# Discussion group 'Agents and agency'

Lorentz Center Workshop 'To Be Announced'

August 27, 2015

**As part of the Lorentz Center workshop 'To Be Announced! Synthesis of Epistemic Protocols' (17 - 21 Aug 2015) some themes were addressed in discussion groups. This is the report of the discussion group on the theme 'Agency'. This group was active during the first half of the week.**

*0) [added to the list] What is agency? What are are the constituents of agency?*

The term covers the ability of an agent (and more generally of a group of agents) to achieve a state of affairs, plus the agent's knowledge of that ability; plus perhaps the fact that the action is goal-oriented. The latter can be related to the agent's rationality. Actions might be nondeterministic.

*1) Typically agents are primitive. They are names. Can we define agents by their epistemic properties or their roles?*

It is not enough to identify an agent (better: an agent type) with its current, static epistemic state (a model or a set of models): the definition should also include the agent's disposition to act in particular circumstances. In other words, two identical agents should not only have the same epistemic state, they should also have the same DEL action models (or, more general: actions). The result identifying an agent can therefore either be viewed as the so-called 'forest' of epistemic state plus (possibly iterated) action execution, or, alternatively, as a triple

- the agent's static epistemic state (including higher-order knowledge);
- the agent's update mechanism;
- the agent's goals, possibly represented as protocols = sequences of action models.

Action models are actually too informative: we only need the agent's view of the action model to determine the agent type. More abstractly, we could identify agents with their update mechanisms, abstracting away from their static knowledge. Things get more complex when we want to define groups of agents (however, group agency is computed by intersection of individual agency in logics such as CL, ATL and in STIT theory).

Additionally to the above, a strategy selecting some along all possible actions might further determine an agent (a sub-forest of the forest, a protocol extension).

The discussion veered between the different but related issues of how to uniquely (up to bisimulation) identify an agent by way of a description in terms of what it knows and does (e.g., a characteristic formula of the forest), and how to distinguish agent *types* for which partial descriptions already suffice, and that are helpful in classying large numbers of agents into small numbers of types.

*2) What does agent A need to know about B to be able to reason about B's knowledge? Perhaps better: to successfully reason about B's knowledge ?*

At the most general level, to be able to reason about $B$'s knowledge agent $A$ should know how to apply MP, i.e., the logic should have the K-axiom. (Unless, of course, we model resource-bounded or boundedly rational agents.) In a language having dynamic modalities such as in PAL this might also involve perfect recall, such that one can infer $K_i[p!]K_jp$ (which is equivalent to $K_i(p \to K_j[p!]p)$, in turn equivalent to $K_i(p \to \top)$ ). But all this is not knowledge about $B$ yet: more specifically, agent $A$ also makes the assumption that $B$'s factual knowledge is identical to A's. Moreover, $A$ may also need knowledge about $B$'s goals and $B$'s update mechanism. As far as group agency is concerned, the distinction between different kinds of group knowledge is relevant: distributed knowledge, 'one agent in the group knows', shared knowledge, and common knowledge.

*3) How is agency related to control, to ensuring consequences?*

According to STIT theory, 'being agentive' is just about agents ensuring outcomes (not necessarily so at will, cf. 'hitting the bull's eye'). Agency is then the capability to be agentive.

*4) We assume a fixed set of agents typically. How do we get beyond this, for example, can we assume an arbitrary number of agents, an infinite number of agents, or a dynamically varying number of agents?*

Yoram Moses's example nicely illustrates that knowledge about the boundedness of the set of agents can make a difference. We wondered how one may add or remove an agent in logics such as DEL or ATL, as is required e.g. in game theory when deleting a dominated player. This might have to be done for agent types. $A$ way to do this could be to extend DEL models by appropriate update operations, but the details would have to be worked out and might not be that obvious.

*5) Our models view agents from the outside, how do we incorporate agents' view of the world from inside agents?*

This relates to Guillaume Aucher's distinction between external and internal view. Given a model with an external view one might obtain an external view by dropping

points that are not accessible for the agent under concern. We end up with a mix of an S5 relation (equivalence relation) at the root and otherwise KD45 relations. Beyond, it also relates to knowledge based programs. For example in gossip protocols there are two versions, an external version and one where it is up to the agents to coordinate (either offline or online).