Changing for the better : preference dynamics and agent diversity

Liu, F.

Publication date
2008

Link to publication

Citation for published version (APA):
Liu, F. (2008). *Changing for the better : preference dynamics and agent diversity*. ILLC.

Summary

This thesis investigates two main issues concerning the behavior of rational agents, preference dynamics and agent diversity.

We take up two questions left aside by von Wright, and later also the multitude of his successors, in his seminal book *Logic of Preference* in 1963: *reasons* for preference, and *changes* in preference. Various notions of preference are discussed, compared and further correlated in the thesis. In particular, we concentrate on extrinsic preference. Contrary to intrinsic preference, extrinsic preference is reason-based, i.e. one's preference for one option over another has a reason. A logical model is proposed and its properties are determined. Dynamics come in naturally, since reasons for the preferences can change. Logical systems and formal results regarding dynamical preference change are then presented.

Preference arises from comparisons between alternatives. A first option is to compare situations. Abstractly speaking, preferences are in this case between propositions, viewed as sets of possible worlds. The reasons can then be based on a 'betterness' relation over possible worlds. Propositional preference arises as a lift from this primitive relation. A standard modal logical approach is taken and we use a modality for the betterness relation in the language. We then model preference change by techniques from dynamic epistemic logic (DEL), where a typical action, e.g. a suggestion or a command can change the betterness ordering of the worlds, and thereby the propositional preference. Dynamic reduction axioms are obtained to encode exactly how such a change takes place. We obtain a complete dynamic preference logic.

A second option is to compare objects as such. Concretely, properties of the objects often determine the preference over the objects. Properties are now the reasons. Inspired by Optimality Theory (OT), we propose a *priority sequence*, an ordering of properties. Various ways of getting a preference from the priority sequence are investigated, though we mostly follow the OT approach. We use a fragment of first-order logic to describe the situation. Here, on the dynamic side, it is priority change that leads to preference change. Using the DEL methodology

again we propose a complete set of reduction axioms concerning the possible dynamic operations on priority sequences.

Not surprisingly, the above two views are closely related. After all, possible worlds can be thought of as objects! On the basis of a systematic comparison of the two views, we develop a *two level perspective*, in which the models themselves are structured in layers. In particular, correspondence results between the changes at the level of the possible worlds and the changes at the level of the priority sequences are proved. We end up by sketching a two-level preferential predicate logic to describe more complex circumstances in which situations and objects are compared simultaneously.

But we do not see this as the whole story. Preference does not live by itself, it is often intermingled with epistemic notions of knowledge and belief. One can have different intuitions about how this entanglement operates. A few options are discussed, and proposals for logical models are presented. When moving to dynamics, we now see a picture of knowledge update, belief revision and preference change taking place symbiotically, often unconsciously as in real life.

The resulting picture in the thesis is one of agents that process information and adjust beliefs and preferences in many different ways. There is no logically prescribed unique norm for doing this. The second part of the thesis takes this general phenomenon of diversity of agents as its focus, since it raises many issues for logical systems and the idealized agents which they normally presuppose. In reality, agents can differ across a wide spectrum of cognitive abilities and habits: in their memory capacity, observation power, inferential power, introspective ability, and revision policies when facing new information.

Two kinds of agents, perfect recall agents, and memory-free agents are studied thoroughly, in the setting of playing games. We show how current dynamic logics can be 'parametrized' to allow for this memory diversity, with new characterizations of agent types resulting in complete dynamic-epistemic logics. The other dimension of the reality of diverse agents is that, however different they are, they often do manage to coordinate with each other successfully. While this theme has been prominent in game theory and multi-agent systems, it has received hardly any attention in logic. We analyze this interaction between different agents by looking at concrete scenarios which model their types explicitly.

Finally, we analyze the issue of agent diversity in its generality, discussing what dynamic logics would have to look like to become a full-fledged account of agents of different capacities and tendencies that pursue and sometimes achieve their goals in irreducibly social settings.