# Changing for the better : preference dynamics and agent diversity

Liu, F.

**Publication date**
2008

# Chapter 1

# Introduction

Preference is what colors our view of the world, and it drives the actions that we take in it. Moreover, we influence each other's preferences all the time by making evaluative statements, uttering requests, commands, and statements of fact that exclude or open up the possibility of certain actions. A phenomenon of this wide importance has naturally been studied in many disciplines, especially in philosophy and the social sciences. This dissertation takes a formal point of view, being devoted to logical systems that describe preferences, changes in preference, and behaviors of different agents in dynamic contexts. I will plunge right in, and immediately draw your attention to the first time when preference was fully discussed by a logician.

### Preference logic in the literature

**What von Wright considered, and what he did not** In his seminal book *The Logic of Preference: An Essay* from 1963, von Wright started with a major division among the concepts that interest moral philosophers. He divided them into the following three categories (though there may be border-line cases):

- *deontological* or *normative*: notions of right and duty, command, permission and prohibition,
- *axiological*: notions of good and evil, the comparative notion of betterness,
- *anthropological*: notions of need and want, decision and choice, motive, end and action.

The intuitive concept of preference itself was said to 'stand between the two groups of concepts': It is related to the axiological notion of betterness on one side, but it is related just as well to the anthropological notion of choice.

While considering the relationship between preference and betterness, von Wright distinguished two kinds of preference relations: *extrinsic* and *intrinsic* ones. He explains the difference with the following example:

> "... a person says, for example, that he prefers claret to hock, because his doctor has told him or he has found from experience that the first wine is better for his stomach or health in general. In this case a *judgement of betterness serves as a ground or reason* for a preference. I shall call preferences, which hold this relationship to betterness, *extrinsic*.
>
> It could, however, also be the case that a person prefers claret to hock, not because he thinks (opines) that the first wine is better for him, but simply because he likes the first better (more). Then his liking the one wine better is not a reason for his preference. ..."
>
> ([Wri63], p.14)

Simply stated, the difference is principally that $p$ is preferred *extrinsically* to $q$ if it is preferred *because* it is better in some explicit respect. If there is no such reason, the preference is intrinsic.

Instead of making the notion of betterness the starting-point of his inquiry,[1] von Wright took a more "primitive" intrinsic notion of preference as 'the point of departure', providing a formal system for it which has generated a whole subsequent literature (cf. [Han01a]).

We are by no means claiming that the division between intrinsic and extrinsic preference is the only natural way of distinguishing preferences. One can also study varieties of moral preference, aesthetic preference, economic preference, etc. However, in this thesis, I will follow von Wright's distinction. Our first main goal is to extend the literature on intrinsic preferences with formal logical systems for the *extrinsic notion of preference*, allowing us to spell out the reasons for a preference. On the way there, we will also make new contributions to the literature on intrinsic preferences.

Besides the extrinsic notion of preference that was removed from von Wright's agenda, there is another important issue which he left open. More precisely, he writes the following:

> "The preferences which we shall study are a subject's intrinsic preferences on one occasion only. Thus we exclude both *reasons* for preferences and the possibility of *changes* in preferences."
>
> ([Wri63], p.23)

Clearly, our preferences are not static! One may *revise* one's preferences for many legitimate (and non-legitimate) reasons. The second main issue dealt with in this thesis is how to model preference change in formal logics. This leads to new dynamic versions of existing preference logics, and interesting connections with belief revision theory.

[1] [Hal57] did propose logic systems for the notion of betterness.

**What others considered afterwards, and what they did not** Following von Wright's work, many studies on preference were carried out over the last few decades. Due to its central character, at the interface between evaluation, choice, action, moral reasoning, and games, preference has become a core research theme in many fields, which have often led to logical theory. In what follows I will summarize the main issues or directions taken by other researchers. My purpose is not to give an overview of the vast literature (I give some basic references for that), but only to point out some issues that are relevant to the present thesis, and some particular proposals that have inspired it.

**Preference in logic and philosophy** Formal investigations on preference logic have been mainly carried out by philosophical logicians. The best survey up to 2001 can be found in the Chapter *Preference Logic* by Sven Ove Hansson in the *Handbook of Philosophical Logic*.

This literature added several important notions to von Wright's original setting. In particular, a distinction which has played an important role is that between preference over *incompatible* alternatives and preference over *compatible* alternatives, based on early discussions in [Wri72]. The former is over *mutually exclusive* alternatives, while the latter does not obey this restriction. Here is a typical example:

> "In a discussion on musical pieces, someone may express preferences for orchestral music over chamber music, and also for Baroque over Romantic music. We may then ask her how she rates Baroque chamber music versus orchestral music from the Romantic period. Assuming that these comparisons are all covered by one and the same preference relation, some of the relata of this preference relation are not mutually exclusive."
>
> ([Han01a], p.346-347)

Most philosophical logicians have concentrated on exclusionary preferences. However, in this thesis we will consider both. As we will see, one of our logical systems is for preference over objects, which are naturally considered as exclusive incompatible alternatives. But we will also work with preferences between propositions, which can be compatible, and indeed stand in many diverse relationships.

Also, most researchers have been particularly interested in the question whether certain *principles* or 'structural properties' are reasonable for preference. Here economists joined logicians, to discuss the axioms of rational preference. Many interesting examples have been proposed to argue for or against certain formal principles, resulting in different logical systems (cf. [Tve69], [Sch75], [Lee84], etc.). However, a general critical result in [Han68] is worth being noticed. In this paper, the author showed that many axioms proposed for a general theory of preference imply theorems which are too strange to be acceptable. But it is often possible to

restrict their domain of application to make them more plausible. In general, our logical systems will not take a strong stand on structural properties of preference, beyond the bare minimum of reflexivity and transitivity (though we note that the latter has been questioned, too: Cf. [Hug80], [Fis99]).

There are also obvious relationships between preference and *moral* or more generally, *evaluative* notions like "good" and "bad". Several researchers have suggested definitions for "good" and "bad" in terms of the dyadic predicate "better". A widespread idea is to define "good" as "better than its negation" and "bad" as "worse that its negation", as in [Wri63] and [Hal57].[2] Alternatively, [CS66b] presents indifference-related definitions for "good" and "bad", and then defines things as follows: "a state of affairs is good provided it is better than some state of affairs that is indifferent, and ... a state of affairs is bad provided some state of affairs that is indifferent is better than it". [Han90a] generalized the previous proposals, and presented a set of logical properties for "good" and "bad". Interestingly, precisely the opposite view has been defended in the logical literature on semantics of natural language. [Ben82] defines binary comparatives like "better" in terms of context-dependent predicates "good", and [Roo07] takes this much further into a general analysis of comparative relations as based on a 'satisfying' view of achieving outcomes of actions.[3] Either way, we will not pursue this particular line of analysis in this thesis, although one might say that our later analysis of preference as based on constraints has some echoes of the linguistic strategy deriving binary comparatives from unary properties.

The connection between preference and moral reasoning is clear in *deontic logic*, another branch of philosophical logic going back to von Wright's work, this time to [Wri51]. While obligation is usually explained as truth in all 'deontically accessible worlds', the latter are really the 'best worlds' in some moral comparison relation. Not surprisingly, then, preference relations were introduced in standard deontic logic to interpret conditional obligations. For modern preference-based deontic logics, see [Han90b], [Tor97]. Preference was introduced particularly to help solve some of the persistent 'deontic paradoxes'. Here are a few examples: [CS66a] gave a moral deontic interpretation of the calculus of intrinsic preference, to solve the *problem of supererogation* - 'acting beyond the call of duty'.[4] [TT98] extended the existing temporal analysis of Chisholm's Paradox of conditional obligation (see [Eck82], too) using a deontic logic that combines temporal and preferential notions. Also, [TT99] provided better solutions to many paradoxes by combining preferential notions with *dynamic updates*: making this dynamics even more explicit will be one of our main themes.

[2] Quantitative versions of these ideas are found in [Len83].

[3] It would be of interest to contrast their formal 'context-crossing principles' with Hansson's proposals.

[4] Non-obligatory well-doing is traditionally called supererogation. Many of the great deeds of saints and heroes are supererogatory.

**Preference in decision theory and game theory** The notion of preference is also central to decision theory and game theory: given a set of feasible actions, a rational agent or player compares their outcomes, and takes the action that leads to the outcome which she most prefers. Typically, to make this work, outcomes are labeled by quantitative utility functions - though there are also foundational studies based on qualitative preference ordering ([Han68]). Moving back to logic, [Res66] brought together the concepts of preference, *utility* and of *cost* that play a key role in the theoretical foundations of economics, studying primarily the metric aspect of these concepts, and the possibility of measuring them. For modern discussions in this line, see [Bol83] and [Tra85]. In terms of axiomatization, the standard approach takes weak preference ("better or equal in value to") as a primitive relation, witness [Han68] and [Sen71].

In particular, economists have studied connections between *preference* and *choice* ([Sen71], [Sen73]), treating *preference* as almost identical with *choice*. Preference is considered to be 'hypothetical choice', and choice to be *revealed preference*. Recently, revealed preference has become prominent in understanding the concept of equilibrium in game theory (cf. [HK02]). Differently from standard logical models, preference is then attached to observed outcomes. Preferences of players have to be constructed, so that the observed outcomes can be rationalized by the chosen equilibrium notion employing these constructed preferences.

But, one has to be careful with such identifications of notions across different fields. Preference is not really the same as choice. Many researchers have remarked on that. Already in [Wri63], it was pointed out that 'it is obvious that there can exist intrinsic preferences, even when there is no question of *actually* choosing between things."([Wri63], p.15). Choice must involve actual action, but preference need not. In this thesis, we will not pursue the connection between preference and its emergence in general action, though our dynamic framework for describing preference change can presumably be extended to deal with the latter scenario.[5]

**Preference in computer science and AI** From the 1980s onward, and especially through the 1990s, researchers in computer science and AI have started paying attention to preference as well. Their motivations are clear: 'agents' are central to modern notions of computation, and agents reason frequently about their preferences, desires, and goals. Thus, representing preferences and goals for decision-theoretic planning has become of central significance. For instance, [CL90] studied general principles that govern agents' reasoning in terms of their belief, goals and actions and intentions. The well-known '$BDI$ model' was first presented in [RG91] to show how different types of rational agents can be modeled by imposing conditions on the persistence of an agent's beliefs, desires or

[5]A related area of formal studies into preferences for agents, and how these can be merged, is *Social Choice Theory*: Cf. [Fis73].

intentions, and its further development can be found in [LHM96], [HW03], and [Woo00]. Other work in qualitative decision theory illustrates how planning agents are driven by goals (defined as desires together with commitments) performing sequences of actions to achieve these (cf. [Bou94], [DT99], [Tho00]). Of interest to logicians, general properties of the language of preference representation have become important, such as striking a balance between expressive power and succinctness (see [CMLLM04] and [CEL06]).[6]

Further occurrences of preference are found in the AI literature on common sense reasoning, witness the treatment of circumscription, time, 'inertia', and causality in [Sho88]. Interestingly, further crucial notions from von Wright have made their way directly into this literature. In particular, his idea that preferences can often only be stated *ceteris paribus* has been taken up in [DSW91] and [DW94], which studied preference "all else being equal". The other main sense of 'ceteris paribus', as "all else being normal", was taken up in [Bou93], where preference relations are based on what happens in the most likely or "normal" worlds. A recent development of ceteris paribus preference in a modal logic framework is [BRG07]. The eventual systems of our thesis can deal with the latter, though not (yet) with the former.

### Some specific influences on this dissertation

In terms of new methods for the logic of preferences, we now mention a few sources here that have influenced this dissertation. The authors in [LTW03] propose a logic of desires whose semantics contains two ordering relations of preference and normality, respectively. They then interpret desires as follows: "in context $A$, I desire $B$" iff "the best among the most normal $A \wedge B$ worlds are preferred to the most normal $A \wedge \neg B$ worlds". Such combinations are typical of what we will deal with eventually. But before getting to these entangled scenarios, we also employ tools from straight preference logic, in particular, the *modal preference logics* proposed by [Bou94], and following him, [Hal97]. Halpern started with just a betterness ordering over possible worlds, and showed how to extend this to sets of possible worlds. He then gave a complete axiomatization of this logic over partial orders. This sets the model for the basic 'static' completeness results we will need later.

But there are yet more influences on our work from the computational literature. One obvious one is *propositional dynamic logic* for sequential programs and general actions ([HKT00]), which will be our main model for describing the dynamics of preference change. Our semantics and complete logics will follow especially the modern format of *dynamic epistemic logic* ([DHK07]). We will

[6] Indeed, preferences are also found in the more 'hard core' theory of computation, e.g., in describing evolutions of computational systems, which need to be compared as to some measure of 'goodness'. Substantial examples of this trend are [Mey96] on dynamic logic with preference between state transitions, and [Ser04] on a general calculus of system evolution.

elaborate on this paradigm in more detail below, and in the main body of the thesis. But there are even further sources. Interestingly, the recent computational literature also takes up themes from social choice theory, such as *aggregation of preferences*, as a matter of crucial interest to describing the behavior of societies of agents, cooperative or competing. One sophisticated study of this sort, which brings together social choice theory, preference logic, and algebraic logic, is [ARS02]. We will use their techniques for preference merge triggered by hierarchies of agents to shed light on the array of notions involved in intrinsic and extrinsic preferences.

This concludes our survey of major developments in preference logics as relevant to this thesis. The account is by no means complete, however. For instance, many further connections between preference, belief revision, and the foundations of *economics* are found in [Rot01]. And also, it will soon be clear that our treatment of extrinsic preferences, generated by further outside considerations, also owes much to *linguistics*, viz. the area of Optimality Theory ([PS93]), which describes grammatical sentences and successful utterances in a rule-free manner, in terms of optimal satisfaction of syntactic, semantic, and pragmatic *constraints*. How constraints induce preference, and how they can enter preference logic, will be a major theme in what follows.[7] But for now, we summarize where we stand.

Our starting point is the preference logic of von Wright, and some major distinctions that he made. We identified two major issues that [Wri63] left out, viz. *reason-based extrinsic preference*, and the *dynamics of preference change*. Of course, we are not claiming that nobody paid any attention to these two issues over the past decades. But it does seem fair to say that most authors took the notion of intrinsic preference only, and concentrated on its properties.[8] Next, we have only found a few papers treating changes in preference as such. [BEF93] is a first attempt at using dynamic logic for this purpose. Also, influenced by *AGM*-style belief revision theory, [Han95] proposed postulates for four basic operations in preference change.

Against this background, this thesis will show how these two crucial aspects of reasoning with preference can be treated in a uniform logical framework, which borrows ideas from several different areas: (a) the subsequent development of preference logic, (b) the computational literature on agents, (c) linguistic optimality theory, and (d) recent developments in the theory of belief revision and dynamic epistemic logic.

Having reviewed what has been done by others, here is what is new in this thesis. Basically, I will study a number of old issues that are still open, and a

[7] These ideas are even extended into models for brain function in cognitive science (cf. [Smo04]).

[8] Still, 'reasons for preference' are a theme in decision theory and economics, witness the brief survey in [HGY06].

few new issues that have not yet been considered. Also, I will study these issues only from a *formal logical point of view*. In what follows I introduce my guiding intuitions, and the main ideas.

### On intuitions and ideas

**Reasons for preference** In many situations, it is quite natural to ask for a reason when someone states her preference to you. It may be a matter of justification for her, but as for you, you simply want more explanation or information (sometimes, in order to judge whether it is rational for her to have that preference). So preference can come with a reason, and this is what von Wright called 'extrinsic preference'. Let us return to the example used by [Wri63] to explain this notion:

> A person prefers claret to hock, *because* his doctor has told him or he has found from experience that the first wine is better for his stomach or health in general.

Here, *the first wine being better for his health* is the reason for his preference of claret to hock. Similar examples abound in real life: one prefers some house over another *because* the first is cheaper and of better quality than the second.

Conceptually, reasons stand at a different level from preferences, and they form a *base* or *ground* for their justification. Reasons can be of various kinds: from general principles to more 'object-oriented' facts. In many cases, one can combine more than one reason to justify one single preference. Thus, in the house example, not only the price of the house matters, but also the quality. In such cases, reasons may have their own structure, and different considerations may be ordered according to their importance. One may think for instance that the quality of a house is more important than its price.

**Preference change** There is more to be said about the above example. Let us first add a twist of imagination to make it dynamical:

> Suppose that before he sees the doctor, he *preferred hock to claret*. Now the doctor tells him "the first wine is better for your health". He then *changes* his preference, and will now *prefer claret to hock*!

Again such things often occur in real life. We change our preferences on the basis of new information that we have received. And the new preference emerges for a new reason. Actually, this way of thinking immediately links us to information dynamics in general. Accordingly, I will use the methodology of modeling information dynamics to deal with preference change in this dissertation. A few more words are in order here. The idea behind information dynamics is this: agents receive new information and update their knowledge or beliefs accordingly. This

style of thinking can be traced back to the early 1980s, e.g., the well-known $AGM$ postulates handling belief change ([AGM85]). But the approach I am taking here is what recently developed under the name of *dynamic epistemic logic* ($DEL$). It has a certain Amsterdam flavor, which inspired me through the following works: [Pla89], [Vel96], [Ben96], [BMS98], [Ger99], and [DHK07], as well as up-to-date work on belief revision by [Ben07a] and [BS08]. Readers will see in the later chapters how I apply techniques from these works to the dynamics of preference. This choice of approach also distinguishes my proposals from the $AGM$-style preference change presented in [Han95].

We know that reasons and preferences live at different levels. Moreover, reasons provide an explanation for preference. Thus one can travel between the two levels, as reasons lead to a preference, and preference can be seen as derived from reasons. Since dynamics can take place at both levels, we will also investigate how to relate the changes at the two levels to each other.

**Beliefs as a reason, too** There is one issue we have not yet considered in the above, namely, *uncertainties*. When someone tries to give a reason for her preference, in some situations, she may not have precise information to offer. Instead, she may say things like 'I believe that it is going to rain, so I prefer bringing my umbrella'. Under such circumstances, one's preference relies on one's *beliefs*, and beliefs come in as an extra reason for preference. People may have different preferences *because* they have different beliefs. Thus the notion of preference becomes richer, and similarly changes in preference as well: preference change may now also be caused by belief change.

The literature on preference logic has not considered intertwined belief and preference yet. But such entanglements are standard in other areas, in particular, decision theory, which has a tradition in modeling decision making under uncertainty ([Sav54], [Jef65]). Here most models rely on a numerical representation where utility and uncertainty are commensurate. For instance, an agent may not know the outcomes of his actions, but may use a probability distribution over outcomes instead. The expected value of an action can be then computed from utility and probability, as explained in any textbook. What is relevant to our preceding discussion is this. The main reason to represent worlds probabilistically in decision theory is to be able to use the *beliefs* as a base for decision making. By contrast, we will use beliefs as well, but mostly in a qualitative approach without numerical calculations.[9]

**We are diverse human beings** Preferences notoriously differ, and this variety seems typical of human behavior and interaction. But this diversity extends to other features of agent behavior. For instance, consider the reasons people have for preferences, and the ways these might change. Here, too, different people

[9]We refer to [Liu06b] for some numerical counterparts to our qualitative proposals.

may react quite differently. In particular, when belief is involved, this naturally leads to various policies for changing beliefs, a diversity which is at the heart of belief revision theory. For instance, a 'radical agent' may change her preference immediately, taking her reasons from some partial information received, whereas a 'conservative agent' will stick to past beliefs and past preferences for longer, waiting for more input. In addition to these differences in preference, and belief policies, agents may also have differences in their even more basic logical capacities for information handling: in particular, their memory capacity, and tendencies to forget crucial information obtained earlier.

This diversity of agent behavior seems an essential fact of life to us, and one of the most striking features human interaction is how it still leads to coordinated, and often very successful behavior. Such phenomena have been studied to some extent in belief revision theory, witness the host of belief revision policies in [Rot06]. Another area of diversity studies is in models for inferential information and computational restrictions on agents abilities (cf. [BM07], [Egr04], [BE07]). But these aspects of diversity have not yet been studied in their totality, and we will make an attempt in this thesis to provide a more comprehensive model of agents whose preferences, beliefs, and information may vary, as well as the dynamic rules which change these.

Given these considerations, the challenge is to understand how, despite our differences, we live in one society, interacting with each other successfully. The following questions then arise:

- What major aspects can agents differ in?
- How differently do they update their knowledge and beliefs when facing new information?
- How do they interact with each other, say in games, despite these differences - say, by learning each other's 'type' of behaviour?

These questions have come up in several areas. For instance, game theorists have studied 'bounded rationality' in the study of cooperative behavior([OR94], [Axe84]), while, as we said, formal epistemologists have tried to parameterize agents' inferential or computational powers. Moreover, the variety of human behavior versus idealized norms has been emphasized in the study of reasoning in cognitive psychology ([Gol05], [HHB07]).

But more in particular, the preceding questions pose a serious challenge to the dynamic logics for preference change and belief that we have developed. Do they leave room for significant differences in agent behavior across the appropriate range of variation that can be found in practice? It may seem that they do not, since 'the valid reduction axioms' for knowledge after update seem written in stone. Even so, this thesis will show that dynamic logics do allow for the proper variation, by providing formal logical models for variety inside dynamic epistemic

logic, which address the preceding issues as well as others. We will relate them to the study of games, and general processes in a temporal setting.

### Connections to related areas

As stated at the outset, this thesis is squarely within the logical tradition. Nevertheless, beyond obvious comparisons to be made with the older literature on preference logic, I believe that my results may be of interest to some of the other areas mentioned here. For instance, the qualitative perspective on preference and preference change may be of interest to decision theorists looking for qualitative models. Likewise, since preference comes with its own intuitions, dynamic logics of preference can be inspirational for dynamic logics of beliefs. A case in point is [BS08] whose account of new belief modalities and reduction axioms for them was influenced by [BL07]. Furthermore, since the models proposed here are abstract and general, they can be applied to neighbors of preference logic such as deontic logic. I believe that norm change and obligation change can be modeled in a similar way to preference change, and indeed, a number of such studies have been made, including [Zar03], and in particular, a recent series of papers by Tomoyuki Yamada, of which [Yam07] is a representative sample. Indeed, vice versa, their work has also influenced mine. Finally, in the philosophy of action, our treatment of the difference between intrinsic preference and extrinsic preference may provide a synthesis between so-called "recognitional" and "constructivist" views of practical reasoning.[10] Our two notions of preference explain such a difference in a precise way. For further connections between preference logic and the philosophy of action, see the two dissertations [Gir08], [Roy08] which touch this one at various points mentioned in subsequent chapters.

Finally, I will briefly state the structure of the thesis in slightly more technical terms, showing how my intuitions and ideas are formalized in logics.

### Structure of the thesis

This thesis is organized as follows:

In Chapter 2, a first model for extrinsic preferences is proposed. Models consist of a universe of possible worlds, representing the different relevant situations, endowed with a basic objective order of 'betterness'. The latter supplies 'reasons for preference' when we 'lift' this order to one among propositions, viewed as sets of possible worlds.[11] There are various kinds of 'lifting', of which we consider in

[10] According to the "recognitional" view, rational practical reasoning consists in trying to figure out which of the available options are good things to do, and then choosing accordingly. According to the "constructivist" view, rational practical reasoning consists in complying with certain conditions of purely formal coherence or procedural rationality. For more details on the debate, see [Wed03].

[11] Note that betterness is a preference over incompatible alternatives.

particular the $\forall\exists$-version saying that every $\varphi$ world has at least one better $\psi$ alternative world. These lifts, and many other types of statement can be described in a standard modal language over betterness models. As for the dynamics of preference change, this is triggered as follows. Statements like suggestions or commands 'upgrade' agents' current preferences by changing the current betterness order among worlds. A complete logic of knowledge update plus preference upgrade is presented that works with dynamic-epistemic-style reduction axioms. The result is an intertwined account of changing preferences and also changing knowledge as triggered by factual information. This system can also model changing obligations, conflicting commands, or 'regret' about possibilities that have already been ruled out epistemically. Beyond specific examples, we present a general format of relation transformers for which dynamic-epistemic reduction axioms can be derived automatically.

Chapter 3 provides a second model for extrinsic preferences. This time, the aim is to analyze preferences over objects, again, comparing incompatible alternatives. For this purpose, inspired by linguistic optimality theory, the primary structure is an ordered 'priority sequence' of 'constraints', i.e., relevant properties of objects. It supplies reasons for preference by comparing objects as to the properties they have or lack in this sequence. Typically, the relationship between reasons and the resulting extrinsic preference is characterized by so called 'representation theorems' in this chapter.[12] Intuitively, these results say that one can always find a reason for some given object preference. Next, in the realistic case where agents only have incomplete information, here, too, we add epistemic structure. In particular, we introduce beliefs that help form preferences. Three definitions are proposed to describe how different kinds of agents get their preference under uncertainties. Changes of preference are then explored with two different reasons: either changes in the priority sequence, and also through belief change. Both can lead to preference change.

In Chapter 4, I primarily draw a comparison between the two approaches in Chapters 2 and 3, both qua semantics and qua syntax. First, abstract *structured models* are introduced to merge 'reasons' (a set of ordered propositions) and a correlated 'betterness order' over possible worlds. I then study general ways of deriving world preferences from an ordered set of propositions, as well as the opposite direction: ways of lifting a world preference relation to an ordering over propositions. Interestingly, when we go back and forth between these, we find several tight correspondences between concrete order-changing operations at the two levels, and some specific definability results are proved. The general context behind these are partially ordered 'priority graphs' from the literature on preference merge, which seem the most elegant mathematical framework behind our specific proposals. We prove definability results at this level, too, and draw a comparison with 'priority product update' in the dynamic epistemic logic of

[12] As usual, these results may be viewed as structural versions of completeness theorems.

belief revision. Then, we briefly look at the different formal languages used in our various systems, and contrast and compare them. Next, in line with both Chapters 2 and 3, I extend the setting from pure preference to intertwining of preference, knowledge, and beliefs. Several new concepts of preference will be defined, in a sequence of modal languages of ascending strength. Finally, I compare with a new proposal of combining all systems studied so far into one grand 'doxastic preferential predicate logic' of both object and world preference.

In Chapter 5, I move to a setting where habits of preference and belief change are just one aspect of general diversity of agents. Agents are *not* all the same, and nevertheless, they manage to coordinate with each other successfully. I start with the observation that dynamic epistemic logic presupposes that every agent remembers all the actions she has taken before. But then I show that this is a negotiable assumption, which can be dropped from the framework. In particular, *memory-bounded agents* are defined and their behavior is captured in a new dynamic epistemic completeness theorem with a key reduction axiom different from the usual one. Next, following ideas from my master of logic thesis [Liu04], I consider different policies in belief revision, and suggest how a continuum of these, too, can be incorporated into dynamic epistemic logic. These logics allow for co-existence of different memory capacities and revision policies, and hence, through different modal operators, they can describe the interplay of diverse agents. Throughout the chapter, imperfect information games, viewed as finite trees of possible actions with epistemic uncertainties, are used as a playground.

Finally, in Chapter 6, diversity of agents is discussed in a more abstract and systematic way. The major sources of diversity are considered first, such as inferential powers, introspective ability, powers of observation, memory capacity, and revision policies. I then show how these can be encoded in dynamic epistemic logics allowing for individual variation among agents along many dimensions. Furthermore, I explore the interaction of diverse agents by looking at some concrete scenarios of communication and learning. A logical methodology to deal with these issues is proposed as well.

Chapter 7 concludes the dissertation and identifies some major further issues for research that come to light once we put our chapters together into one account of diverse preference-driven agents.

**Origins of the material** Material from these chapters has been presented at several colloquia and conferences, including ESSLLI 2005 (Edinburgh), ESSLLI 2006 (Malaga), LOFT 2006 (Liverpool), and Luxembourg Workshop on Norm Change 2007. As for publications, Chapter 2 is the published joint paper ([BL07]). Chapter 3 is an extension of the joint paper ([JL06]) as submitted for publication organized after the *Workshop on Modeling Preference Change* in Berlin, 2006. Chapter 4 is largely new, and partly a product of the 'dynamics seminar' at ILLC Amsterdam. Chapter 5 is an updated and extended version of the published joint

paper ([BL04]). Chapter 6 is an extension of the accepted paper ([Liu06a]) of the *Workshop on Logics for Resource Bounded Agents* in Malaga, 2006, and it will appear in the *Journal of Logic, Language and Information*.