly autonomous agents with unknown life spans might not have a final state), and always implies the possibility of failure of its prediction task (hence the term "expectation"). The social interaction structures triggered by a certain utterance within a certain context of preceding communications is called the empirical *semantics* of this utterance (Lorentzen and Nickles 2002). Technically, this semantics is given as a sub-tree of the EN (the EN which represents the current expectation structures). Within the EN, this sub-tree starts with the node which corresponds to the utterance[4]. The path leading to the sub-tree's root corresponds to the context of the utterance.
Since in social systems theory interaction systems have the distinguishing property that the participants are co-present (i.e., the communication flow is public), we can safely assume that social interaction structures represent the *common ground* (shared knowledge) of the discourses in the respective social system.

Social interaction systems

3.17

The way social interaction structures are actually processed is captured in our model of *social interaction systems* (SIS). They capture the current expectation structures and changes to these expectations obtained from new perceptions. The latter are technically represented by an expectation update function which maps an older EN to another, more topical EN after a new message event has been observed. The *semantics observer* maintains a model of the SIS, and in the following, we refer to this model if we speak about the SIS.

Definition 2. A *(Social) Interaction System* at time t is a structure $SIS_t = (\mathcal{M}, f, \varpi_t, \rho)$ where

- $\mathcal{M}$ is the formal language used for agent actions (according to table 1),
- $f : \mathcal{EN}(\mathcal{M}) \times \mathcal{M} \to \mathcal{EN}(\mathcal{M})$ is the *expectations update function* that transforms any expectation network EN to a new network upon experience of an action $m \in \mathcal{M}$. $f(\perp, m)$ returns the so-called *initial EN*, transformed by the observation of m. This initial EN can be used for the pre-structuring of the social system using given e.g. social norms or other *a priori* knowledge which can not be learned using f. Any ENs resulting from an application of f are called *Social Interaction Structures*.
  As a non-incremental variant we define $f : \mathcal{M}^+ \to \mathcal{EN}(\mathcal{M})$ to be
  $f(m_0 \sqcup m_1 \ldots \sqcup m_t) = f(\ldots(f(f(\perp, m_0), m_1)\ldots), m_t)$,
- $\varpi_t = m_0 \sqcup m_1 \ldots \sqcup m_t \in \mathcal{M}^*$ is the list of all actions observed until time t. The subindices of the m_i impose a linear order on the actions corresponding to the times they have been observed[5],
- $\rho \in \mathbb{N}$ is a duration greater of equal to the expected life span of the SIS. We require this to calculate the so-called *spheres of communication* (see below). If the life time is unknown, we set $\rho = \infty$. Although a sphere of communication denotes the ultimate boundaries of trustability for a single communication, even with $\rho = \infty$ initially certain limits of trustability and sincerity become visible in empirical semantics by means of the extrapolation of interaction sequences. Suppose, e.g., a certain agent x takes opinion a in discourses with agent y, but opinion $\neg a$ in all interactions with agent z. Since this "opinion switching" shows regularities, our algorithm will reveal it.

We refer to events and EN nodes as *past*, *current* or *future* depending on their timely position (or the timely position of their corresponding node, respectively) before, at or after t. We refer to $EN_t = f(\varpi_t)$ as the *current EN* from the semantics observer's point of view, if the semantics observer has observed exactly the sequence $m_0 m_1 \ldots m_t$ of events so far.
The intuition behind our definition of SIS_t is that a social interaction system can be characterised by how it would update an existing expectation network upon newly observed actions $m \in \mathcal{M}$. The EN within SIS_t can thus be computed through the sequential application of the structures update function f for each action within ϖ, starting with a given expectation network which models the observers' *a priori* knowledge. ϖ_{t-1} is called the *context* (or *precondition*) of the action observed at time t.
To simplify the following formalism, we demand that an EN ought to be implicitly complete, i.e., to contain *all* possible paths, representing all possible event sequences (thus the EN within an interaction system is always infinite and represents all possible world states, even extremely unlikely ones). If the semantics observer has no *a priori* knowledge about a certain branch, we assume this branch to represent uniform distribution and thus a very low probability for every future decision alternative ($\frac{1}{|\mathcal{M}|}$), if the action language is not trivially small.
Note that any part of an EN of an SIS does describe exactly one time period, i.e., each node

within the respective EN corresponds to exactly one moment on the time scale in the past or the future of observation or prediction, respectively, whereas this is not necessarily true of ENs in general. For simplicity, and to express the definiteness of the past, we will define the update function f such that the *a posteriori* expectabilities of past events (i.e., observations) become 1 (admittedly leading to problems if the past is unknown or contested, or we would like to allow contested assertive ECAs *about* the past). There shall be exactly one path pc in the current EN leading from start node $\triangleright_{en}$ leading to a node pc_t such that $|pc| = t$ and $\forall i, 0 \leq i \leq t : Label(pc_i) = m_i$. The node pc_i and the ECA m_i are said to *correspond* to each other.

3.18

Building on this formal< definition, the *empirical semantics* of a sequence of messages (and other events) at time t ϖ_t (respectively, the empirical semantic of the message m_t within a context of preceeding events ϖ_{t-1}) is formally defined as the probability distribution $\Delta_{EN,\varpi}$ represented by the EN sub-tree starting with the node within EN_t that corresponds to ϖ_t:

$$\Delta_{EN,\varpi}(w') = \frac{\prod_{i,1 \leq i \leq |w'|} Expect(w'_i | \varpi_t w'_1 \ldots w'_{i-1})}{\sum_{m \in M^+} \prod_{i,1 \leq i \leq |m|} Expect(m_i | \varpi_t m_1 \ldots m_{i-1})}$$

for all $w' :\Leftrightarrow \varpi_t \sqcup w' \in M^+$. The w'_i denote single event labels along w', i.e., $w' = w'_1 \sqcup w'_2 \sqcup \ldots$ (for m analogously).

Projections

3.19

As defined in table 1, ECAs consist of two parts: The uttering agent, represented as an agent identifier, and the ECA content in the form of *projections*. Each *projection* is a set of EN node pairs which are derived from the following two syntactical elements (cf. table 1)[6]. We describe *projections* and thus ECAs only informally here for lack of space. Refer to Nickles et al (2004a) for a formal definition. Note that the *projections* of ECAs correspond somewhat to the linguistic meaning of the term "semantics".

- *Conditions* chooses, using an EN path (without expectabilities), a possibly infinite set of EN states which have to become reality in order to make the uttering agent start to act towards its uttered goal (e.g. in "*If I deliver the goods*, you must pay me the money"). As shown in table 1, conditions are given as a linear list of node labels. This path must match with paths in the current EN, either beginning with $\triangleright$, or starting at nodes after the node which corresponds to the ECA. The end nodes of all matches in the EN are called the *condition nodes* of the ECA *projections*. So, if the node list is empty, the only condition node is the node corresponding to the ECA. Path matching is always successful, since in our model, an EN implicitly contains all possible paths, although with a probability close to zero for most of them.

- *GoalStates* chooses, using an EN path (without expectabilities), the (possibly infinite) set of states of the *expectation network* the uttering agent is expected to strive for. The uttered *GoalStates* path must match with a set of paths within the EN such that the last node of each match is a node of an EN branch that has a condition node from *Conditions* as its root. Both in *Conditions* and *GoalStates* paths, wildcards "?" for single actions are allowed.

Rational hulls

3.20

We consider it to be reasonable that communications subsequent to a certain communication by the same person are consistent with this communication and support the reaching of the projected goals of this communication *at least for some significant amount of time*[7]. This span of a communication in terms of consistent subsequent events is called *sphere of communication* (cf.

Figure 1). Such a sphere ends as soon as the respective agent contradicts himself or stops trying to achieve his projected goals. Theoretically, each ECA could have its own *sphere of communication*. For simplicity, we assume that the initial *sphere of communication* of any ECA eca is simply $p - time(eca)$, where the first operand is the expected time of the last observed utterance within the SIS, and the second is the utterance time of the projecting ECA. Independent of this value, the actual *spheres of communication* are implicitly evolving during communication.

3.21

As stated before, the actions expected to be performed within the respective *sphere of communication* in order to make *projections* come true (i.e., assuming intentions which are not necessarily honest) is called the *rational hull* of the ECA. Thus, the determination of the rational hulls of observed ECAs plays a crucial role in determining the empirical semantics of agent communication languages (ACLs). The expected agent actions determined by such limited and revisable rational hulls can be seen as the actual pragmatics and meaning "behind" the more normative and idealistic concept of social commitments, and *spheres of communication* can be seen as fine-grained utterance-level models of trust (Nickles et al 2005b; Fischer and Nickles 2006).

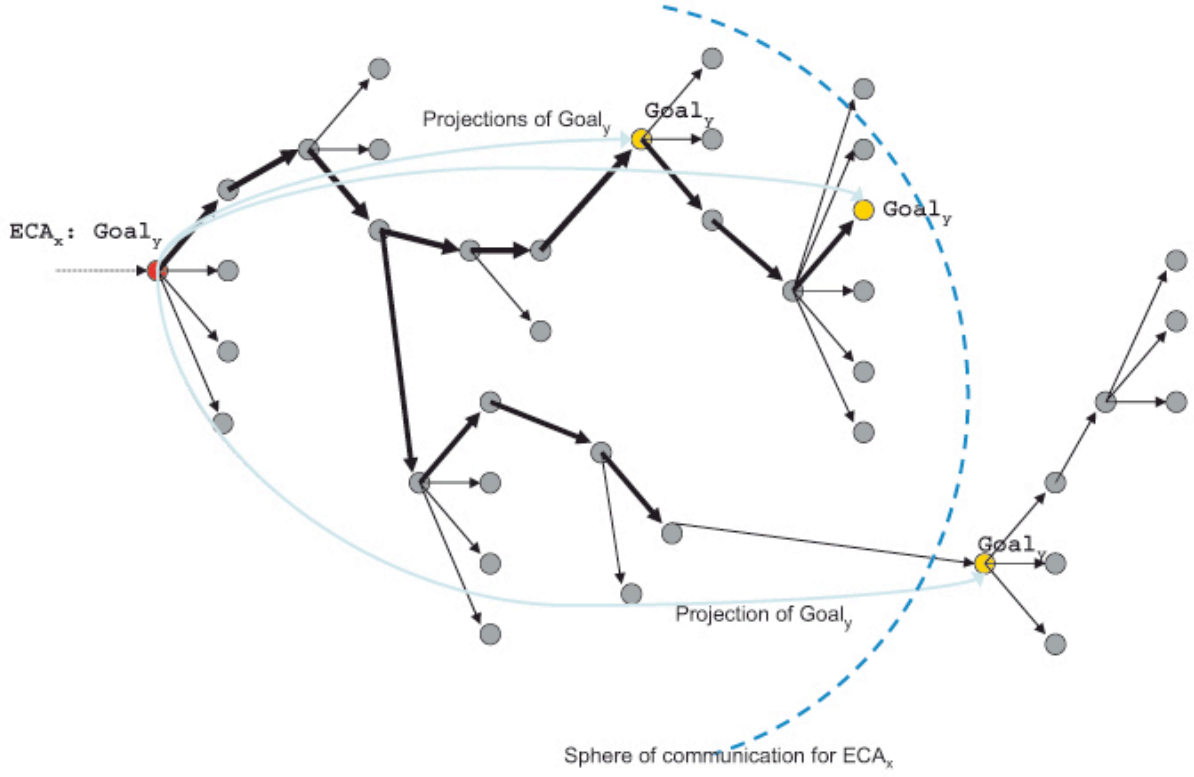

Figure 1. An EN with *projections* and a *sphere* of communication

3.22

We assume the manifestation of the following attitudes by means of ECAs *within the respective spheres of communication* and contextualised by means of other ECAs. They can be seen as "public intentional stances" (so-called *ostensible mental attitudes*) of the uttering agents, and restrict what can be expected about rational hulls. As a simple example, the attitude expressed by $assert(a)$ would be inconsistent with the attitude of $assert(\neg a)$, and is thus very unlikely *given overlapping spheres of communication* (but of course it is not so unlikely a self-interested agent utters inconsistent ECAs in different contexts, e.g. facing different communication partners).

- *Information of other agents about desired states of communication.* This information is given by *projections* as described above.

- *Mutual support and consistency among multiple ECAs.* The supporting functionality communication has regarding other communications by the same agent is defined by the rational hulls of the supported elementary communication acts, which will become implicitly more expectable too if supporting rational hulls increase their own expectabilities (i.e., every ECA supports other ECAs of the same agent or is neutral in this regard). In this, the mutual consistency of multiple assertive ECAs is a special case of this mutual support.
- *Manifestation of understanding.* In case agents "understand" each other, ECAs cannot express contradiction to the fact that other ECAs pursue the two previous intentions (i.e., Agent 1 does not need to believe Agent 2 is right, but she needs to believe at least that Agent 1 *wants* to be right in a specific case). We do not consider misunderstanding in this work.

Capturing these intentions, and given the set of *projections* for each ECA eca `uttered by an agent a, we calculate the semantics of ECAs using the following two principles.

Alleged rationality ascribed by the observer

3.23

After uttering eca, an agent a is expected to choose an action policy such that, *within the respective sphere of communication*, his actions maximise the probability of the projected state(s). Again, it is important to see that this expectable rational behaviour does not need to reflect the "true", hidden intentions of the observed agents, but is an external ascription made by the observer instead. Let $p \in projections(eca, EN_t)$ be a *projection*. Then, considering that p would be a useful state for the uttering agent to be in, the rule of rational choice proposes that for every node v_d with $agent(v_d) = a$ along the path $v_t \ldots p$ leading from the current node v_t to p, $Expect(in(v_d)) = 1$ for the incoming edge of v_d, and that the expectabilities of the remaining outgoing edges of the predecessor of v_d are reduced to 0 appropriately (if no other goals have to be considered).
Again, for lack of space we cannot give the precise formalism for imposing these rules here (cf. Nickles et al 2004a).

3.24

Figure 1 shows an EN modelling the future of some communication process. ECA_X is an utterance which encodes $Goal_Y$. This goal itself stands for several (seemingly) desirable states of the EN (yellow nodes). Since within the *sphere of communication* of ECA_X it is expected that the uttering person rationally strives for these states, certain EN paths leading to these states become more likely (bold edges) because the actions along these EN paths are followed more likely than others by subsequent communications. Such paths need to be (more or less) rational in terms of their expected "utility" (e.g., in comparison with competing goal states), and they need to reflect experiences from analogous behaviour in the past. Not much could be said about the "true utilities" the agents assign internally to EN states, nevertheless.

Empirical stationarity assumption and rationality-biased empirics

3.25

The empirical learning of future behavioural trajectories as a kind of observable desideratum of a communication system provides the basis of empirical semantics that can be used to anticipate how agents will act and react to others' actions. However, if we were to use the rule of rational choice without existing empirical data, we would encounter at least three problems: 1) Predicting actions according to the rule of rational choice requires some given evidence about subsequent actions of other agents. In case this previous evidence is missing, the rule of rational choice would just "convert" one uniform distribution to another uniform distribution. Therefore, we have to provide an initial probability distribution that the rule of rational choice can be applied to[8]. 2) the set of *projections* for a single ECA might be infinite. Most of the expectabilities along the paths leading from the current node to these EN branches sum up to very low probabilities for the respective *projection*. Thus, a pre-selection of likely paths will be

necessary. And by far most important 3), the rule of rational choice does not consider individual behavioural characteristics like (initially opaque) goal preferences, insincerity and trustability, but treats all *projections* uniformly. Such information needs thus to be obtained from past agent practice, as well as individual strategies towards the *projections* and limits of trustability and sincerity. For these reasons, we combine the application of the rule of rational choice with the assumption of some stationarity of past event trajectories, i.e., the assumption that previously observed action sequences repeat themselves in the future in a similar context. We use this assumption to retrieve a probability distribution the rule of rational choice can be applied to and weighted with subsequently.

3.26

In order to learn EN stationarity from previous observations, we follow the so-called *variable-memory approach* to higher-order Markov chains using *Probabilistic Suffix Automata* introduced for *L-predictable* observation sequences (Ron et al 1996), in form of *Prediction Suffix Trees* (PST). This approach efficiently models Markov chains of order L (i.e., with a model memory size of L), allowing for rich stochastical models of observed sequences. The applicability of this approach to our scenario is based on the heuristical assumption that many Social Interaction Systems are *short-memory systems*, which allow the empirical prediction of social behaviour from a relatively short perceived event sequence.

3.27

Putting together the rule of rational choice and the assumption of empirical stationarity, a definition for the SIS update function is gained (cf. Nickles et al 2004a for details). It applies the results of the calculation of rational hulls to the entire "raw" EN resulting from the PST by means of a recursive top-down tree traversal which is limited by a maximum search depth. In that, each expectability obtained from the PST is weighted with the corresponding "utility" obtained from the temporarily assumed intentional stance of the respective observed agent. This step finally yields the updated EN, representing the empirical semantics. The updating (or the complete re-generation of the EN from all events observed so far, if no incremental learning algorithm is provided as in Nickles et al (2004a)) has to be repeated for newly observed events (*interaction system evolution*) until the communication process (i.e., the interaction system) comes to its end.

The Individualistic View: Individual Agents and Interaction Frames

4.1

From the standpoint of individual agents, the problem of learning and applying empirical expectation structures presents itself quite differently from the way we have described it so far in our description of system-level communication structure learning and usage. Agents (at least the knowledge-based, deliberative agents that operate on explicit, symbolic representations of the world as the ones we discuss here) are *intentional systems*, and thus any activity — including reasoning about empirical semantics — must be embedded within their more general, goal-oriented deliberation and means-ends reasoning processes. Therefore, at the agent level the problem of deriving and using empirical models of social interaction structures becomes a problem of *strategic reasoning about interaction*. Capturing the empirical semantics of communication in a system is no more an end in itself (or a mere step of pre-processing resulting in an expectation network that is later used for various purposes by either the human designer or some computational system observer), it is intertwined with strategic agent decision making.

4.2

This has various implications on how the empirical semantics approach has to be used from the perspective of an individual agent. Firstly, quite differently from the "birds-eye" view of a global system observer, individuals do not avail of any information about interaction processes other than those they have witnessed themselves or been explicitly given information about by others who have done so. Secondly, even within this restricted amount of empirical data, it is in the best interest of agents to focus their (boundedly rational) reasoning and learning activities on those interaction processes that matter to them. Thirdly, agents' actual participation in interaction processes implies that they decompose a more complex flow of observed interaction into manageable chunks of interaction they can reason about efficiently so as to be able to

generate a timely response. In other words, agents situated in a co-inhabited environment that demands a certain responsiveness of them do not have the luxury of performing computationally expensive update operations on complex models of communicative expectations. Finally, any prediction made for the purposes of making optimal communication and action decisions in the context of an ongoing interaction process must be explicitly combined with the agent's sub-social rational behaviour. At the end of the day, interaction with other agents is only useful for the agent to the extent that it can aid in furthering her own goals.

The InFFrA architecture

4.3

Bearing these requirements in mind, we have developed a social reasoning architecture for reasoning about interaction processes using the empirical semantics approach called InFFrA (the Interaction Frames and Framing Architecture). As we have explained in section 2, we use Mead's model of the construction of the "self" in combination with Goffman's concepts of *frames* and *framing* as a foundation for the architecture. Frames are used as data structures to represent classes of interaction patterns because Mead's theory only describes individual actions and does not provide any abstraction that enables the composition of more complex models of interaction processes. Conversely, we have to complement Goffman's rather abstract description of the framing process (i.e. the socially meaningful yet from the point of view of the individual social actor strategic application of frame knowledge) with the Meadian model of social action. This model provides a more fine-grained view of the reasoning process that unfolds while agents make communication and action decisions during an interaction and lends itself to modelling the concept of framing in computational terms.[9]

4.4

The InFFrA architecture comes in two flavours: As an abstract architecture which provides a meta-model for frame-based social reasoning and provides an abstract definition of frames and the framing process (Rovatsos et al 2002), and as a concrete computational model called m^2inffra of an instance of abstract InFFrA that is based on empirical semantics and uses a combination of hierarchical reinforcement learning (Barto and Mahadevan 2003), case-based reasoning (Kolodner 1993) and cluster validation techniques. Since both the abstract architecture (Rovatsos et al 2002) and the concrete computational model (Fischer and Rovatsos 2004; Rovatsos et al 2004; Fischer et al 2005) have been described in previous accounts[10] we will only provide a brief description here which only covers those aspects that are necessary to understand how InFFrA re-interprets the empirical semantics approach from an agent-centric, individualistic perspective.

Abstract architecture

4.5

InFFrA is an abstract framework for reasoning about and learning different classes of interactions, characterised by so-called interaction frames (or simply frames, for short). Each of these frames describes a category of interaction situations in terms of

- roles held by the interacting parties and relationships between them,
- trajectories that describe the observable surface structure of the interaction, and
- context and belief conditions that need to be fulfilled for the respective frame to be enacted.

4.6

Furthermore, InFFrA defines framing as the activity of constructing, adapting and strategically applying a set of interaction frames from the point of view of a locally reasoning agent in accordance with her private goals. Roughly speaking, framing consists of four phases:

1. Interpreting the current interaction situation in terms of a perceived frame and matching it against the normative model of the active frame which determines what the interaction

should look like.

2. Assessing the current active frame (based on whether its conditions are currently met, whether its surface structure resembles the perceived interaction sequence, and whether it serves the agent's own goals).

3. Deciding on whether to retain the current active frame or whether to re-frame (i.e. to retrieve a more suitable frame from one's frame repository or to adjust an existing frame model to match the current interaction situation and the agent's current needs) on the grounds of the previous assessment phase.

4. Using the active frame to determine one's next (communicative/social) action, i.e. apply the active frame as a prescriptive model of social behaviour in the current interaction encounter.

4.7

InFFrA provides a generic model of frames and framing without making any specific requirements for the representations and algorithms that are actually used in concrete implementations. Also, it does not pre-assume that the empirical semantics approach is used to derive frame models from observation and to make predictions in terms of the expected continuation of an ongoing encounter and allows for different uses of the architecture (including, for example, strict execution of hand-coded, immutable frames). Yet it lays out the general design of concrete frame-based social reasoning architectures and identifies the "primitives" each concrete architecture needs to provide definitions for.

Computational model

4.8

Starting from abstract InFFrA, the $\mathrm{m}^2\mathrm{inffra}$ model has been proposed as one possible concrete instance of the general framework that can be readily (and has been) implemented and includes learning and generalisation capabilities as well as methods for boundadly rational decision making in interaction situations.

4.9

Essentially, $\mathrm{m}^2\mathrm{inffra}$ can be seen as the result of imposing a number of constraints on general SISs to yield a simple, computationally lightweight representation of expectation structures which allows for an application of learning and decision-making algorithms that are appropriate for implementation in reasonably complex socially intelligent agents. Using the EN/SIS terminology, these restrictions can be described as follows:

- $\mathrm{m}^2\mathrm{inffra}$ only considers two-party turn-taking interaction episodes called *encounters* which have clear start and termination conditions. In other words, agents interact by initiating a "conversation" with a single other agent, exchanging a number of messages in a strictly turn-taking fashion, and can unambiguously determine when this conversation is finished.

- All social reasoning activity is conducted within the horizon of the current encounter. This implies that subsequent agent dialogues are not related to each other, which is equivalent to limiting the EN that would result from constructing an expectation structure from interaction experience to a certain depth.

- ECAs are equivalent to elementary utterances, i.e. each utterance is considered an independent and self-contained communicative action (there are no composite ECAs) that is taken as a primitive in forming expectations.

- The trajectory of every $\mathrm{m}^2\mathrm{inffra}$ frame is a sequence of message patterns (i.e. speech-act like messages which may contain variables for sender, receiver and (parts of their) content) that describes the surface structure of a particular class of interaction encounters. Thus, an individual frame trajectory corresponds to a path on an EN, and while the whole set of frames (or *frame repository*) an agent may dispose of is equivalent to a tree,

whenever the agent activates a single frame he disregards all other possible paths of execution and reasons only about the degrees of freedom provided by a single frame (at least until the next re-framing process). This helps to greatly reduce the complexity of the expectation structure reasoned about at least as long as the currently active frame can be upheld (e.g. it is not applicable/desirable anymore).

- In addition to its trajectory model, each frame keeps track of the number of encounters that matched (prefixes of) that trajectory (which is the simplest possible method to derive transition probabilities in the EN view), and lists of corresponding variable substitutions/logical conditions to record the values variables had in previous enactments of the frame and the conditions that held true at the time of enactment.

- Agents maintain a set of these frames instead of an EN but the size of this frame repository is implicitly bounded because agents apply generalisation techniques (which make use of heuristics from the area of cluster validation) to represent similar encounters by a single, (by virtue of replacing instance values by variables) more abstract frame whenever this seems appropriate (Fischer 2003). What this means is that agents are not allowed to grow arbitrarily large ENs and are instead forced to coerce their experience into a frame repository of manageable size.

- The SIS initialisation and update mechanism is fairly simple. Agents start out with a frame repository specified *a priori* by the human designer, and simply add every new encounter they experience to this repository in the form of a new frame unless it can be subsumed under an existing frame as a new substitution/condition pair or the generalisation methods mentioned above suggest abstracting from some existing frame conception to accommodate the new observation (in which case it also simply becomes a substitution/condition pair in the newly created, more abstract frame). While it is possible in theory to observe third-party encounters in which one is not directly involved as a participant this method is not used at present.

- We assume that each agent can assess the usefulness of any sequence of messages and physical action (i.e. ground instance of any frame trajectory) using a real-valued *utility function*. This facilitates the application of decision-theoretic principles (note, however, that this utility estimate need not be equivalent to the rewards received from the environment and is just thought to provide the agent with hints as to which possible future interaction sequences to prefer).

- The agent's decision-making process is modelled as a two-level Markov Decision Process (MDP) (Puterman 1994). At the *frame selection* level, agents pick the most appropriate frame according to its long-term utility and the current state. For this purpose, we apply the *options* framework (Precup 2000) for hierarchical reinforcement learning to interpret encounter sequences as macro-actions in the MDP sense and to approximate the value of each frame in each state through experience. We assume that the rewards received from the environment after an interaction encounter depend only on the physical (i.e. environment-manipulating) actions that were performed during that encounter (however, non-physical actions are assigned a small negative utility to prevent endless conversations that do not result in physical action). At the "lower" *action selection* level, the agent seeks to optimise her choices given the degrees of freedom that the current frame still offers. These are defined by the variables contained (and still unbound by the encounter so far) in the remaining steps of the trajectory model of the active defined. Here, using a (domain-dependent) similarity measure over messages and considering past cases as stored in the substitution/condition lists of the active frame, we are able to derive probabilities for the possible outcomes of the frame (and for the moves the other party might make within it). Together with the utility estimates for each of these predictions, agents can then choose the action that maximises the expected utility of the encounter to be performed in the next step.

To make the workings of the m^2inffra model more concrete, we go through a simple example:

$$
\begin{aligned}
F = \Big\langle & \big\langle \xrightarrow{5} \mathtt{request}(A,B,X) \xrightarrow{3} \mathtt{do}(B,X) \big\rangle, \\
& \big\langle \{can(B,X)\}, \\
& \ \{can(B,\mathtt{pay}(S))\} \big\rangle \\
& \big\langle \xrightarrow{2} \langle [A/\mathtt{a}], [B/\mathtt{b}], [X/\mathtt{pay}(\$100)] \rangle, \\
& \ \xrightarrow{1} \langle [A/\mathtt{b}], [B/\mathtt{a}], [X/\mathtt{pay}(S)] \rangle \big\rangle
\end{aligned}
$$

This frame reflects the following interaction experience: A asked B five times to perform (physical) action X, out of which B actually did so in three instances. In two of successful instances, it was $\mathtt{a}$ who asked and $\mathtt{b}$ who headed the request, and the action was to pay $100. In both cases, $can(\mathtt{b},\mathtt{pay}(\$100))$ held true. In the third case, roles were swapped between $\mathtt{a}$ and $\mathtt{b}$ and the amount S remains unspecified (which does not mean that it did not have a concrete value, but that this was abstracted away in the frame). Note that in such frames it is neither required that the reasoning agent is either of $\mathtt{a}$ or $\mathtt{b}$, nor that all the trajectory variables are substituted by concrete values. Also, trajectories may be specified at different levels of abstraction. Finally, any frame will only give evidence of successful completions of the trajectory, i.e. information about the three requests that were unsuccessful have to be stored in a different frame.

In the $\mathrm{m}^2\mathrm{inffra}$ reasoning cycle, the reasoning agent A enters the framing loop whenever sub-social (e.g. BDI) reasoning processes generate a goal that requires actions to be taken that A cannot perform herself. (If A is already engaged in an ongoing conversation, this step is skipped.) From all those frames contained in her frame repository $\mathcal{F} = \{F_1, \ldots F_n\}$ she then picks the frame that (1) achieves the goal[11], (2) is executable in the current state of affairs and (3) has proven reliable and utile in the past (this is done using the reinforcement learning methods described above). Let us assume that frame is the example frame F used above. In a decision-making step that does not mark the initiation of a new encounter (i.e. if the interaction has already started), A would also have to ensure that the frames considered for selection match the current "encounter prefix", i.e. the initial sequence of messages already uttered.

Once the frame has been selected, A has to make an optimal choice regarding the specific choices for variables the frame may contain. In the case of F this is trivial, because if X is already instantiated with the action A wants B to perform, then the frame leaves no further degrees of freedom. However, if, for example, the frame contained additional steps and/or variables(e.g. an exchange of arguments before B actually agrees to perform X), A would compute probabilities and utility estimates for each ground instance of the "encounter postfix" (the steps still to be executed along the current frame trajectory) to be able to chose that next action to perform which maximises the expected "utility-to-go" of the encounter.

The process of reasoning about specific action choices *within* the bounds of a single frame of course involves reasoning about the actions the other party will perform, i.e. it has to be borne in mind that some of the variables in the postfix sequence message patterns will be "selected" by the opponent.

As the encounter unfolds, either of the two parties (or both) may find that the current active frame is not appropriate anymore, and that there is a need to *re-frame*. Three different reasons may lead to re-framing which spawns a process that is similar to frame selection at the start of an encounter:

1. The other party has made an utterance that does not match the message pattern that was expected according to the active frame.

2. At least one of the physical actions along the postfix sequence is not executable anymore

because some of its pre-conditions are not fulfilled (and not expected to become true until they are needed).

3. No ground instance of the remaining trajectory steps seems desirable utility-wise.

While the first two cases are straightforward in the sense that they clearly necessitate looking for an alternative frame, the last step largely depends on the "social attitude" of the agent, and is closely related to issues of social order as discussed in section 2. Obviously, if agents were only to select frames that provide some positive profit to them, cooperation would be quite unlikely, and also they would also be prone to missing opportunities for cooperation because they do not "try out" frames to see how profitable they are in practice.

To balance social expectations as captured by the current set of frames with the agent's private needs, we have developed an entropy-based heuristics for trading off *framing utility* against *framing reliability* (Rovatsos et al 2003). Using this heuristics, the agent will occasionally consider frames that do not yield an immediate profit, if this is considered useful to increase mutual trust in existing expectations.

Finally, agents terminate the encounter when the last message on the trajectory of the active frame has been executed (unless the other party sends another message, in which we have to re-frame again). Whenever no suitable frame can be found in the trajectory that matches the perceived message sequence, this sequence is stored as a new frame, i.e. agents are capable of learning frames that are new altogether.

4.10

The m²inffra architecture has been successfully implemented and validated in complex *argumentation-based negotiation* scenarios. It can be seen as a realisation of the empirical semantics approach for agent-level social reasoning architectures thus illustrating its wide applicability.

Evaluation

5.1

Interaction trajectories like those produced by InFFrA agents can be used as empirical input for the calculation of the systemic empirical and empirical-rational semantics as described in Section 3, and this allows for the evaluation of InFFrA results using *expectation networks*, i.e., from a top-down systemic perspective. Of course, since this text is concerned with Socionics approaches, the results make statements about social systems for *artificial* agents. The applicability of our concepts and methods to "human" social systems is still unclear, and out of the scope of our work.

5.2

In order to achieve the experiment practically, all InFFrA messages are syntactically transformed into either single communication acts (in the sense of the systemic view on empirical semantics) in form of ECAs (respectively projections as the content of ECAs), or "physical" (i.e., non-symbolic) actions. What is interesting in this regard is that it is indeed possible to map all symbolic InFFrA messages to a single type of communication act, namely *projections* without a loss of meaning. While *projections* essentially represent the *ostensible* (i.e., communicated) objectives of agents, InFFrA-internally a similar mapping of the various communication act types to a relatively small set of internal representations is done, each reflecting the *actual* (mental) objectives of the agents associated with the respective communication act.

Application scenario

5.3

As a concrete InFFrA-based application, we will focus on the InFFrA-based LIESON simulation framework (Rovatsos 2004) for initial experiments. LIESON simulates knowledge-based agents that seek to maximise the popularities of web sites through intelligent link modification and intelligent communication with other agents.

5.4

In this system, agents represent Web site owners who hold different views of the contents of other Web sites *private ratings*. At the same time, they can express their opinion about others' sites by attaching numerical weight labels to links laid toward these sites, so that these link weights function as *public ratings*. The physical actions available to an agent `A` are `addLink(A,B,R)`, `deleteLink(A,B)` and `modifyRating(A,B,R')` to add a link with public rating `R` to agent `B`, delete an existing link, or modify its current rating value to a new value `R'`.

5.5

The primary goal of agents in LIESON is to increase the dissemination of their own opinion through appropriate linkage structures, and for this purpose they negotiate with each other over mutually beneficial linkage. For this purpose, they have to (i) increase their own popularity which depends on the quality (rating value) and number of incoming links, (ii) to increase (decrease) the popularity of favoured (disliked) sites, and (iii) decrease (increase) the difference between their own ratings towards third-party party sites and those expressed by favoured (disliked) sites. This last aspect follows the intuition that the more two sites "like" each other, the more should they strive to express similar opinions regarding third parties so as to increase Web transparency for Web users.

5.6

The utility function in LIESON computes the popularity of each site on the grounds of a hypothetical model of Web user behaviour, according to which the probability of following a link is proportional to the numerical weight attached to a link. The total utility each agent receives after each simulation round is based on these site popularities and takes aspects (i)-(iii) above into account.

5.7

What is interesting about this utility function is that it yields very low utilities to all agents for empty, full negative and full positive linkage. This means that if agents do not lay any links at all, or if they lay links to every other site using uniformly maximal or uniformly minimal rating values for all links their performance will be very poor. On the other hand, if they truthfully link to every agent and display their true private rating of that site with every link ("honest linkage") or use a "politically correct" (PC) linkage scheme which is identical to honest linkage except that no links with negative rating values are laid, their performance is very high. Interestingly, PC linkage provides a substantially higher average utility than honest linkage, i.e.~agents are better off concealing their discontent toward other sites.

5.8

LIESON agents reason about their actions along the following lines: Using their local link network knowledge, they project the usefulness of a number of physical actions and prioritise them using a goal/action queue in a BDI-like fashion. Then, they choose the topmost queue element for execution (unless its consequences have already been achieved or it is not applicable under current circumstances) and either (i) execute it themselves if this is possible or (ii) request its execution by an agent who can perform it (in the linkage scenario, this agent can always be uniquely identified). After such a request, the InFFrA component takes control of agent action until the initiated dialogue is terminated and processes the frame repository and the perceived messages as described in the previous sections.

5.9

In the simulations on which the data was generated that we discuss in this paper, we used a set of simple *proposal-based negotiation frames* that all agents were equipped with at the beginning and which allow for accepting and rejecting requests, making counter-proposals ("I can't do `X` as requested, but I can do `Y` for you instead") and reciprocally conditional proposals ("I will do `X` as you requested, if you do `Y` in return").

Note that in these simulations, agents have the possibility to reject any proposal, so in principle they can avoid any undesirable agreement. However, this does not imply that they will adhere to the frames, because they might be insincere and not execute actions they have committed themselves because their private desirability considerations suggest different utility values from those expected when the agreement was reached (or simply because they calculated that lying

is more profitable than keeping one's promises).

5.10

The following list explains the meaning of LIESON messages and their *elementary communication act* (ECA) counterparts (from section 3):

addLink(agent1,agent2,w)
Add link from agent1's site to agent2's site with weight w
--> do(addLink(agent1, agent2, w))

deleteLink(agent1, agent2)
Remove link from agent1's site to agent2's site
--> do(deleteLink(agent1, agent2))

modifyRating(agent1, agent2, w)
Modify the weight of an existing link
--> do(modifyRating(agent1, agent2, w))

request(agent1,agent2,act(...))
agent1 asks agent2 to perform act(...)
--> project(agent1, agent2, do(act(...))

accept(agent1,agent2,act(...))
agent1 agrees to perform act(...)
--> project(agent1, agent1, act(...)) I.e., "accept" means to project the fulfilment of a previously requested action (a request of agent1 to do something by herself, so to say). This can also be done implicitely by performing the requested action.

reject(agent1,agent2,act(...))
agent1 rejects a request or proposal
--> project(agent1, agent2, not act(...))

propose(agent1,agent2,act(...))
agent1 proposes to perform by herself act(...)
--> project(agent1, agent2, do(act(...))

Further performatives used in LIESON are not significant for the purposes of this article.

5.11

It should be noted that these communication acts are, from the perspective of the *expectation network* (EN) learning algorithm i) trigger actions for expectations the uttering agents raise themselves regarding their own future behaviour, ii) regarding the desired behaviour of other agents, corresponding to the message contents, and iii) trigger actions for the calculation of the expectations of the observer (i.e., the maintainer of the EN). i), ii) and iii) are generally not identically of course, except in the case of completely sincere, reliable and cooperative agents. But in any case determining an EN from sequences of such acts reflects the self-induced expectation structures of the communication system, not some normative or pre-defined meaning of the communication acts.

5.12

From LIESON protocols and other empirical discourse data, a *semantics observer* can basically obtain two kinds of ENs automatically using either empirical or empirical-rational semantics: ENs of type A are obtained *without* any assumption of observable agent rationality (purely-empirically, so to say), whereas ENs of type B are yielded from the data using additional assumptions about ostensible agent rationality as described in section 3, i.e., ENs of type A depict empirical semantics whereas ENs of type B depicts empirical-rational semantics. In both cases, the respective EN models the expected, uncertain continuation of the protocol for arbitrary time steps in the future. Practically, ENs of type A model an agent behaviour at which the agents repeat the observed sequence in a stereotypical manner, whereas type B ENs reflect the preferableness of such action sequences which have been seemingly successful in the past regarding the achievement of communicated goals (including self-commitments), even if such sequences did not show up more frequently or more recently in the observed sequence.

5.13

As an example typical for the data obtained by LIESON, consider the following sequence of LIESON agent actions:

Step 1: Request 1
```
request(agent1, agent2,addLink(agent2,agent1,3))
```

Step 2: Denial
```
reject(agent2,agent1, addLink(agent2,agent1,3))
```

Step 3: Counter proposal
```
request(agent2, agent1, addLink(agent1,agent2,4))
```

Step 4: Accept counter proposal (implicit)
```
addLink(agent1,agent2,4)
```

Step 5: Request 1 (again)
```
request(agent1, agent2, addLink(agent2,agent1,3))
```

Step 6: Accept request 1 (implicit)
```
addLink(agent2,agent1,3)
```

Step 7: Request 2
```
request(agent1, agent2, addLink(agent2,agent1,3))
```

Step 8: Accept request 2 (implicit)
```
addLink(agent2,agent1,3)
```

5.14

ENs of type B predict that, after having experienced this sequence, agent 1 has learned that fulfilling the counter-proposal of agent 2 (to perform action `addLink(agent1,agent2,4)`) is likely a prerequisite for making agent 2 fulfill request 1. In contrast to a corresponding type A EN, here the counter proposal as well as the denial of request 1 becomes superfluous in a significant number of cases, since (in the EN-based communication model as well as in reality) agent 1 anticipates the communicated *projections* of agent 2.

Experiments

5.15

In order to estimate the performance of the EN-based prediction algorithm applied to action sequences like this, we performed several experiments. ENs of both types were retrieved from prefixes of the whole protocol as empirical evidence, and then parts of the rest of the protocols (i.e., the actual continuations of the conversation) were compared with paths within the ENs to yield an estimation for the prediction achievement. The comparisons were performed automatically, since the resulting ENs were mostly to complex to evaluate them manually.

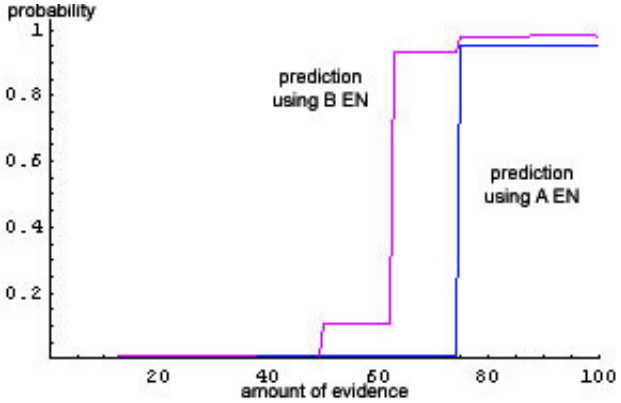

Figure 2. Prediction of communication sequences (example 1)

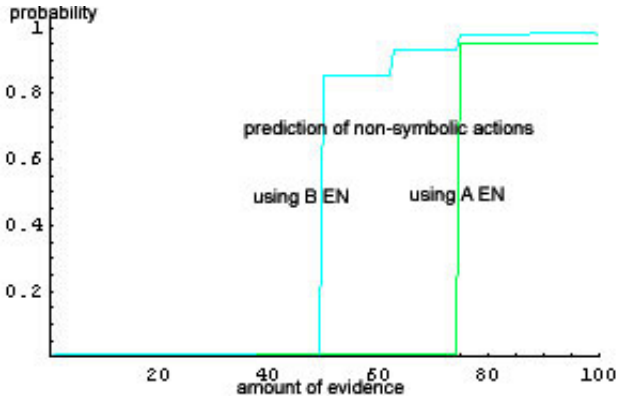

Figure 3. Prediction of non-symbolic actions (example 1)

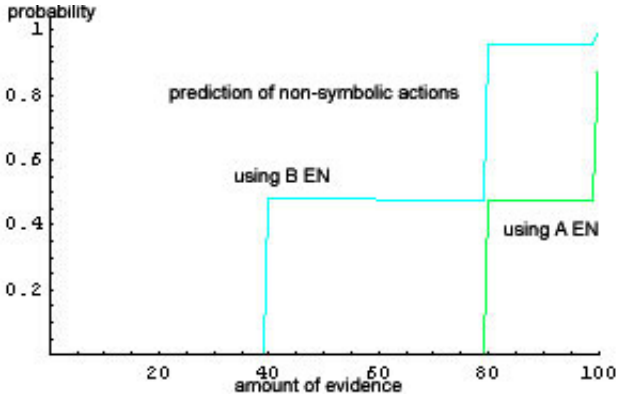

Figure 4. Prediction of communication sequences (example 2)

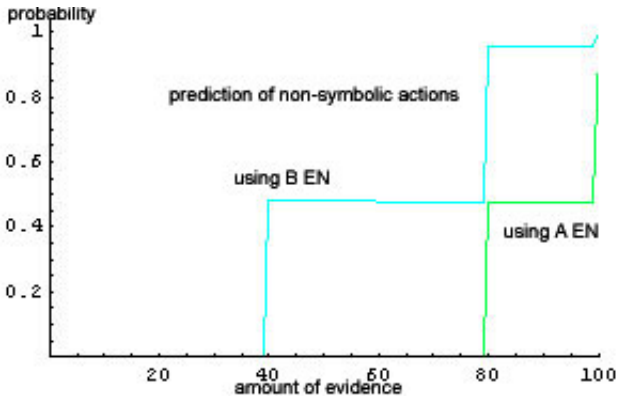

Figure 5. Prediction of non-symbolic actions (example 2)

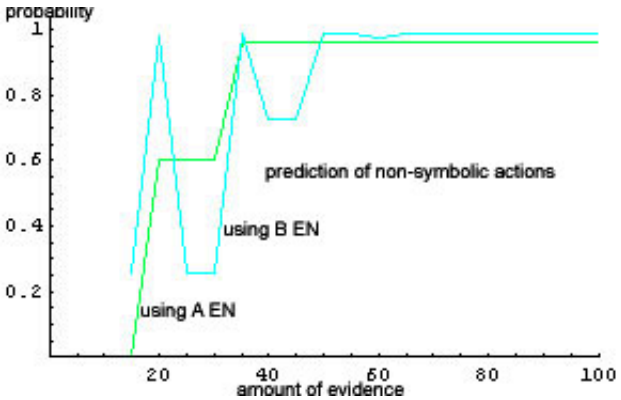

Figure 6. Prediction of non-symbolic actions (example 3)

5.16

Figures 2 and 3 show the results for this general setting applied to the example sequence of actions shown above (steps 1–8). The observations used as evidence to retrieve the ENs range from 0% to 100% of the protocol, taken as a prefix. In Figure 2, the values plotted on the y-axis are the probabilities of the last 20% of the complete protocol (i.e., the respective probability that the event sequence described by the protocol suffix will actually happen, e.g., that agent 2 complies to the requests of agent 1 in the example above). These probabilities are stated by an

EN which was retrieved only from evidence in form of the respective protocol prefix (x-axis = prefix length in percent). Thus, the values on the y-axis denote how good the respective EN predicts the protocol suffix depending on the length of a sequence of observations. The blue curve in Figure 2 was gained from a type A EN (i.e., purely-empirically), and the purple curve there was yielded using a type B EN (i.e, empirical-rationally). The blue curve simply reflects that the predicted sequence did not occur in the evidence sequence before approx. the 75% mark, but occured afterwards once, from which is concluded that it will likely occur again later. In contrast, the purple curve is caused by the fact that from approx. 50% evidence on, the empirical-rational EN anticipates that the agent 1 can increase the likelihood that agent 2 will comply with `request(agent1, agent2,addLink(agent2,agent1,3))` by accepting the counter proposal (`request(agent2, agent1, addLink(agent1,agent2,4))`).

5.17

Figure 3 shows similar results, but in contrast to Figure 2, here values plotted on the y-axis are the maximum occurrence probabilities (according to the respective EN) of non-symbolic actions (i.e., do(), addLink()) within the last 20% of the example protocol, calculated after the observations of protocol prefixes with lengths indicated by the percentages on the x-axis. Such results are of course quite far from the full empirical semantics (the complete EN), but they are an indicator for EN performance in cooperative settings. In the concrete example, from experiencing a certain percentage of the sequence, an EN is calulated, and over all occurrences of the non-symbolic action `addLink(agent2,agent1,3)` within this EN, the maximum expectation is taken and plotted on the y-axis. This is done both purely-empirically (green curve) and empirically-rational (cyan curve). Figure 3 thus shows the overall probability of fulfilling agent 1's request. This yields roughly the same result as above (Figure 2), but is for technical reasons much easier to compute in case the predicted sequence (the 20% at the end of the example sequence here) is very long. Figure 4 and 5 show such a case (rather extreme, for clearness), obtained from a protocol of 236 LIESON actions. Here, Figure 5 shows a good predictability, whereas Figure 4 predicts the protocol remainder only if at least 80% of the protocol is given as evidence, i.e., if the predicted sequence and the evidence overlap.

5.18

We have applied the same experimental setting to several LIESON protocols, and found basically two classes of results: Those in the style of Figure 2-5 can be interpreted as above, i.e., two or more agents make requests to each other and eventually comply to the respective other's request when the conditions for compliance are given, that is the other agent complies reciprocally. More generally speaking, the communications evoke certain expectations about desired and proposed behaviour, and the following interaction validates these expectations more or less.

5.19

Very different results were obtained in case such self-induced expectations are disappointed in the further course of communication, as in Figure 6 (especially if the agents try to cheat on each other, or do not fulfil their own proposals for other reasons, e.g., because the environmental conditions for reaching the respective agent goals are not given any more). Here, the "empirical-rational" graph (cyan) shows a quite unstable prediction performance, whereas the "purely-empirical" curve (green) shows a relatively good conduct. A possible explanation (which could not be verified in this and similar experiments for sure because the protocol was to large to evaluate it "manually") is that the type B ENs over-emphasize self-induced expectations (which appear to fail often here), whereas the purely markovian type A ENs are more robust in such cases because they simply ignore the ostensibly-rationale public stances of the agents.

5.20

What can already quite safely be concluded from our experiments is the following:

- In case of predominantly reliable interaction processes, empirical-rational ENs perform significantly better than the purely-empirical ones in terms of their prediction power, and both EN types show an increasing prediction power with an increasing amount of evidence, as anticipated.
- In our other experimental settings, the purely-empircal ENs predicted the future either equally well, or more reliably than empirical-rational ENs in their current implementation.

This does not mean that these settings are "irrational", of course, but shows simply that learning ENs of type A does not apply social choice towards projected goals, and thus cannot, in the short them, be mislead easily by, e.g., fraudulence. Although both types of ENs observe the fact that the interaction system's self-induced expectations adapt themselves to fraudulence and other kinds of non-compliance to commitments in the longer term, B type ENs apparently blur this adoption process by overemphasizing the seemingly-rational intentional stances of the agents, given the rather short protocols in our experiments. If, e.g., in certain contexts some agent fails to perform actions in favour of other agents (which she had previously announced), the other agents will eventually stop issueing such requests in similar contexts. Since all kinds of ENs are able to model dynamic social contexts, they should correctly predict this context-dependant behaviour, but it takes time to recognize such a behavioural change from the viewpoint of an external observer. The LIESON protocols available in our experiments were not long enough to show how this adoption process performs for type B ENs.

- In many cases, InFFrA agents show up a behaviour which coincides extremely well with the behaviour predicted by ENs. A more speculative conclusion from this is that the InFFrA-internal kind of empirical semantics is in parts similar to the (purely-)empirical semantics determined by the EN learning algorithms (which can be verified comparing InFFra with EN at the programming level).

5.21

In future work, it is certainly necessary to evaluate the EN learning algorithms with really long LIESON protocols in order to find further evidence for the suitability of our approach in highly unstable settings, and to improve our algorithms if necessary.

Conclusion

6.1

This paper proposes the empirical semantics approach to agent communication semantics as a way to conceive of the meaning of communication processes in terms of their expected continuations. We have presented two different formal frameworks based on this idea, one that realises the viewpoint of a systemic observer aiming at the approximation of the communicative behaviour of multiple black-box agents as empirical expectations, and another which reasons about empirical semantics from the perspective of an individual agent, focusing on individually rational, goal-based interaction behaviour. From a Socionics point of view, these approaches highlight some very promising insights for future theoretical development. To combine individual-oriented and system-oriented views on social order in a common model focusing on meaning generation (empirical semantics) via expectations seems to be a viable path to overcome traditional barriers in sociological theory. Experiments showing the compatability of both perspectives may therefore lead the way to a more integrated view on the different levels of sociality. But note in this respect, that both our systemic and our individualistic perspective consider at the moment only artificial agents, not humans. Our insights and results are thus valid for such agents only. Further implications pertaining to "human" social systems would certainly demand further experiments, with real-world data or more "human-like" agents. Although already very promising results have been achieved, and several applications and related research fields were exposed (e.g. Nickles et al 2004c; Rovatsos et al 2004a; Fischer et al 2005; Nickles et al 2005b), there is still a long way ahead to fully unfold and realise the presented approach, being a truly novel paradigm in the field of Distributed Artificial Intelligence.

Acknowledgements

We thank the anonymous reviewers for their profound and very helpful comments. This work has been supported by Deutsche Forschungsgemeinschaft (DFG) under contracts no. BR 609/11-2 and MA 759/4-2. We would also like to thanks former members of the *ConStruct* project for their very valuable contributions to this work, namely Kai F. Lorentzen, Kerstin Beck, Felix Brandt and Ulrike Löwer, as well as to our colleagues from the other projects within the DFG Priority Program *Socionics* for their excellent cooperation.

Notes

[1] In Luhmann's view, codes are binary distinctions on which a system bases all its operations and observations, whereas a systems programs structure and organize the attribution of events to both sides of the binary code.

[2] Of course, the (expectation of) triggered behaviour can trigger (the expectation of) other agent's behaviour and so on.

[3] Actually, two different paths can have the same semantics in terms of their expected continuations, a fact which could be used to reduce the size of the EN by making them directed graphs with more than one path leading to a node instead of trees as in this work.

[4] To be precise, a single utterance might be split into several so-called elementary communication acts, each corresponding to a dedicated EN node.

[5]We assume a discrete time scale with $t \in \mathbb{N}$, and that no pair of actions is performed at the same time (quasi-parallel events achieved through a highly fine grained time scale), and that the *expected* action time corresponds with the depth of the respective node within in the EN.

[6]A future version of our framework might allow the utterance of whole ENs as projections, in order to freely project new expectabilities or even introduce novel event types not found in the current EN.

[7]This time span of projection trustability can be very short though — think of *joke questions*.

[8]This probability distribution must also cover projected events and assign them a (however low) probability even if these events are beyond the spheres of communication, because otherwise it would be impossible to calculate the rational hull.

[9]We regard the identification of the necessity of combining these two (admittedly closely related) theories to be able to produce adequate computational models of reasoning about interaction as one of the major insights of our research that sociologists can benefit from. This nicely illustrates the bi-directional benefits of transdisciplinary collaboration in the Socionics research programme.

[10]In particular, Rovatsos et al (2004) describes both models and their theoretical foundations in detail and includes an account of an extensive experimental validation of the approach.

[11]In an AI planning sense, agents will also activate frames that achieve sub-goals towards some more complex goal, but we ignore this case here for simplicity.

References

AUSTIN J L (1962) *How to Do Things with Words.* Oxford: Oxford University Press, 1962.

BARTO A G and Mahadevan S (2003) Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems* , 13(4):41–77, 2003.

COHEN P R and Levesque H J (1990) Performatives in a Rationally Based Speech Act Theory. In *Proceedings of the 28th Annual Meeting of the ACL*, 1990.

COHEN P R, Levesque H J (1995) Communicative Actions for Artificial Agents. In *Proceedings of ICMAS-95*, 1995.

FISCHER F and Nickles M (2006) Computational Opinions. *Proceedings of the 17th European Conference on Artificial Intelligence* (ECAI-2006), IOS press. To appear.

FISCHER F and Rovatsos M (2004) Reasoning about Communication: A Practical Approach based on Empirical Semantics. *Proceedings of the 8th International Workshop on Cooperative Information Agents* (CIA-2004), Erfurt, Germany, Sep 27-29, LNCS 3191, Springer-Verlag, Berlin, 2004.

FISCHER F (2003) Frame-based learning and generalisation for multiagent communication. Diploma Thesis. Department of Informatics, Technical University of Munich, 2003.

FISCHER F, Rovatsos M and Weiss G (2005) Acquiring and Adapting Probabilistic Models of Agent Conversation. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems* (AAMAS), Utrecht, The Netherlands, 2005.

GAUDOU B, Herzig A, Longin D, and Nickles M (2006) A New Semantics for the FIPA Agent Communication Language based on Social Attitudes. *Proceedings of the 17th European Conference on Artificial Intelligence* (ECAI-2006), IOS press. To appear.

GOFFMAN E (1974) *Frame Analysis: An Essay on the Organisation of Experience*. Harper and Row, New York, NY, 1974.

GRICE H. P. (1968) Utterer's Meaning, Sentence Meaning, and Word-Meaning. *Foundations of Language*, 4: 225-42, 1968.

GUERIN F and Pitt J (2001) Denotational Semantics for Agent Communication Languages. In *Proceedings of Agents'01*, ACM Press, 2001.

KOLODNER J L (1993) *Case-Based Reasoning* . Morgan Kaufmann, San Francisco, CA, 1993.

LABROU Y and Finin T (1997) Semantics and Conversations for an Agent Communication Language. In *Proceedings of IJCAI-97*, 1997.

LORENTZEN K F and Nickles M (2002) Ordnung aus Chaos — Prolegomena zu einer Luhmann'schen Modellierung deentropisierender Strukturbildung in Multiagentensystemen. In T. Kron, editor, *Luhmann modelliert. Ans‰tze zur Simulation von Kommunikationssystemen*. Leske & Budrich, 2002.

LUHMANN N (1995) *Social Systems*. Stanford University Press, Palo Alto, CA, 1995.

MATURANA H R (1970) Biology of Cognition. Biological Computer Laboratory Research Report BCL 9.0. Urbana, Illinois: University of Illinois, 1970.

MEAD G H (1934) *Mind, Self, and Society*. University of Chicago Press, Chicago, IL, 1934.

MÜLLER H J , Malsch Th and Schulz-Schaeffer I (1998) SOCIONICS: Introduction and Potential. *Journal of Artificial Societies and Social Simulation* vol. 1, no. 3, http://jasss.soc.surrey.ac.uk/1/3/5.html

NICKLES M and Weiss G (2003) Empirical Semantics of Agent Communication in Open Systems. *Proceedings of the Second International Workshop on Challenges in Open Agent Environments*. 2003

NICKLES M, Rovatsos M and Weiss G (2004a) Empirical-Rational Semantics of Agent Communication. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS'04), New York City, 2004.

NICKLES M, Rovatsos M, Brauer W and Weiss G (2004b) Towards a Unified Model of Sociality in Multiagent Systems. *International Journal of Computer and Information Science* (IJCIS), Vol. 5, No. 1, 2004.

NICKLES M, Froehner T and Weiss G (2004c) Social Annotation of Semantically Heterogeneous Knowledge. Notes of the *Fourth International Workshop on Knowledge Markup and Semantic Annotation* (SemAnnot-2004), Hiroshima, Japan, 2004.

NICKLES M and Weiss G (2005) Multiagent Systems without Agents: Mirror-Holons for the Compilation and Enactment of Communication Structures. In Fischer K, Florian M and Malsch

Th. (Eds.): *Socionics: Its Contributions to the Scalability of Complex Social Systems*. Springer LNAI, 2005.

NICKLES M, Rovatsos R and Weiss G (2005a) Expectation-Oriented Modeling. In *International Journal on Engineering Applications of Artificial Intelligence* (EAAI) Vol. 18. Elsevier, 2005.

NICKLES M, Fischer F and Weiss G (2005b) Communication Attitudes: A Formal Approach to Ostensible Intentions, and Individual and Group Opinions. In *Proceedings of the Third International Workshop on Logic and Multi-Agent Systems* (LCMAS), 2005.

NICKLES M, Fischer F and Weiss G (2005c) Formulating Agent Communication Semantics and Pragmatics as Behavioral Expectations. In Frank Dignum, Rogier van Eijk, Marc-Philippe Huget (eds.): *Agent Communication Languages II*, Springer Lecture Notes in Artificial Intelligence (LNAI), 2005.

PARSONS T (1951) *The Social System*. Glencoe, Illinois: The Free Press, 1951.

PITT J and Mamdani A (1999) A Protocol-based Semantics for an Agent Communication Language. In *Proceedings of IJCAI-99*, 1999.

PRECUP D (2000) Temporal Abstraction in Reinforcement Learning. PhD thesis, Department of Computer Science, University of Massachusetts, Amherst, 2000.

PUTERMAN M L (1994) *Markov Decision Processes* . John Wiley & Sons, New York, NY, 1994.

RON D, Singer Y and Tishby N (1996) The Power of Amnesia - Learning Probabilistic Automata with Variable Memory Length. In *Machine Learning* Vol. 25, p. 117–149, 1996

ROVATSOS M, Weiss G and Wolf M (2002) An Approach to the Analysis and Design of Multiagent Systems based on Interaction Frames. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS'02), Bologna, Italy, 2002.

ROVATSOS M, Nickles M and Weiss G (2003) Interaction is Meaning: A New Model for Communication in Open Systems. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS'03), Melbourne, Australia, 2003.

ROVATSOS M, Fischer F and Weiss G (2004a) Hierarchical Reinforcement Learning for Communicating Agents. *Proceedings of the 2nd European Workshop on Multiagent Systems* (EUMAS), Barcelona, Spain, 2004.

ROVATSOS M, Nickles M and Weiss G (2004b) An Empirical Model of Communication in Multiagent Systems. In Dignum F (Ed.), *Agent Communication Languages*. Lecture Notes in Computer Science, Vol. 2922, 2004.

ROVATSOS M (2004) Computational Interaction Frames. Doctoral thesis, Department of Informatics, Technical University of Munich, 2004.

SINGH M P (2000) A Social Semantics for Agent Communication Languages. In *Proceedings of the IJCAI Workshop on Agent Communication Languages*, 2000.

SINGH M P (1993) A Semantics for Speech Acts. *Annals of Mathematics and Artificial Intelligence*, 8(1–2):47–71, 1993.

Return to Contents of this issue