Achieving Multiagent Organisation by Organising Agent Experience

Michael Rovatsos and Gerhard Weiß

{rovatsos,weissg}@cs.tum.edu
Institut für Informatik
Technische Universität München
80290 München, Germany
tel. +49-89-289-22407
fax. +49-89-289-28483

Abstract. There are two contrary types of approaches to multiagent design and modelling, each having its specific shortcomings: first, top-down (normative) approaches that emphasise social level control; and second, bottom-up (constructivist) approaches that emphasise agent autonomy. This paper introduces a "middle way" between these two types. A novel social reasoning architecture called INFFRA is described that integrates normative and constructivist properties. This architecture is based on the sociological concepts of *frame* and *framing* and focuses on the cognitive processing of social and organisational knowledge. It is presented at a conceptual level, and first suggestions are made as to its concrete implementation, as well as comments on open problems that we find challenging for our future research.

1 Introduction

The fundamental dilemma between *autonomy*, which is undoubtedly a key aspect of multiagent systems (MAS) research that has largely contributed to the success and appeal of the field, and *social-level control*, that aims at ensuring society-level coherence, has long been identified in the MAS community. In recent years, two contrary lines of research have evolved that reflect this dilemma. The one line encompasses *top-down* approaches to multiagent system design and modelling. The key characteristic of these approaches is that they offer a "structuralist" view of multiagent societies by enforcing pre-designed social structures on individual agents in a strictly *normative* fashion. Prominent examples of such social structures are pre-designed interaction protocols (e.g., [1, 8, 23]), social laws (e.g., [20]), protocols applied to a priori determine interaction paths (e.g., [1, 8, 23]), and roles employed to a priori specify required capabilities (e.g., [10]), rights and duties (e.g., [23]), and/or tasks [7] to be fulfilled by the agents. Top-down approaches are motivated by the fairly wide consensus among multiagent researchers that a normative use of social structures is a useful means to harness the dangers of incoherent, chaotic system behaviour, and so it is not surprising that much effort has been invested in developing top-down design methods, frameworks and tools ([2] provides an overview). Yet, many researchers reject this normative "top-down imposition", arguing that approaches following this line are nothing but

behaviour-constraining models of sociality. In other words, the key critique of these approaches is that they put too much emphasis on an agent-independent organisational level and, with that, inherently tend to seriously restrict agents in their autonomy. The other line of research is that of *bottom-up* approaches to multiagent system design and modelling. These are characterised by the "constructivist" view of artificial societies that they offer and that emphasises the autonomy and the motives of *individualistic* rational agents. According to this view, the focus of agent research should be on the development of rational reasoning mechanisms that enable individual agents to derive information about others and to construct models of others' goals and intentions. Two examples of such approaches are [13, 18], and a good overview of many related decision- and game-theoretic approaches is presented in [19]. A key problem with these approaches, however, is that they fail to explain in sufficient detail how socially intelligent behaviour can emerge, under various conditions, as a necessary result from the individual agents' "reasoning about others" and "models of others". In other words, while attempting to avoid any restrictions on the agents' choices, these approaches run the risk to not achieve global coherence. The reason for this simply is that individually derived "models of others" lag far behind the richness of "normative" social and organisational knowledge as used by top-down approaches and thus are not really helpful in reducing the agents' uncertainty about the "social world" in which they are embedded.

An obviously urgent question thus is how the top-down (normative) line and bottom-up (constructivist) line of multiagent design and modelling can be combined while at the same time avoiding their respective major drawbacks. In other words, the question is how to use knowledge about different kinds of social structures within a multiagent system so as to obtain social order without unnecessarily restricting agent autonomy. Undoubtedly, this question constitutes a long-term scientific challenge, since its answering requires to bring together thoughts and ideas that have evolved independently over many years. For that reason it should be clear that a complete answer can not be expected at this stage of research in the field. Instead, what is needed in a first step is a profound conceptual basis for more specific theoretical and experimental work. The work described in this paper is intended to provide such a conceptual basis. A novel social reasoning architecture called INFFRA (INteraction Frames and FRAming) is introduced that offers a "middle way" between both "top down" and "bottom up": like conventional top-down approaches, this architecture respects the importance of social and organisational knowledge in the process of achieving social order; and like conventional bottom-up approaches, this architecture respects the importance of the individual agents' behavioural freedom. INFFRA significantly differs from available top-down and bottom-up approaches in that it *re-interprets* the notion of "social structures" and "organisations" in two fundamental ways:

1. with respect to their *locus* – by viewing them as mental models of the social context that agents have rather than as rules that uniformly apply throughout an entire MAS, and
2. with respect to their *purpose* – by assuming them to be descriptive models of agent *experience* rather than prescriptive models of behaviour.

This means that according to InFFra the whole process of structuring and organising interaction processes into roles, protocols, relationships etc. takes place in the agent's

mind through active reasoning (and is not simply "adopted" by the agent), where the evolving structural and organisational patterns can only be used by agents as models of past regularities of behaviour rather than as "models of committed and ascertained future behaviour". The *key idea* underlying INFFRA thus is that social and organisational knowledge evolves within autonomous individuals through interaction, and structures future experience to the end of reducing contingencies for the individual. More specifically, according to INFFRA this integration of social-level order and individual-level autonomy is achieved through the use of special mental representations of organisational knowledge ("frames") and their application and adaptation ("framing") by autonomous, self-interested agents. The concepts of frames and framing are borrowed from the work of the sociologist Erving Goffman [11], and so INFFRA also offers the advantage of being well founded in a contemporary sociological theory.

The remainder of the paper is structured as follows. Section 2 introduces Goffman's concepts of "frame" and "framing". In Section 3 we develop a computational model of frames, which are used as the central data structure in the social reasoning architecture based on framing that is described in Section 4. Section 5 rounds up with some conclusions and an outlook on directions for further research on the topic.

2 First principles: sociological background, computational model

2.1 Sociological concepts

Frames [11] are one of the key concepts in the sociological analyses of everyday life that Erving Goffman engaged in throughout his research. In short, a frame can be seen as the answer to the simple question *"what is going on here?"* that each human poses to herself in any interaction situation, consciously or unconsciously. That is, it provides "framing" information about a particular class of interaction situations that will allow the participant to act "appropriately", i.e. in a competent, routine fashion.

In a MAS context, we can view them as "data structures" that contain sufficient organisational knowledge to structure interaction for the individual that employs them. In that, they re-construct what is *in-between agents* rather than what is *inside agents*, as most mentalistic approaches do. While offering this advantage of being "genuinely social", they are still linked to the mental processes of the agent which is using them, and hence allow for a construction of purely deliberative agent architectures.

Of course, attempting to use such frames as interaction models in order to manage interactions effectively raises the question of how they will be employed by agents in practice, provided that an adequate, sufficiently complex and rich "database" of such frames is available to the interactant which has been acquired through experience. In accordance with the imagery of knowledge frames putting the situation "in the picture", Goffman calls this process "framing" to suggest that the issue we should focus on in analysing social interaction is *which* actors use *what* frames *how* and *under which conditions*.

Inspired by this (crudely simplified interpretation of a particular) micro-social theory, we can attempt to make use of it in computational terms. In order to translate the concepts of "frames" and "framing" into tractable computational models, we must first

explain what kinds of information would be needed inside such frames in order to capture the properties of a class of interactions that make it distinct for the framing agent and thus may reduce the "search space" concerning the agent's expectations and the expectations that others have of it.

2.2 Required properties of a computational model

If a frame is to "feed" the interaction with sufficient information, it must exhibit the following properties:

1. *Common knowledge:* It must be allegedly *shared* knowledge among the interacting agents. When one agent uses it, it must assume its peer(s) to have the same information. Only if this is assumed *a priori* can the "mentalistic" models of belief (with the related epistemic problems of attaining equal belief states [12]) be abandoned in favour of models of social expectations.
2. *Relevance:* The interactional knowledge captured by the frame must be grounded in agents' experience, it must occur repeatedly, and it must describe the relevant aspects of the interaction in question adequately.
3. *Generalisation:* It must generalise from particular enactments of a class of interactions. By being expressive enough to abstract from individual actors, situational contexts and actual actions performed during an interaction it provides an expressive means of capturing the characteristics of a whole class of interactions in a single data model.
4. *Instrumentalisation:* The knowledge captured by it must relate to the agent's private goals and preferences. If the agent is to gain from using frame knowledge, that knowledge has to be processed in a goal-directed manner in order to fulfil its purpose.

As concerns requirements 1 and 4, they can be realised by separating frame knowledge into *common attributes* and *private attributes*, the former employed in reasoning "as if" interaction partners had the same knowledge, the latter expressing the agent's current stance towards the frame in question (that is not visible to others).

The conditions stated in 2 and 3 refer to "qualitative" aspects of the knowledge a frame captures. They require that agents identify what matters to them through their experience of past encounters, and that they be able to store those pieces of information efficiently in expressive representations. This will in part have to be ensured by application-specific heuristics and learning algorithms, but we have also developed a generic computational model of frames to support this process, which is discussed at length in the following section.

3 Interaction frames

Interaction frames, as we use them, must be thought of as data structures that consist of two parts, the common attributes section and the private attributes section, which follow the intuition just laid out. The common attributes section incorporates all knowledge that is necessary to carry out and interpret a particular class of interactions, while private attributes are used to store the status ascribed to common attributes in the current situation by the agent who is employing the frame.

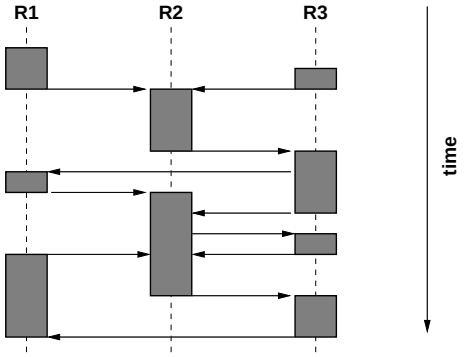

Fig. 1. A trajectory model: three actors R_1, R_2, R_3 (actually roles, cf. below) perceive each other's actions and react upon them. Agents' activities (shaded boxes) evoke reactions in their peers along the vertical time-line. Arrows denote explicit messages, while actions themselves might include "physical" actions that are publicly observable.

3.1 Common attributes

Trajectories Common attributes are formed around *trajectories* [21] as models of "guided doings" [11], i.e. temporally ordered action sequences of actors that relate to each other. These must have "communication-like" properties, which means that they are (1) explicitly *performed* by individuals, (2) *perceived* by the counter-parties involved and (3) *reacted* upon. They have the semantics of "gestures" [15], since they evoke certain reactions and generate an "image" of the actors performing them.

The most genuine kind of such trajectories is, of course, communicative action, i.e. messages in the context of communication protocols, but, in principle, any action can count as part of a trajectory as soon as it spawns some reaction in another agent.

These trajectories constitute the core of an interaction frame by expressing what the frame is *about*: a particular kind of interaction processes. In order to generalise from overtly specific descriptions that reduce the trajectory model to a single situation that will never be repeated, we use generalised *trajectory models* (with variables for actors and individual attributes of trajectories [17]) that merely describe properties of a whole class of interaction instances thus covering a (suitably) wide range of situations.

We do not intend to restrict ourselves to a single formalism for expressing such trajectories at this point, since the adequacy of particular formalisms may vary between applications, and we therefore only use abstract graphical models of protocol-like trajectories as the one depicted in Figure 1.

To the least, employed formalisms should enable a description of agents' *actions*, *time* and *branching* to account for points of choice. Good candidates include multi-modal branching-time logics with deontic operators, probabilistic sequential diagrams, etc.

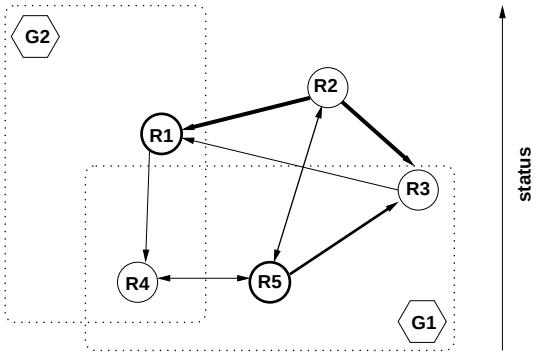

Fig. 2. Role and relationship model.

Roles & Relationships Like trajectories generalise from individual courses of action, roles abstract from individuals. We suggest that a powerful concept of *role models* as data structures must encompass three kinds of attributes that defi ne a role R:

1. *Behavioural attributes*
   (a) Expected behaviour: Information about what R usually does or does not.
   (b) Skills: Information concerning the capabilities of R (possible actions).
2. *Intentional attributes*
   (a) Goals, Tasks, Intentions: The teleological grounding of R's actions.
   (b) Values/Preferences: The valuations R has of certain events or states of affairs.
   (c) Beliefs: Knowledge states of R.
3. *Social attributes (Relationships)*
   (a) Dependencies: Actions of other roles that are needed for R to do something.
   (b) Aggregation/Representation: What roles R consists of or pertains to (in case of groups, formal organisations etc.), and information about who is seen to act on behalf of such aggregates.
   (c) Acquaintance: Information regarding the knowledge R has of other roles and agents.

Since the literature on modelling roles abounds [8, 10, 14, 23], we will not describe the formalisms we use in our model here (a detailed account can be found in [17]). Instead, we will only introduce an informal, simple graphical notation for capturing role and relationship knowledge, cf. Figure 2: It shows agent roles as rounded nodes and group roles as hexagons with a boundary box around members, possibly overlapping; these are interlinked through relationship arcs (for various types of relationships). The vertical line may be used as a "status" scale, if a one-dimensional measure has been defi ned (e.g. by computing the total of existing dependencies for each role).

Contexts Having abstracted from actions and actors, there is also a need to abstract from *situations* in which the frame becomes relevant, and this is achieved by using *context models*. These consist of two parts:

1. *Relevance Conditions*
   (a) Activation conditions: conditions, under which the frame will be adopted by the participating parties (e.g. receiving a request for information).
   (b) Deactivation conditions: statements about events and states of affairs that make agents abandon the frame (e.g. a peer terminating a discussion).
2. *Enactment Conditions*
   (a) Preconditions: conditions that are necessary for the frame to be carried out correctly (e.g. existence of disagreement in a negotiation frame).
   (b) Postconditions: conditions that are always ensured after a frame has been completed (e.g. a new task allocation).
   (c) Sustainment conditions: conditions that must hold throughout the enactment of a frame (e.g. availability of communication channels).

As Figure 3 suggests, the "scope" of a trajectory is defined by embedding it into a context model: most importantly, context models define clear-cut conditions for *adoption* and *abandonment* of the frame. Choosing appropriate conditions for these contexts has a huge impact on the usefulness of the frame, since it guides the search in a frame "repository" that the agent has at its disposition. Once activation conditions are met, the agent will use the frame and the trajectory knowledge will become normative for her by influencing action choices, until deactivation conditions become true, whereupon the agent abandons the frame and attempts "re-framing" (this is explained in more detail in Section 4).

Enactment conditions are a somewhat weaker concept than relevance conditions, but equally useful. While they don't trigger activation and deactivation, they supply information about what is needed to carry out the frame properly and about what the frame achieves. Especially *postconditions* can be of vital importance to the agent, since they make explicit what the consequences of frame enactment are, and thus guide the agent's search for a "useful" frame – a "social operator" in the terminology of AI planning, if one wants (cf. "Instrumentalisation" in Section 2.2).

Beliefs In comparison to roles, norms and contexts, the *beliefs model* part of the common attributes section in a frame plays a subordinate role with respect to our theoretical intuitions. Although it may contain beliefs that are necessary to execute and interpret the frame properly, e.g. causal or conceptual knowledge (as Figure 4 suggests), it can be neglected (or even not be filled with data at all) as long as the inter*action* itself occurs as expected. Frames still reserve a slot for such frame-related epistemic knowledge, mainly to cater for situations in which the designer may wish to "locate" certain knowledge (e.g. meta-level knowledge such as ontologies) within the boundaries of a certain frame.

Links and history. *Links*, that relate an entire frame to other frames by such relationships as aggregation, inheritance in the object-oriented sense and also by semantic relations (such as "F is an alternative to G", "F is a variant of G by sharing the same set of roles" etc.), and *histories*, that relate a frame to previous or subsequent frames by recording the modifications performed when deriving new frames ("F was derived

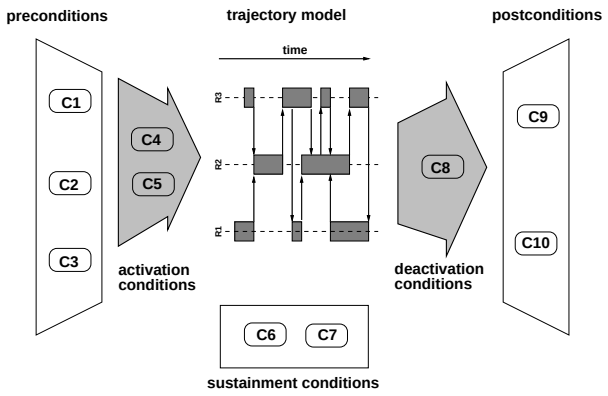

Fig. 3. Context model with embedded trajectory that shows conditions C_i in shaded (relevance) and white (enactment) boxes/arrows.

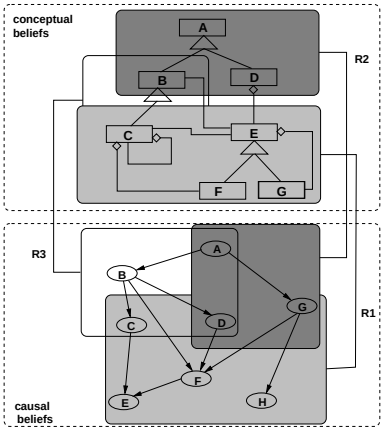

Fig. 4. Two-part belief model with conceptual and causal beliefs. Roles' beliefs are depicted as shaded sub-areas of the networks.

from G by adding precondition C to its context model") are both captured by building up *frame repositories*. These are databases that comprise various frames (in the form shown in Figure 5) linked to each other which are used by the agent (as information that is local to its mental processes) to manage the framing process described in Section 4.

The whole of an agent's frame repository thus constitutes that agent's "social world" [21], i.e. we assume in the following that it is the locus of all *social reasoning* the agent conducts, and that this social reasoning can be clearly distinguished from the agent's *sub-social reasoning capabilities* such as intentional processes like planning, scheduling, deliberating and physical processes like perception and physical action.

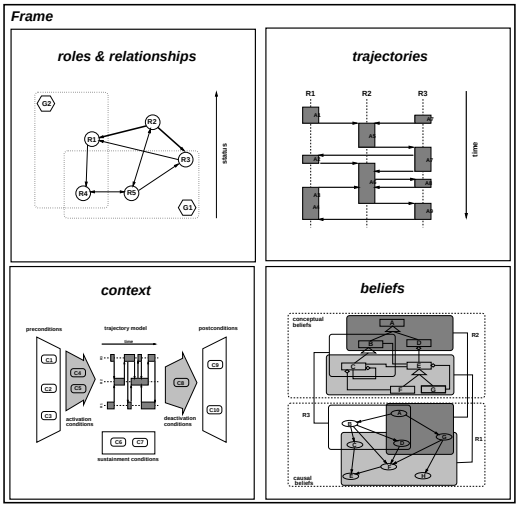

Fig. 5. Integrated frame data structure.

3.2 Private attributes

As mentioned in Section 2.2, frames must not be thought of as *passive* data structures that contain information about certain types of interaction processes. They rather constitute models that both *are constructed by* and *support the* decision-making processes of agents. Therefore, the common attributes that represent shared interactional knowledge must be supplemented with local information about the individual *experiences* and *evaluations* of the agent using them if they are to aid in the process of *organising* social experience. This is achieved by the private attributes section of the frame data structure, which basically contains "status" slots for each of the common attributes. More specifically, these are:

1. *role assignment status*,
2. *trajectory status*,
3. *activation status* and
4. *belief status*.

All of them contain *mappings* for all facts in the respective common attribute and *assessments* concerning the private evaluations of the current state of affairs. To keep things simple, we will not extend the notation of Figure 5, since status data can be simply added to the four slots already introduced.

As for the concrete processes, by which the resulting values of private attributes are determined, they are the result of the *framing* process and depend on the particular architecture employed to achieve framing. One such architecture is introduced in the next section.

4 INFFRA – A social reasoning architecture based on *framing*

Designing a social reasoning architecture based on the notion of *framing* means both describing how frames are constructed by an agent and how they are used in practice. Framing is a very complex activity that involves (1) tracking the enactment of activated frames, (2) choosing whether to retain the current frame or to change frame when appropriate, (3) modifying frame knowledge with experience and (4) relating these three activities to one's private goals in order to make them part of individually rational decision-making.

4.1 Overview

The top-level view of a framing process can be described in a simple way by the following steps that an agent has to perform in each reasoning cycle:

1. **Situation interpretation:** Interpret recent percepts in terms of a *perceived frame*.
2. **Frame matching:** Compare the *perceived frame* with the *activated frame*. Determine a *difference model* describing the results of this comparison.
3. **Framing assessment:** Assess the usability of the active frame.
4. **Framing decision:** If the current frame seems appropriate, continue with 6. Else, proceed with 5.
5. **Re-framing:** Determine a new frame.
   (a) *Search:* Search the *frame repository* for more suitable frames. If candidates are found, proceed with 5(b). Else, proceed with 5(c).
   (b) *"Mock-activation":* Activate one of the candidates above as *trial frame*, go back to 1.
   (c) *Adjustment:* Iteratively modify frames in the *frame repository* and continue with 5(a).
6. **Frame enactment:** Derive commitments that result from the *activated frame*.
7. **Behaviour generation:** Influence action decisions by applying these commitments. Return to 1.

The intuition behind this process is a very simple one: as long as the *activated frame* (the frame currently "in use") seems appropriate given the current interaction situation (the *perceived frame*), it is maintained and influences the behaviour generating processes that the agent employs at the sub-social level (by generating commitments that stem from the frame data). If it fails to match the needs of the situation or those of the agent, the *frame repository* is searched for alternatives, and if none exist, frames are adapted until promising alternatives are found. These are iteratively instantiated as *trial frames* and tested against current conditions ("as if" they had been activated). If a suitable alternative has been found, it is finally activated (if not, the *re-framing* process never terminates).

If successful, this process ensures that the following conditions hold throughout all interactions:

1. Some frame is always active – the agent always has its model of the social context.

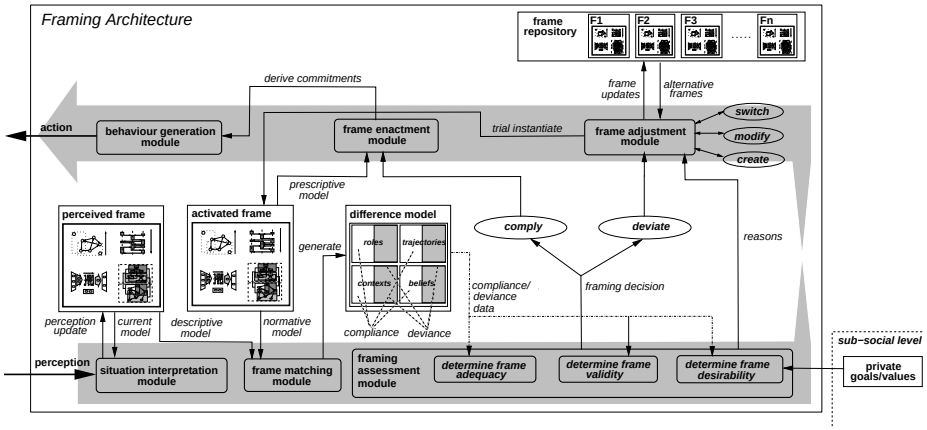

Fig. 6. Detailed view of the framing-based agent architecture. The main line of reasoning between perception and action (shown as a shaded arrow) captures both the sub-processes involved and the temporal order in which they occur.

2. The activated frame is always deemed appropriate both considering the agent's own motives and goals and with respect to the observed course of interaction.
3. The frame repository is only significantly modified when needed[1].

The top-level view of the INFFRA architecture given above has already introduced its central data structures (active frame, perceived frame, trial frame, difference model, frame repository). In the remainder of this section, we will focus on the processing that is performed on them inside INFFRA.

4.2 The architecture in detail

A detailed view of the entire architecture is given in Figure 6. It includes both the data structures just introduced and several functional components that embody the functionality required to conduct the reasoning steps described on page 10. The entire reasoning component must be thought of as a *social-level* reasoner that is added to the mental models and processes that already exist in the agent to manage sub-social activities, e.g. a conventional BDI architecture. The way in which it interacts with these "local reasoning" functions in the *framing assessment*, *re-framing* and *frame enactment* steps will be described in the following paragraphs, which also include details for all the functional modules that appear in Figure 6.

Situation interpretation module This module obtains the current *perceived frame* and incoming percepts as inputs, and outputs a frame that is used as a *descriptive* model of the current situation for matching purposes.

[1] Not counting the tracking of private status attributes that are always added to frame data.

In order to be compared against the active frame, current percepts should always be interpreted in terms of a *perceived frame*, i.e. they should be clustered and classified into roles, trajectories, contexts and beliefs, so that the currently ongoing interaction is captured adequately. A simple implementation might consist of a perception-tracking function that simply records ongoing interaction with respect to a time window of given length. The observed percepts might be categorised into "context" and "interaction", e.g. by classifying certain agents (and the physical environment) as belonging to one of these two categories. For the sake of simplicity, void role and relationship models can be used in first prototypes, since perceived behaviour need not be separated by a role-behaviour/trajectory-behaviour dichotomy for the sake of matching. Likewise, beliefs need not make part of perceived frames in simple implementations – it suffices if they are included in the frames pertaining to the frame repository.

Frame matching module This module is the locus of matching the *perceived frame* (as a descriptive model of interaction "as is") with the *active frame* (a normative picture of what the interaction "should be like"), whereupon it produces the *difference model*, which contains lists of observations that *conform* with the active frame and of percepts that *deviate* from the expected course of action (in the process of re-framing, it compares "trial frames" with the perceived frame). Finally, it outputs frame knowledge of the active frame as *prescriptive* information to the "frame enactment module" which is responsible for biasing decision-making according to the data contained in that frame.

Framing assessment module The assessment module constitutes, together with the adjustment module, the core of the framing architecture. It evaluates the data obtained from the difference model with respect to three measures:
Frame adequacy. This measure computes the degree to which the context model is satisfied by the current situation, which is important in order to determine whether the frame can and should be used in the current context, with two respects: first, failure to meet context preconditions and sustainment conditions jeopardises correct execution of the actions prescribed by the frame; second, the occurrence of deactivation conditions implies initiating a re-framing procedure, and occurrence of activation conditions spawns the adoption of particular frames. Thus, assessing context adequacy is the "first filter pass" that has to be performed prior to any further evaluation of the active frame.
Frame validity. This measure that is computed after the frame has been found to be adequate in the current context expresses to which degree the observed interaction process matches the trajectory data. Mainly, it is used to infer whether the interacting parties meet the expectations induced by the frame, i.e. whether they are "doing the right thing". The result of assessing frame validity should be a detailed description of who fails to comply with which expectation for what reason. Obviously, if this description proves the active frame to be a "false" interpretation of what is going on, re-framing must be attempted.
Frame desirability. Given that the active frame is both adequate and valid, it remains to be evaluated regarding the agent's private goals. As mentioned before, social reasoning is supposed to *serve* individual rationality, so a frame that fails to reflect this rationality must be abandoned. Implementing this functionality implies comparing

1. the post-conditions ensured by enacting the frame,
2. the restrictions the role models of the frame induce on agent behaviour and
3. the actions and events occurring inside the frame trajectory model

to the agent's goals and preferences with in order to maintain a frame only if it seems desirable.

Obviously there are various possibilities for combining these three measures in order to make an appropriate framing decision. In any case, the assessment module should output both the framing decision to the enactment/adjustment module (in the case of comply/deviate, respectively) and the reasons for the re-framing decision to the adjustment model, in case a better frame must be sought for (this information about deviance, unsatisfied private goals, etc. can then be used to better guide the search for a new frame).

Frame adjustment module Frame adjustment (which controls *re-framing* by embodying search, adjustment and mock-activation control) always occurs when the agent has decided to cancel activation of the current frame. Using the reasons for this decision, the frame repository data and operators that allow for *switching* to, *modifying* and *creating* new frames (which are used in this order: first, an existing alternative is sought for, then existing frames are adequately adapted, and only if both procedures fail to produce satisfactory results, an entirely new frame is created) *trial frames* are generated, which are used as "mock instances" for the active frame data structure. Just like the active frame, any *trial frame* can be matched against the perceived frame, and if the difference model resulting from this trial instantiation is reasonably "small", the trial frame becomes the active frame. In terms of machine learning terminology, it can be regarded as the current *learning hypothesis* chosen from the *hypothesis space* the frame repository constitutes, with frame modification operators defining the process of *search* in this hypothesis space.

The variety of algorithms that could be used in this module for frame modification (i.e. the interesting case) is vast: however, it should be possible to represent most of these as special cases of a generic search process using *attribute modification operators* (that alter common attributes of a given repository frame) as search operators for node expansion with prioritisation according to (1) statements in the agent's private goal base and (2) statements that express recently perceived observations.

As concerns switching, a "matching measure" (as used in case-based reasoning algorithms [16]) is one possibility to guide search, and, obviously, generating a completely new frame is the "last option" that can always be applied if the other two mechanisms fail, yet at the risk of not being shared by any peer, since a new frame is not supported by prior experience[2].

Frame enactment module When a framing choice has been made (regardless whether this is the product of frame compliance or that of re-framing), the frame enactment

[2] A point that has to be paid much attention to is, here, the question of how to avoid "infinite trials", by re-searching for new alternatives over and over. It is quite possible that meta-framing reasoning components are needed here, but at present, we have not developed a theory that extends our approach with this respect in a natural way.

module is used to generate commitments for the agent that result from the employed frame. In practice, this means that directives are output to the "behaviour generation module" at the right point in time (according to the temporal model of the trajectory) which inform the agent about social obligations, permissions and prohibitions that it should respect.

Additionally, the frame enactment module fills the private attributes' slots in the active frame by recording the information obtained from the assessment module (the assessment module contains all information about matching results and also all private evaluations that are necessary to derive private attribute values).

So apart from *implementing* the expectations that result from the framing decision, this module also takes over the functionalities needed to record framing experience.

Behaviour generation module This final functional component obtains current social "constraints" on action computed by the frame enactment module and influences the action choices of the agent. Figure 6 and the above description of the architecture seem to suggest that action choices made by the agent at a sub-social level are simply overruled by these social-level decisions, and, indeed, this is the intuition we have in general. However, alternatives to this "strict" overriding of local reasoning can be conceived of, and so we include this module to cater for richer models of combining individual with social choices.

5 Conclusions

The dilemma between applying a *top-down* approach in MAS design by using rich models of sociality and between the appeal of enabling true autonomy that is put forward by proponents of *bottom-up* approaches has lead us to attempt a "middle way" between both views.

On the one hand, we concede that using organisational knowledge that is located above the level of simple interaction protocols and individualistic reasoning mechanisms and that captures social role, action, context and belief models is both useful and necessary if agents are expected to cope with the contingencies of interaction in artificial societies. Yet, on the other hand, our approach re-interprets organisational models as models of *experience* rather than of *top-down design*, thus attempting to focus more on the possibilities of designing how such models are *processed* by the agent rather than how they can be successfully designed by human designers. From this follows that we concentrate on how individual autonomy and social normativity can be successfully combined to yield true social intelligence.

We have identified *frames* and *framing* as useful sociological concepts in the development of a social reasoning architecture: in contrast to top-down sociological theories, these concepts assume a *symbolic interactionist* perspective (Gasser [9, 10] was the first to introduce concepts from this school of research in DAI, cf. his comments in [2]) that allows for a very natural combination of the cognitive and the social level of intelligence – we have underpinned this hypothesis by showing that both notions can be used as starting points for developing realistic computational models.

We contribute to the literature on social reasoning architectures by introducing the framing architecture INFFRA that contains a number of interesting features:

1. a notion of *focus* and *contextuality* in agent interaction (that reduces decision-making complexity by using micro-trajectorial models),
2. a multi-dimensional abstraction from interaction experience by generalising from actors (roles), actions (trajectories), and situations (contexts) and
3. a treatment of social and organisational models as models of experience, inseparably linked to agents' private goals and preferences.

To our knowledge, there exists no social reasoning architecture or methodology that makes use of a combination of these aspects. Also, frames and framing provide concepts that can be used at least at three different levels: for *software engineering* purposes, where designers conduct "soft social design" by designing appropriate frame repositories, for *meta-frame negotiation*, where frames are the target of a negotiation process between agents that aim to find consensus on "how they are going to interact", and, finally, (in the totally open view we have suggested in this paper) for *social learning* as the process of evolving "interaction cases" and shared social and organisational knowledge.

Finally, we have presented a new perspective on *social learning* as an adaptation process that focuses around models of "interaction cases" that are assumed to be *shared knowledge*, which nicely contrasts efforts that aim at "learning about others" in terms of finding out what the other thinks. Although it still remains to be proven that concrete algorithms will successfully tackle the problem of "learning frames" as genuinely social models of action, we believe that transcending the mentalist position is certainly what is needed in future multiagent learning research. INFFRA can be seen as a first step in this direction.

As mentioned in the introduction, it is a long-term scientific challenge to achieve useful and conceptually clear combinations of elements of bottom-up and top-down MAS design. Since our endeavour to develop such theories and systems has just started, it is not surprising that a lot of work lies ahead. Currently, we are working on proof-of-concept implementations that are expected to illustrate the adequacy of our approach. The main challenge that we are currently faced with is to successfully combine the social reasoning architecture we have laid out here with sub-social goal-directed reasoning. In particular, Goffmanian theories do not give *prescriptive* advice of how to employ frames in order to "succeed in social life", and, hence, we have to come up with our own theory of integrating the framing architecture with e.g. BDI agents, game-theoretic utility models, etc. Given the novel perspective on social rationality that INFFRA offers, it has a large potential as a suitable architecture to study this and other issues.

References

1. M. Barbuceanu and M.S. Fox. Capturing and modeling coordination knowledge for multi-agent systems. *International Journal of Cooperative Information Systems*, 5(2–3):275–314, 1996.
2. K. M. Carley and L. Gasser. Computational organization theory. In [22], pages 201–253.
3. C. Castelfranchi. Engineering social order. In *Working Notes of the First International Workshop on Engineering Societies in the Agents' World (ESAW-00)*, 2000.

4. C. Castelfranchi, F. Dignum, C. M. Jonker, and J. Treur. Deliberate normative agents: Principles and architecture. In *Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, Orlando, FL, 1999.
5. R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, London, 1995.
6. R. Conte and C. Castelfranchi. From conventions to prescriptions: Toward an integrated theory of norms. In *Proceedings of the ModelAge'96 Workshop*, Sesimbra, Italy, January 1996.
7. K. S. Decker. TÆMS: A framework for environment centered analysis and design of coordination mechanisms. In G.M.P. O'Hare and N.R. Jennings, editors, *Foundations of Distributed Artificial Intelligence*, pages 429–448. Wiley, New York et al., 1996.
8. J. Ferber and O. Gutknecht. AALAADIN: A meta-model for the analysis and design of organizations in multi-agent systems. Technical Report LIRMM 97189, Université Montpellier II, 1997.
9. L. Gasser. Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence*, 47:107–138, 1991.
10. L. Gasser, C. Braganza, and N. Hermann. MACE: A flexible testbed for distributed AI research. In M. Huhns, editor, *Distributed Artificial Intelligence*, pages 119–152. Pitman, London, 1987.
11. E. Goffman. *Frame Analysis: An Essay on the Organisation of Experience*. Harper and Row, New York, NY, 1974.
12. J. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.
13. S. Kalenka and N. R. Jennings. Socially responsible decision making by autonomous agents. In K. Korta, E. Sosa, and X. Arrazola, editors, *Cognition, Agency and Rationality*, pages 135–149. Kluwer, 1999.
14. E. A. Kendall. Role modelling for agent system analysis, design, and implementation. In *Proceedings of the First International Symposium on Agent Systems and Applications*, 1999.
15. G. H. Mead. *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. University of Chicago Press, Chicago, 1934.
16. T. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.
17. M. Rovatsos. Interaction frames for artificial agents. Report FKI-244-01, Technical University of Munich, 2001.
18. M. Rovatsos and J. Lind. Hierarchical common-sense interaction learning. In E. H. Durfee, editor, *Proceedings of the Fifth International Conference on Multi-Agent Systems (ICMAS-00)*, Boston, MA, 2000.
19. T. W. Sandholm. Distributed rational decision making. In [22], pages 299–330.
20. Y. Shoham and M. Tennenholtz. Social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, 94:139–166, 1995.
21. A. L. Strauss. *Continual Permutations of Actions*. Aldine de Gruyter, New York, NY, 1993.
22. G. Weiß, editor. *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. The MIT Press, Cambridge, MA, 1999.
23. M. J. Wooldridge, N. R. Jennings, and D. Kinny. The Gaia methodology for agent-oriented analysis and design. *Autonomous Agents and Multi-Agent Systems*, 3(3):285–312, 2000.