# Changing for the better : preference dynamics and agent diversity

Liu, F.

**Publication date**
2008

# Chapter 7

# Conclusions and Future Work

### Conclusions

This dissertation has started from two issues concerning the functioning of rational agents that have been largely left aside since [Wri63]: *reasons* for preference, and *changes* in preference. Extrinsic reason-based preference was chosen as our main topic, and two models have been proposed for it in Chapters 2 and 3, respectively. Those models differ in their point of departure: object comparison versus priority order of propositions, but they have a common feature, in that preference and reasons come together.

In Chapter 2, I have shown how preference over propositions, derived from a primitive betterness relation over possible worlds, can be studied with techniques from dynamic epistemic logic. In particular, dynamic reduction axioms encode exactly how propositional preferences change when some new evaluative trigger such as a suggestion or command changes the betterness order. This brings a new methodology to traditional preference logic, while at the same time extending the scope of $DEL$.

In Chapter 3, preference over objects was studied in a fragment of first-order logic $FOL$. Preference over objects is now derived from a primitive base order of propositions in a priority sequence. This shows how logic can deal with basic ideas from Optimality Theory and related areas of 'optimal choice' in computer science and the social sciences. Moreover, I have shown how this, too, is compatible with $DEL$ methodology, proposing dynamic operations on priority sequences with a complete set of reduction axioms.

In Chapter 4, a comparison between the two models of Chapters 2 and 3 showed that they are systematically related, and that they may fit together in various elegant mathematical ways. One example is a view of preference definition and preference change as related to more general preference merge between orderings coming from different sources. Another example is a grand two- or even three-level doxastic preferential predicate logic that can deal with the various notions of preference encountered in our intuitive daily reasoning.

In both Chapters 2 and 3, I have also shown how static preference representation and preference dynamics can live together with epistemic and doxastic structure, thereby doing justice to the intuitive entanglement of preference with knowledge and belief. I have brought these strands together in Chapter 4, showing how this all fits with a sequence of modal logics describing various 'degrees of entanglement'. This also allowed for further connections with belief revision theory, although we have by no means exhausted this analogy. Cf. the dissertation [Gir08] for a complementary agenda of logical themes at this rich interface.

In Chapters 5 and 6, I have then developed a logical perspective on much more general diversity of agents, trying to locate all aspects in which they can differ. Chapter 5 high-lights the dramatic difference between agents with perfect recall and agents with bounded memory. I have shown how both can be captured in dynamic epistemic logics, thereby dispelling the idea that $DEL$ can only account for idealized agents, and making for connections with the theory of games with imperfect information. This diversity is then extended to logics defining policies for belief revision in both Chapters 5 and 6.

Thus I have proposed the basic ingredients for dynamic logics of agents with information update, belief revision, and preference upgrade. When adding these together in realistic settings, one must analyze the *interactions* of diverse agents. At the end of Chapter 6, I make a first step in this direction. I analyze a couple of scenarios in which different types of agents interact with each other, and propose a model for analyzing these.

**Future work**

This thesis proposes a rich model of diverse preference-driven rational agents based on dynamic logic. In doing so, many new questions have arisen, which have been noted along the way.

Some obvious open problems within the logical sphere are links between our various systems that are yet to be developed. First, we need to understand combined systems incorporating object relation transformers and constraint dynamics in greater generality, and Chapter 4 contained many leads in this direction. Next, the transition from Chapters 2, 3, 4 to the themes in Chapters 5, 6 suggests a more systematic merge of limitations on information dynamics (memory, inference, observation) and similar limitations on preference dynamics. To put it briefly, *how to model preference change for bounded agents?* We believe that our thesis supplies the right ingredients for doing so, but we have not done it yet.

Next, I have mostly considered preferences for single agents, while rational agency clearly involves *groups*. The current framework extends easily to interactive multi-agent systems in a purely formal manner. In Chapter 3 we made a start in investigating concepts of cooperation and competition by interpreting them in terms of reasons for preferences of the different agents. We also made a brief excursion on preference merge, and hence 'group preferences' in Chapter 4.

It should also be noted that preference change does not just involve myopic single steps. It often takes place in longer scenarios over *time*. To fully understand the temporal dynamics of preference, we need to integrate time into the current framework, as has been done for dynamic epistemic and doxastic logic in [BP06], [Bon07], and other publications.

Finally, going beyond the narrower ILLC world of logic and computation, there are also evident broader questions relating our present logical framework to other approaches.

In particular, I have adopted a *qualitative* approach to preference representation. But in areas like decision theory and social choice theory, usually, numerical utility functions represent preference. And likewise, for modeling beliefs under uncertainties, numerical probabilities are used widely. In the area of belief revision, and to some extent also $DEL$ these days, this is a well-known interface. Can the logical systems for preference proposed in this thesis support *quantitative* utilities in a natural manner? I made a first attempt in [Liu06b], using $DEL$ methodology to upgrade numerical 'plausibility values', using ideas from [Auc03], but much more remains to be done.

Moreover, this thesis has provided quite abstract models of reasons for preference and changes in preference. *Applying* these models to concrete scenarios in areas like decision theory and game theory seems a reasonable test for our proposals. For instance, in games, a player may have an initial preference over moves, but then, observing what her opponent plays may make her change her mind. In such a scenario, both preference and beliefs play a role, often at the same time (see [Ben06b]). Such considerations also affect how players compare propositions about the future course of a game. Thus, the usual solution procedure of Backward Induction involves a mixture of relative plausibility and preference between outcomes, as has been pointed out in [Ben02] and [DZ07]. Likewise, our models should be confronted with those in the philosophy of action, where preference supports rationality. The dissertation [Roy08] takes static preference logic and $DEL$-style dynamics to this arena in modeling information update and intentions, but it does not yet contain a full-fledged account of belief revision and preference change.

This concludes our summary of what this thesis has done, and what may be, and perhaps should be, done next.