Changing for the better : preference dynamics and agent diversity

Liu, F.

Publication date
2008

Chapter 3

Preference, Priorities and Belief

3.1 Motivation

The notion of preference occurs frequently in game theory, decision theory, and many other research areas. Typically, preference is used to draw comparison between two alternatives explicitly. Studying preference and its general properties has become a main logical concern after the pioneering seminar work by [Hal57] and [Wri63], witness [Jen67], [Cre71], [Tra85], [DW94], [Han01a], [BRG07] etc., and more recently work on dynamics of preference e.g. [Han95] and [BL07]. Let us single out immediately the two distinctive characteristics of the approach to preference we take in this chapter.

- Most of the previous work has taken preference to be a primitive notion, without considering how it comes into being. We take a different angle here and explore both preference and its origin. We think that preference can often be rationally derived from a more basic source, which we will call a *priority base*. In this manner we have two levels: the priority base, and the preference derived from it. We hope this new perspective will shed light on the reasoning underlying preference, so that we are able to discuss *why* we prefer one thing over another. There are many ways to get preference from such a priority base, a good overview can be found in [CMLLM04].

- In real life we often encounter situations in which no complete information is available. Preference will then have to be based on our beliefs, i.e. do we believe certain properties from the priority base to apply or not? Apparently, this calls for a combination of doxastic language and preference language. We will show a close relationship between preference and beliefs. To us, both are mental attitudes. If we prefer something, we believe we do (and conversely). In addition, this chapter is also concerned with the dynamics of preference. By means of our approach, we can study preference changes, whether they are due to a change in the priority base, or caused by belief revision.

Depending on the actual situation, preference can be employed to compare alternative states of affairs, objects, actions, means, and so on, as listed in [Wri63]. One requirement we impose is that we consider only mutually exclusive alternatives. In this paper, we consider in first instance preference over *objects* rather than between propositions (compare [DW94]). Objects are, of course, congenitally mutually exclusive. Although the priority base approach is particularly well suited to compare preference between objects, it can be applied to the study of the comparison of other types of alternatives as well. In Section 3.7 we show how to apply the priority base approach to propositions. When comparing objects, the kind of situation to be thought of is:

3.1.1. Example. Alice is going to buy a house. For her there are several things to consider: the cost, the quality and the neighborhood, strictly in that order. All these are clear-cut for her, for instance, the cost is good if it is inside her budget, otherwise it is bad. Her decision is then determined by the information whether the alternatives have the desirable properties, and by the given order of importance of the properties.

In other words, Alice's preference regarding houses is derived from the priority order of the properties she considers. This chapter aims to propose a logic to model such situations. When covering situations in which Alice's preference is based on incomplete information belief will enter into the logic as an operation.

There are several points to be stressed beforehand, in order to avoid misunderstandings: First, our intuition of priority base is linked to graded semantics, e.g. spheres semantics by [Lew73]. We take a rather syntactical approach in this chapter, but that is largely a question of taste, one can go about it semantically as well. We will return to this point several times. Second, we will mostly consider a linearly ordered priority base. This is simple, giving us a quasi-linear order of preference. But our approach can be adapted to the partially ordered case, as we will indicate at the end of the paper. Third, when we add a belief operator to the preference language (fragment of FOL), it may seem that we are heading into doxastic predicate logic. This is true, but we are not going to be affected by the existing difficult issues in that logic. What we are using in this context is a very limited part of the language. Finally, although we start with a two level perspective this results on the preference side in logics that are rather like ordinary propositional modal logics. The bridge between the two levels is then given by theorems that show that any models of these modal logics can be seen as having been constructed from a priority base. These theorems are a kind of completeness theorems, but we call them *representation theorems* to distinguish them from the purely modal completeness results.

The following sections are structured as follows: In Section 3.2, we start with a simple language to study the rigid case in which the priorities lead to a clear and unambiguous preference ordering. In Section 3.3 we review some basics about ordering. Furthermore, a proof of a representation theorem for the simple language without beliefs is presented. Section 3.4 will consider what happens when the agent has incomplete information about the priorities with regard to the alternatives. In Section 3.5 we will look at changes in preference caused by two different sources: changes in beliefs, and changes of the sequence of priorities. Section 3.6 is an extension to the multi-agent system. We will prove representation theorems for the general case, and for the special cases of cooperative agents and competitive agents. In Section 3.7 we apply our approach to preference over propositions. Finally, we discuss how to generalize our approach to partially ordered preferences, and we end the chapter with a few conclusions.

3.2 From priorities to preference

As we mentioned in the preceding, there are many ways to derive preference from the priority base. We choose one of the mechanisms, the way of Optimality Theory (OT), as an illustration because we like the intuition behind this mechanism. Along the way, we will discuss other approaches as well, to indicate how our method can be applied to them as well.

Here is a brief review of some ideas from optimality theory that are relevant to the current context. In optimality theory a set of conditions is applied to the alternatives generated by the grammatical or phonological theory, to produce an optimal solution. It is by no means sure that the optimal solution satisfies all the conditions. There may be no such alternative. The conditions, called *constraints*, are strictly ordered according to their importance, and the alternative that satisfies the earlier conditions best (in a way described more precisely below) is considered to be the optimal one. This way of choosing the optimal alternative naturally induces a preference ordering among all the alternatives. We are interested in formally studying the way the constraints induce the *preference ordering* among the alternatives. The attitude in our investigations is somewhat differently directed than in optimality theory.[1]

Back to the issues of preference, to discuss preference over objects, we use a first order logic with constants $d_0, d_1 \ldots$; variables $x_0, x_1, \ldots$; and predicates $P, Q, P_0, P_1, \ldots$. In practice, we are thinking of finite domains, monadic predi-

[1]Note that in optimality theory the optimal alternative is chosen unconsciously; we are thinking mostly of applications where conscious choices are made. Also, in optimality theory the application of the constraints to the alternatives lead to a *clear* and *unambiguous* result: either the constraint clearly is true of the alternative or it is not, and that is something that is not sensitive to change. We will loosen this condition and consider issues that arise when changes do occur.

cates, simple formulas, usually quantifier free or even variable free. The following definition is directly inspired by optimality theory, but to take a neutral stance we use the words priority sequence instead of constraint sequence.

3.2.1. DEFINITION. A *priority sequence* is a finite ordered sequence of formulas (priorities) written as follows:

$$C_1 \gg C_2 \cdots \gg C_n \quad (n \in \mathbb{N}),$$

where each of C_m $(1 \leq m \leq n)$ is a formula from the language, and there is exactly one free variable x, which is a common one to each C_m.

We will use symbols like $\mathfrak{C}$ to denote priority sequences. The priority sequence is linearly ordered. It is to be read in such a way that the earlier priorities count strictly heavier than the later ones, for example, $C_1 \wedge \neg C_2 \wedge \cdots \wedge \neg C_m$ is preferable over $\neg C_1 \wedge C_2 \wedge \cdots \wedge C_m$ and $C_1 \wedge C_2 \wedge C_3 \wedge \neg C_4 \wedge \neg C_5$ is preferable over $C_1 \wedge C_2 \wedge \neg C_3 \wedge C_4 \wedge C_5$. A difference with optimality theory is that we look at *satisfaction* of the priorities whereas in optimality theory *infractions* of the constraints are stressed. This is more a psychological than a formal difference. However, optimality theory knows multiple infractions of the constraints and then counts the number of these infractions. We do not obtain this with our simple objects, but we think that possibility can be achieved by considering composite objects, like strings.

3.2.2. DEFINITION. Given a priority sequence of length n, two objects x and y, *Pref(x,y)* is defined as follows:

$$Pref_1(x,y) ::= C_1(x) \wedge \neg C_1(y),$$
$$Pref_{k+1}(x,y) ::= Pref_k(x,y) \vee (Eq_k(x,y) \wedge C_{k+1}(x) \wedge \neg C_{k+1}(y)), k < n,$$
$$Pref(x,y) ::= Pref_n(x,y),$$

where the auxiliary binary predicate $Eq_k(x,y)$ stands for $(C_1(x) \leftrightarrow C_1(y)) \wedge \cdots \wedge (C_k(x) \leftrightarrow C_k(y))$.[2]

In Example 3.1.1, Alice has the following priority sequence:

$$C(x) \gg Q(x) \gg N(x),$$

where $C(x)$, $Q(x)$ and $N(x)$ are intended to mean 'x has low cost', 'x is of good quality' and 'x has a nice neighborhood', respectively. Consider two houses d_1 and d_2 with the following properties: $C(d_1), C(d_2), \neg Q(d_1), \neg Q(d_2), N(d_1)$ and $\neg N(d_2)$. According to the definition, Alice prefers d_1 over d_2, i.e. $Pref(d_1, d_2)$.

Unlike in Section 3.4 belief does not enter into this definition. This means that $Pref(x,y)$ can be read as x *is superior to* y, or *under complete information* x *is preferable over* y.

[2]This way of deriving an ordering from a priority sequence is called *leximin ordering* in [CMLLM04].

3.2.3. Remark. Our method easily applies when the priorities become graded. Take the Example 3.1.1, if Alice is more particular, she may split the cost C into C^1 very low cost, C^2 low cost, C^3 medium cost, similarly for the other priorities. The original priority sequence $C(x) \gg Q(x) \gg N(x)$ may change into

$$C^1(x) \gg C^2(x) \gg Q^1(x) \gg C^3(x) \gg Q^2(x) \gg N^1(x) \gg \dots \ .$$

As we mentioned at the beginning, we have chosen a syntactic approach expressing priorities by formulas. If we switch to a semantical point of view, the priority sequence translates into pointing out a sequence of n sets in the model. The elements of the model will be objects rather than worlds as is usual in this kind of study. But one should see this really as an insignificant difference. If one prefers, one may for instance in Example 3.1.1 replace house d by the situation in which Alice has bought the house d.

When one points out sets in a model, Lewis' sphere semantics ([Lew73] p.98-99) comes to mind immediately. The n sets in the model obtained from the priority base are in principle unrelated. In the sphere semantics the sets which are pointed out are linearly ordered by inclusion. To compare with the priority base we switch to a syntactical variant of sphere semantics, a sequence of formulas $G_1, \dots, G_m$ such that $G_i(x)$ implies $G_j(x)$ if $i \leq j$. These formulas express the preferability in a more direct way, $G_1(x)$ is the most preferable, $G_m(x)$ the least. In what follows, we will show that the two approaches are equivalent in the sense that they can be translated into each other.

3.2.4. Theorem. *A priority sequence $C_1 \gg C_2 \cdots \gg C_m$ gives rise to a G-sequence of length 2^m. In the other direction a priority sequence can be obtained from a G-sequence logarithmic in the length of the G-sequence.*

Proof. Let us just look at the case that m=3. Assuming that we have the priority sequence $C_1 \gg C_2 \gg C_3$, the preference of objects is decided by where their properties occur in the following list:

$R_1 : C_1 \wedge C_2 \wedge C_3;$
$R_2 : C_1 \wedge C_2 \wedge \neg C_3;$
$R_3 : C_1 \wedge \neg C_2 \wedge C_3;$
$R_4 : C_1 \wedge \neg C_2 \wedge \neg C_3;$
$R_5 : \neg C_1 \wedge C_2 \wedge C_3;$
$R_6 : \neg C_1 \wedge C_2 \wedge \neg C_3;$
$R_7 : \neg C_1 \wedge \neg C_2 \wedge C_3;$
$R_8 : \neg C_1 \wedge \neg C_2 \wedge \neg C_3.$

The $G_i s$ are constructed as disjunctions of members of this list. In their most simple form, they can be stated as follows:

$G_1 : R_1$;
$G_2 : R_1 \vee R_2$;
$\vdots$
$G_8 : R_1 \vee\ R_2 \cdots \vee R_8$.

On the other hand, given a G_i-sequence, we can define C_i as follows,

$C_1 = R_1 \vee R_2 \vee R_3 \vee R_4$;
$C_2 = R_1 \vee R_2 \vee R_5 \vee R_6$;
$C_3 = R_1 \vee R_3 \vee R_5 \vee R_7$.

And again this can be simply read off from a picture of the G-spheres. The relationship between C_i, R_i, and G_i can be seen from the Figure 3.1. □

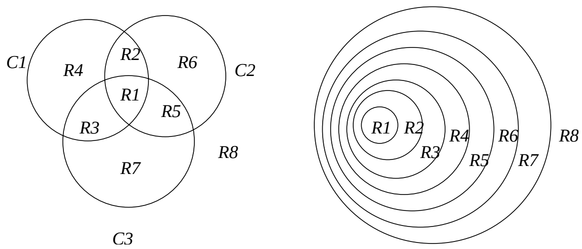

Figure 3.1: C_i, R_i, and G_i

3.2.5. REMARK. In applying our method to such spheres, the definition of $\underline{Pref}(x, y)$ comes out to be $\forall i(y \in G_i \to x \in G_i)$. The whole discussion implies of course that our method can be applied to spheres as well as to any other approach which can be reduced to spheres.

3.2.6. REMARK. As we pointed out at the beginning, one can define preference from a priority sequence $\mathfrak{C}$ in various different ways, all of which we can handle. Here is one of these ways, called *best-out ordering* in [CMLLM04], as an illustration. We define the preference as follows:

$$Pref(x, y) \quad \text{iff} \quad \exists C_j \in \mathfrak{C}(\forall C_i \gg C_j\, ((C_i(x) \wedge C_i(y)) \wedge (C_j(x) \wedge \neg C_j(y))).$$

In this case, we only continue along the priority sequence as long as we receive positive information. Returning the Example 3.1.1, this means that under this option we only get the conclusion that $\underline{Pref}(d_1, d_2)$ and $\underline{Pref}(d_2, d_1)$: d_1 and d_2 are equally preferable, because after observing that $\neg Q(d_1), \neg Q(d_2)$, Alice won't consider N at all.

3.3 Order and a representation theorem

In this section we will just run through the types of order that we will use in the current context. A relation $<$ is a *linear order* if $<$ is irreflexive, transitive and asymmetric, and satisfies *connectedness*:

$$x < y \vee x = y \vee y < x$$

More precisely, $<$ is called a *strict* linear order. A *non-strict* linear order $\leq$ is a reflexive, transitive, antisymmetric and connected relation. It is for various reasons useful to introduce non-strict variants of orderings as well.

Mathematically, strict and non-strict linear orders can easily be translated into each other:

(1) $x < y \leftrightarrow x \leq y \wedge x \neq y$, or
(2) $x < y \leftrightarrow x \leq y \wedge \neg(y \leq x)$,
(3) $x \leq y \leftrightarrow x < y \vee x = y$, or
(4) $x \leq y \leftrightarrow x < y \vee (\neg(x < y) \wedge \neg(y < x))$.

Optimality theory only considers linearly ordered constraints. These will be seen to lead to a *quasi-linear order* of preferences, i.e. a relation $\preccurlyeq$ that satisfies all the requirements of a non-strict linear order but antisymmetry. A quasi-linear ordering contains *clusters* of elements that are 'equally large'. Such elements are $\leq$ each other. Most naturally one would take for the strict variant $\prec$ an irreflexive, transitive, connected relation. If one does that, strict and non-strict orderings can still be translated into each other (only by using alternatives (2) and (4) in the above though, not (1) and (3)). However, $Pref$ is normally taken to be an asymmetric relation, and we agree with that, so we take the option of $\prec$ as an irreflexive, transitive, asymmetric relation. Then $\prec$ is definable in terms of $\preccurlyeq$ by use of (2), but not $\preccurlyeq$ in terms of $\prec$. That is clear from the picture below, an irreflexive, transitive, asymmetric relation cannot distinguish between the two given orderings.

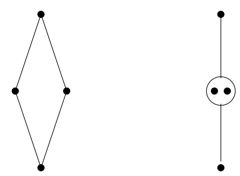

Figure 3.2: Incomparability and indifference.

One needs an additional equivalence relation $x \sim y$ to express that x and y are elements in the same cluster; $x \sim y$ can be defined by

(5) $x \sim y \leftrightarrow x \leq y \wedge y \leq x$.

Then, in the other direction, $x \leq y$ can be defined in terms of $<$ and $\sim$:

(6) $x \leq y \leftrightarrow x < y \vee x \sim y$.

It is certainly possible to extend our discussion to partially ordered sets of constraints, and we will make this excursion in Section 3.8. The preference relation will no longer be a quasi-linear order, but a so-called *quasi-order*: in the non-strict case a reflexive and transitive relation, in the strict case an asymmetric, transitive relation. One can still use (2) to obtain a strict quasi-order from a non-strict one and (6) to obtain a non-strict quasi-order from a strict one and $\sim$. However, we will see in Section 3.4 that in some contexts involving beliefs these translations no longer give the intended result. In such a case one has to be satisfied with the fact that (5) still holds and that $\prec$ as well as $\sim$ imply $\preccurlyeq$.

In the following we will write $Pref$ for the strict version of preference, $\underline{Pref}$ for the non-strict version, and let Eq correspond to $\sim$, expressing two elements are equivalent. Clearly, no matter what the priorities are, the non-strict preference relation has the following general properties:

(a) $\underline{Pref}(x, x)$,
(b) $\underline{Pref}(x, y) \vee \underline{Pref}(y, x)$,
(c) $\underline{Pref}(x, y) \wedge \underline{Pref}(y, z) \rightarrow \underline{Pref}(x, z)$.

(a), (b) and (c) express reflexivity, connectedness and transitivity, respectively. Thus, $\underline{Pref}$ is a quasi-linear relation; it lacks antisymmetry.

Unsurprisingly, (a), (b) and (c) are a complete set of principles for preference. We will put this in the form of a representation theorem as we announced in the introduction. In this case it is a rather trivial matter, but it is worthwhile to execute it completely as an introduction to the later variants. We reduce the first order language for preference to its core:

3.3.1. DEFINITION. Let Γ be a set of propositional variables, and D be a finite domain of objects, the *reduced language* of preference logic is defined as follows,

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \underline{Pref}(d_i, d_j),$$

where p, d_i respectively denote elements from Γ and D.

The reduced language contains the propositional calculus. From this point onwards we refer to the language with variables, quantifiers, predicates as the *extended language*. In the reduced language, we rewrite the axioms as follows:

(a) $\underline{Pref}(d_i, d_i)$,
(b) $\underline{Pref}(d_i, d_j) \vee \underline{Pref}(d_j, d_i)$,
(c) $\underline{Pref}(d_i, d_j) \wedge \underline{Pref}(d_j, d_k) \rightarrow \underline{Pref}(d_i, d_k)$.

We call this axiom system **P**.

3.3.2. THEOREM. (*representation theorem*). $\vdash_{\mathbf{P}} \varphi$ *iff* φ *is valid in all models obtained from priority sequences.*

Proof. The direction from left to right is obvious. Assume formula $\varphi(d_1, \dots, d_n, p_1, \dots, p_k)$ is not derivable in **P**. Then a non-strict quasi-linear ordering of the $d_1, \dots, d_n$ exists, which, together with a valuation of the atoms $p_1, \dots, p_k$ in φ falsifies $\varphi(d_1, \dots, d_n)$. Let us assume that we have a linear order (adaptation to the more general case of quasi-linear order is simple), and also, w.l.o.g. that the ordering is $d_1 > d_2 > \cdots > d_n$. Then we introduce an extended language containing unary predicates $P_1, \dots, P_n$ with a priority sequence $P_1 \gg P_2 \cdots \gg P_n$ and let P_i apply to d_i only. Clearly, the preference order of $d_1, \dots, d_n$ with respect to the given priority sequence is from left to right. We have transformed the model into one in which the defined preference has the required properties.[3] □

3.3.3. REMARK. It is instructive to execute the above proof for the reduced language containing some additional predicates $Q_1, \dots, Q_k$. One would like then to obtain a priority sequence of formulas in the language built up from Q_1 to Q_k. This is possible if in the model $\mathcal{M}$ each pair of constants d_i and d_j is distinguishable by formulas in this language, i.e. for each i and j, there exists a formula φ_{ij} such that $\mathcal{M} \models \varphi_{ij}(d_i)$and $\mathcal{M} \models \neg\varphi_{ij}(d_j)$. In such a case, the formula $\psi_i = \bigwedge_{i \neq j} \varphi_{ij}$ satisfies only d_i. And $\psi_1 \gg \cdots \gg \psi_n$ is the priority sequence as required. It is necessary to introduce new predicates when two constants are indistinguishable. A trivial method to do this is to allow identity in the language, $x = d_1$ obviously distinguishes d_1 and d_2.

Let us at this point stress once more what the content of a representation theorem is. It tells us that the way we have obtained the preference relations, namely from a priority sequence, does not affect the general reasoning about preference, its logic. The above proof shows this in a rather strong way: if we have a model in which the preference relation behaves in a certain manner, then we can think of this preference as derived from a priority sequence without disturbing the model as it is.

3.4 Preference and belief

In this section, we discuss the situation that arises when an agent has only incomplete information, but she likes to express her preference. The language will be extended with belief operators $B\varphi$ to deal with such uncertainty, and it is a

[3] Note that, although we used n priorities in the proof to make the procedure easy to describe, in general $^2log(n) + 1$ priorities are sufficient for the purpose.

small fragment of doxastic predicate logic. It would be interesting to consider what more the full doxastic predicate logic language can bring us, but we will leave this question to other occasions. We will take the standard **KD45** as the logic for beliefs, though we are aware of the philosophical discussions on beliefs and the options of proper logical systems.

Interestingly, the different definitions of preference we propose in the following spell out different "procedures" an agent may follow to decide her preference when processing the incomplete information about the relevant properties. Which procedure is taken strongly depends on the domain or the type of agents. In the new language, the definition of priority sequence remains the same, i.e. a priority C_i is a formula from the language *without* belief operators.

3.4.1. Definition. (decisive preference). Given a priority sequence of length n, two objects x and y, *Pref(x,y)* is defined as follows:

$$Pref_1(x,y) ::= BC_1(x) \wedge \neg BC_1(y),$$
$$Pref_{k+1}(x,y) ::= Pref_k(x,y) \vee (Eq_k(x,y) \wedge BC_{k+1}(x) \wedge \neg BC_{k+1}(y)), k < n,$$
$$Pref(x,y) ::= Pref_n(x,y),$$

where $Eq_k(x,y)$ stands for $(BC_1(x) \leftrightarrow BC_1(y)) \wedge \cdots \wedge (BC_k(x) \leftrightarrow BC_k(y))$.

To determine the preference relation, one just runs through the sequence of relevant properties to check whether one believes them of the objects. But at least two other options of defining preference seem reasonable as well.

3.4.2. Definition. (conservative preference). Given a priority sequence of length n, two objects x and y, *Pref(x,y)* is defined below:

$$Pref_1(x,y) ::= BC_1(x) \wedge B\neg C_1(y),$$
$$Pref_{k+1}(x,y) ::= Pref_k(x,y) \vee (Eq_k(x,y) \wedge BC_{k+1}(x) \wedge B\neg C_{k+1}(y)), k < n,$$
$$Pref(x,y) ::= Pref_n(x,y)$$

where $Eq_k(x,y)$ stands for $(BC_1(x) \leftrightarrow BC_1(y)) \wedge (B\neg C_1(x) \leftrightarrow B\neg C_1(y)) \wedge \cdots \wedge (BC_k(x) \leftrightarrow BC_k(y)) \wedge (B\neg C_k(x) \leftrightarrow B\neg C_k(y))$.

3.4.3. Definition. (deliberate preference). Given a priority sequence of length n, two objects x and y, *Pref(x,y)* is defined below:

$$Supe_1(x,y)^4 ::= C_1(x) \wedge \neg C_1(y),$$
$$Supe_{k+1}(x,y) ::= Supe_k(x,y) \vee (Eq_k(x,y) \wedge C_{k+1}(x) \wedge \neg C_{k+1}(y)), k < n,$$
$$Supe(x,y) ::= Supe_n(x,y),$$

[4] Superiority is just defined as preference was in the previous section.

$Pref(x, y) ::= B(Supe(x, y))$,

where $Eq_k(x, y)$ stands for $(C_1(x) \leftrightarrow C_1(y)) \wedge \cdots \wedge (C_k(x) \leftrightarrow C_k(y))$.

To better understand the difference between the above three definitions, we look at the Example 3.1.1 again, but in three different variations:

A. Alice favors Definition 3.4.1: She looks at what information she can get, she reads that d_1 has low cost, about d_2 there is no information. This immediately makes her decide for d_1. This will remains so, no matter what she hears about quality or neighborhood.

B. Bob favors Definition 3.4.2: The same thing happens to him. But he reacts differently than Alice. He has no preference, and that will remain so as long as he hears nothing about the cost of d_2, no matter what he hears about quality or neighborhood.

C. Cora favors Definition 3.4.3: She also has the same information. On that basis Cora cannot decide either. But some more information about quality and neighborhood helps her to decide. For instance, suppose she hears that d_1 has good quality or is in a good neighborhood, and d_2 is not of good quality and not in a good neighborhood. Then Cora believes that, no matter what, d_1 is superior, so d_1 is her preference. Note that such kind of information could not help Bob to decide.

Speaking more generally in terms of the behaviors of the above agents, it seems that Alice always decides what she prefers on the basis of the limited information she has. In contrast, Bob chooses to wait and require more information. Cora behaves somewhat differently, she first tries to do some reasoning with all the available information before making her decision. This suggests yet another perspective on diversity of agents than discussed in Chapter 6.

Apparently, we have the following fact.

3.4.4. FACT.

- Totality holds for Definition 3.4.1, but not for Definition 3.4.2 or 3.4.3;

- Among the above three definitions, Definition 3.4.2 is the strongest in the sense that if $Pref(x, y)$ holds according to Definition 3.4.2, then $Pref(x, y)$ holds according to Definition 3.4.1 and 3.4.3 as well.

It is striking that, if in Definition 3.4.3, one plausibly also defines $\underline{Pref}(x, y)$ as $B(\underline{Supe}(x, y))$, then the normal relation between $Pref$ and $\underline{Pref}$ no longer holds: $Pref$ is not definable in terms of $\underline{Pref}$ any more, or even $\underline{Pref}$ in terms of $Pref$ and Eq.

For all three definitions, we have the following theorem.

3.4.5. THEOREM. $\underline{Pref}(x,y) \leftrightarrow B\underline{Pref}(x,y)$.

Proof. In fact we prove something more general in **KD45**. Namely, if α is a propositional combination of B-statements, then $\vdash_{\mathbf{KD45}} \alpha \leftrightarrow B\alpha$.

From left to right, since α is a propositional combination of B-statements, it can be transformed into conjunctive normal form: $\beta_1 \vee \cdots \vee \beta_k$. It is clear that $\vdash_{\mathbf{KD45}} \beta_i \rightarrow B\beta_i$ for each i, because each member γ of the conjunction β_i implies $B\gamma$. If $A = \beta_1 \vee \cdots \vee \beta_k$ holds then some β_i holds, so $B\beta_i$, so $B\alpha$. Then we immediately have: $\vdash_{\mathbf{KD45}} \neg\alpha \rightarrow B\neg\alpha$ (*) as well, since $\neg\alpha$ is also a propositional combination of B-statements if α is.

From right to left: Suppose $B\alpha$ and $\neg\alpha$. Then $B\neg\alpha$ by (*), so $B\bot$, but this is impossible in **KD45**, therefore α holds.

The theorem follows since $\underline{Pref}(x,y)$ is in all three cases indeed a propositional combination of B-statements. □

3.4.6. COROLLARY. $\neg\underline{Pref}(x,y) \leftrightarrow B\neg\underline{Pref}(x,y)$.

Actually, we think it is proper that Theorem 3.4.5 and Corollary 3.4.6 hold because we believe that preference describes a state of mind in the same way that belief does. Just as one believes what one believes, one believes what one prefers.

If we stick to Definition 3.4.1, we can generalize the representation result (Theorem 3.3.2). Let us consider the reduced language built up from standard propositional letters, plus $\underline{Pref}(d_i,d_j)$ by the connectives, and belief operators B. Again we have the normal principles of **KD45** for B.

3.4.7. THEOREM. *The following principles axiomatize exactly the valid ones.*

(a) $\underline{Pref}(d_i,d_i)$,
(b) $\underline{Pref}(d_i,d_j) \vee \underline{Pref}(d_j,d_i)$,
(c) $\underline{Pref}(d_i,d_j) \wedge \underline{Pref}(d_j,d_k) \rightarrow \underline{Pref}(d_i,d_k)$,
(1.) $\neg B\bot$,
(2.) $B\varphi \rightarrow BB\varphi$,
(3.) $\neg B\varphi \rightarrow B\neg B\varphi$,
(4.) $\underline{Pref}(d_i,d_j) \leftrightarrow B\underline{Pref}(d_i,d_j)$.

We now consider the **KD45-P** system including the above valid principles, *Modus ponens*(MP), as well as *Generalization* for the operator B.

3.4.8. DEFINITION. A model of **KD45-P** is a tuple $\langle W, D, R, \{\preceq_w\}_{w\in W}, V\rangle$, where W is a set of worlds, D is a set of constants, R is a euclidean and serial accessibility relation on W. Namely, it satisfies $\forall xyz((Rxy \wedge Rxz) \rightarrow Ryz)$ and $\forall x\exists y Rxy$. For each w, $\preceq_w$ is a quasi-linear order on D, which is the same throughout each euclidean class. V is evaluation function in an ordinary manner.

We remind the reader that in most respects euclidean classes are equivalence classes except that a number of points are irreflexive and have R relations just towards the reflexive members (the *equivalence part*) of the class.

3.4.9. Theorem. *The* **KD45-P** *system is complete.*

Proof. The canonical model of this logic **KD45-P** has the required properties: The belief accessibility relation R is euclidean and serial. This means that with regard to R the model falls apart into euclidean classes. In each node $\underline{Pref}$ is a quasi-linear order of the constants. Within a euclidean class the preference order is constant (by $B\underline{Pref} \leftrightarrow \underline{Pref}$). This suffices to prove completeness. □

3.4.10. Theorem. *The logic* **KD45-P** *has the finite model property.*

Proof. By standard methods. □

3.4.11. Theorem. (*representation theorem*). $\vdash_{\mathbf{KD45-P}} \varphi$ *iff* φ *is valid in all models obtained from priority sequences.*

Proof. Suppose that $\nvdash_{\mathbf{KD45-P}} \varphi(d_1, ..., d_n, p_1, ..., p_m)$. By Theorem 3.4.9, there is a model with a world w in which φ is falsified. We restrict the model to the euclidean class where w resides. Since the ordering of the constants is the same throughout euclidean classes, the ordering of the constants is now the same throughout the whole model. We can proceed as in Theorem 3.3.2 defining the predicates $P_1, \ldots, P_n$ in a constant manner throughout the model. □

3.4.12. Remark. The three definitions above are not the only definitions that might be considered. For instance, we can give a variation (*) of Definition 3.4.2. For simplicity, we just use one predicate C.

$$Pref(x, y) ::= \neg B\neg C(x) \wedge B\neg C(y). \quad (*)$$

This means the agent can decide on her preference in a situation in which on the one hand she is not totally ready to believe $C(x)$, but considers it consistent with what she assumes, on the other hand, she distinctly believes $\neg C(y)$. Compared with Definition 3.4.2, (*) is weaker in the sense that it does not require explicit positive beliefs concerning $C(x)$.

We can even combine Definition 3.4.1 and (*), obtaining the following:

$$Pref(x, y) ::= (BC(x) \wedge \neg BC(x)) \vee (\neg B\neg C(x) \wedge B\neg C(y). \quad (**)$$

Contrary to (*), this gives a quasi-linear order.

Similarly, for Definition 3.4.3, if instead of $B(Supe(x, y))$, we use $\neg B\neg(Supe(x, y))$, a weaker preference definition is obtained.

3.5 Preference changes

So far we have given different definitions for preference in a stable situation. Now we direct ourselves to changes in this situation. In the definition of preference in the presence of complete information, the only item subject to change is the priority sequence. In the case of incomplete information, not only the priority sequence, but also our beliefs can change. Both changes in priority sequence and changes in belief can cause preference change. In this section we study both. Note that priority change leads to a preference change in a way similar to entrenchment change in belief revision theory (see [Rot03]), but we take the methodology of dynamic epistemic logic in this context.

3.5.1 Preference change due to priority change

Let us first look at a variation of Example 3.1.1:

3.5.1. Example. Alice won a lottery prize of ten million dollars. Her situation has changed dramatically. Now she considers the quality most important.

In other words, the ordering of the priorities has changed. We will focus on the priority changes, and the preference changes they cause. To this purpose, we start by making the priority sequence explicit in the preference. We do this first for the case of complete information in language without belief. Let $\mathfrak{C}$ be a priority sequence with length n as in Definition 3.2.1. Then we write $Pref_{\mathfrak{C}}(x,y)$ for the preference defined from that priority sequence. Let us consider the following possible changes: we write $\mathfrak{C}^\frown C$ for adding C to the right of $\mathfrak{C}$, $C^\frown\mathfrak{C}$ for adding C to the left of $\mathfrak{C}$, $\mathfrak{C}^-$ for the sequence $\mathfrak{C}$ with its final element deleted, and finally, $\mathfrak{C}^{i\leftrightarrows i+1}$ for the sequence $\mathfrak{C}$ with its i-th and i+1-th priorities switched. It is then clear that we have the following relationships:

$$Pref_{\mathfrak{C}^\frown C}(x,y) \leftrightarrow Pref_{\mathfrak{C}}(x,y) \vee (Eq_{\mathfrak{C}}(x,y) \wedge C(x) \wedge \neg C(y)),$$
$$Pref_{C^\frown\mathfrak{C}}(x,y) \leftrightarrow (C(x) \wedge \neg C(y)) \vee ((C(x) \leftrightarrow C(y)) \wedge Pref_{\mathfrak{C}}(x,y)),$$
$$Pref_{\mathfrak{C}^-}(x,y) \leftrightarrow Pref_{\mathfrak{C},n-1}(x,y),$$
$$Pref_{\mathfrak{C}^{i\leftrightarrows i+1}}(x,y) \leftrightarrow Pref_{\mathfrak{C},i-1}(x,y) \vee (Eq_{\mathfrak{C},i-1}(x,y) \wedge C_{i+1}(x) \wedge \neg C_{i+1}(y)) \vee (Eq_{\mathfrak{C},i-1}(x,y) \wedge (C_{i+1}(x) \leftrightarrow C_{i+1}(y)) \wedge C_i(x) \wedge \neg C_i(y)) \vee (Eq_{\mathfrak{C},i+1}(x,y) \wedge Pref_{\mathfrak{C}}(x,y)).$$

These relationships enable us to describe preference change due to changes of the priority sequence in the manner of dynamic epistemic logic. We now consider the following four operations: $[^+C]$ of adding C to the right, $[C^+]$ of adding C to the left, $[-]$ of dropping the last element of a priority sequence of length n, and $[i\leftrightarrow i{+}1]$ of interchanging the i-th and i+1-th elements. Then we obtain the following reduction axioms:

$$[^+C]Pref(x,y) \leftrightarrow Pref(x,y) \vee (Eq(x,y) \wedge C(x) \wedge \neg C(y)),$$
$$[C^+]Pref(x,y) \leftrightarrow ((C(x) \wedge \neg C(y)) \vee ((C(x) \leftrightarrow C(y)) \wedge Pref(x,y))),$$
$$[-]Pref(x,y) \leftrightarrow Pref_{n-1}(x,y),$$
$$[i \leftrightarrow i+1]Pref(x,y) \leftrightarrow Pref_{i-1}(x,y) \vee (Eq_{i-1}(x,y) \wedge C_{i+1}(x) \wedge \neg C_{i+1}(y)) \vee (Pref_i(x,y) \wedge (C_{i+1}(x) \leftrightarrow C_{i+1}(y))) \vee (Eq_{i+1}(x,y) \wedge Pref(x,y)).$$

Of course, the first two are the more satisfactory ones, as the right hand side is constructed solely on the basis of the previous $Pref$ and the added priority C. Note that one of the first two, plus the third and the fourth are sufficient to represent any change whatsoever in the priority sequence. Noteworthy also is that operator $[C^+]$ has exactly the same effects on a model as the operator $[\sharp C]$ in Chapter 2. We will discuss connections of this sort later in Chapter 4.

In the context of incomplete information when we have the language of belief, we can obtain similar reduction axioms for Definition 3.4.1 and 3.4.2. For instance, for Definition 3.4.1, we need only replace C by BC and $\neg C$ by $\neg BC$. For Definition 3.4.3, the situation is very complicated, reduction axioms are simply not possible. To see this, we return to the Example of Cora. Suppose Cora has a preference on the basis of cost and quality, and she also has the given information relating quality and neighborhood. Then her new preference after 'neighborhood' has been adjoined to the priority sequence is not a function of her previous preference and her beliefs about the neighborhood. The beliefs relating quality and neighborhood are central for her reasoning, but they are neither contained in the beliefs supporting her previous preference, nor in the beliefs about the neighborhood per se.

3.5.2 Preference change due to belief change

Now we move to the other source which causes preference change, namely, a change in belief. Such a thing often occurs in real life, new information comes in, one changes one's beliefs. Technically, the update mechanisms of [BS06a] and [Ben07a] can immediately be applied to our system with belief. As preference is defined in terms of beliefs, we can calculate preference changes from belief change. We distinguish the two cases that the belief change is caused by an update with so-called *hard* information and an update with *soft* information.

Preference change under hard information

Consider a simpler version of the Example 3.1.1:

3.5.2. Example. Let us assume that this time Alice only consider the houses' cost (C) and their neighborhood (N) with $C(x) \gg N(x)$. There are two houses d_1 and d_2 available. The real situation is that $C(d_1), N(d_1), C(d_2)$ and $\neg N(d_2)$. First Alice prefers d_2 over d_1 because she believes $C(d_2)$ and $N(d_1)$. However, now

Alice reads that $C(d_1)$ in a newspaper. She accepts this information. Accordingly, she changes her preference.

Here we assume that Alice treats the information obtained as hard information. She simply adds new information to her stock of beliefs. Figure 3.3 shows the situation before Alice's reading.

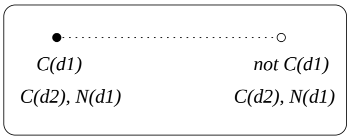

Figure 3.3: Initial model.

As usual, the dotted line denotes that Alice is uncertain about the two situations. In particular, she does not know whether $C(d_1)$ holds or not. After she reads that $C(d_1)$, the situation becomes Figure 3.4. The $\neg C(d_1)$-world is eliminated from the model: Alice has updated her beliefs. Now she prefers d_1 over d_2.

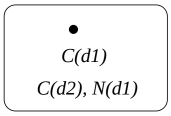

Figure 3.4: Updated model.

We have assumed that we are using the elimination semantics (e.g. [Ben06a], [FHMV95], etc.) in which public announcement of the sentence A leads to the elimination of the $\neg A$ worlds from the model. We have the reduction axiom:

$$[!A]Pref_{\mathfrak{C}}(x,y) \leftrightarrow A \rightarrow Pref_{A\rightarrow\mathfrak{C}}(x,y),$$

where, if $\mathfrak{C}$ is the priority sequence $C_1 \gg \cdots \gg C_n$, $A \rightarrow \mathfrak{C}$ is defined as $A \rightarrow C_1 \gg \cdots \gg A \rightarrow C_n$.

We can go even further if we use conditional beliefs $B^{\psi}\varphi$ as introduced in [Ben07a], with the meaning φ is believed under the condition of ψ. Naturally one can also introduce *conditional preference* $Pref^{\psi}(x,y)$, by replacing B in the definitions in Section 3.4 by B^{ψ}. Assuming A is a formula without belief operators, an easy calculation gives us another form of the reduction axiom:

$$[!A]Pref(x,y) \leftrightarrow A \rightarrow Pref^{A}(x,y).$$

Preference change under soft information

When incoming information is not as solid as considered in the above, we have to take into account the possibilities that the new information is not consistent with the beliefs the agent holds. Either the new information is unreliable, or the agent's beliefs are untenable. Let us switch to a semantical point of view for a moment. To discuss the impact of soft information on beliefs, the models are graded by a plausibility ordering $\leq$. For the one agent case one may just as well consider the model to consist of one euclidean class. The ordering of this euclidean class is such that the worlds in the equivalence part are the most plausible worlds. For all the worlds w in the equivalence part and all the worlds u outside it, $w < u$. Otherwise $v < v'$ can only obtain between worlds outside the equivalence part. To be able to refer to the elements in the model, instead of only to the worlds accessible by the R-relation, we introduce the universal modality U and its dual E. For the update by soft information, there are various approaches, we choose the *lexicographic upgrade* $\Uparrow A$ introduced by [Vel96] and [Rot06], adopted by [Ben07a] for this purpose. After the incoming information A, the ordering $\leq$ is updated by making all A-worlds strictly better than all $\neg A$-worlds keeping among the A-worlds the old orders intact and doing the same for the $\neg A$-worlds. After the update the R-relations just point to the best A-worlds. The reduction axiom for belief proposed in [Ben07a] is:

$$[\Uparrow A]B\varphi \leftrightarrow (EA \wedge\ B^A([\Uparrow A]\varphi) \vee (\neg EA \wedge\ B[\Uparrow A]\varphi)$$

We apply this only to priority formulas φ which do not have belief operators, and obtain for this restricted case a simpler form:

$$[\Uparrow A]B\varphi \leftrightarrow (EA \wedge\ B^A\varphi) \vee (\neg EA \wedge\ B\varphi).$$

From this one easily derive the reduction axiom for preference:

$$[\Uparrow A]Pref(x,y) \leftrightarrow (EA \wedge\ Pref^A(x,y)) \vee (\neg EA \wedge\ Pref(x,y)).$$

Or in a form closer to the one for hard information:

$$[\Uparrow A]Pref(x,y) \leftrightarrow (EA \rightarrow\ Pref^A(x,y)) \wedge (\neg EA \rightarrow\ Pref(x,y)).$$

The reduction axiom for conditional preference is:

$$[\Uparrow A]Pref^{\psi}(x,y) \leftrightarrow (E(A\wedge\psi) \rightarrow\ Pref^{A\wedge\psi}(x,y)) \wedge (\neg E(A\wedge\psi) \rightarrow\ Pref^{\psi}(x,y)).$$

By the fact that we have reduction axioms here, the completeness result in [Ben07a] for dynamic belief logic can be extended to a dynamic preference logic.

We will not spell out the details here.

3.6 Extension to the many agent case

This section extends the results of Section 3.4 to the many agent case. This will generally turn out to be more or less a routine matter. But at the end of the section, we will see that the priority base approach gives us a start of an analysis of cooperation and competition of agents. We consider agents here as cooperative if they have the same goals (priorities), competitive if they have opposite goals. This foreshadows the direction one may take to apply our approach to games. The language we are using is defined as follows.

3.6.1. Definition. Let Γ be a set of propositional variables, G be a group of agents, and D be a finite domain of objects, the *reduced language* of preference logic for many agents is defined in the following,

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \underline{Pref}^a(d_i, d_j) \mid B^a\varphi$$

where p, a, d_i respectively denote elements from Γ, G, and D.

Similarly to $\underline{Pref}^a$ expressing non-strict preference, we will use $Pref^a$ to denote the strict version. When we want to use the extended language, we add variables and the statements $P(d_i)$.

3.6.2. Definition. A *priority sequence* for an agent a is a finite ordered sequence of formulas written as follows: $C_1 \gg_a C_2 \cdots \gg_a C_n \quad (n \in \mathbb{N})$, where each C_m $(1 \leq m \leq n)$ is a formula from the language of Definition 3.6.1, with one single free variable x, but without $\underline{Pref}$ and B.

Here we take decisive preference to define an agent's preference. But the results of this section apply to other definitions just as well. It seems quite reasonable to allow in this definition of $Pref^a$ formulas that contain B^b and $Pref^b$ for agents b other than a. But we leave this for a future occasion.

3.6.3. Definition. Given a priority sequence of length n, two objects x and y, $Pref^a(x, y)$ is defined as follows:

$$Pref^a_1(x, y) ::= B^aC_1(x) \wedge \neg B^aC_1(y),$$
$$Pref^a_{k+1}(x, y) ::= Pref^a_k(x, y) \vee (Eq_k(x, y) \wedge B^aC_{k+1}(x) \wedge \neg B^aC_{k+1}(y)), k < n,$$
$$Pref^a(x, y) ::= Pref^a_n(x, y),$$

where $Eq_k(x, y)$ stands for $(B^aC_1(x) \leftrightarrow B^aC_1(y)) \wedge \cdots \wedge (B^aC_k(x) \leftrightarrow B^aC_k(y))$.

3.6.4. Definition. The preference logic for many agents **KD45-P^G** is consists of the following principles,

(a) $\underline{Pref^a}(d_i, d_i)$,
(b) $\underline{Pref^a}(d_i, d_j) \vee \underline{Pref^a}(d_j, d_i)$,
(c) $\underline{Pref^a}(d_i, d_j) \wedge \underline{Pref^a}(d_j, d_k) \rightarrow \underline{Pref^a}(d_i, d_k)$,
(1.) $\neg B^a \bot$,
(2.) $B^a \varphi \rightarrow B^a B^a \varphi$,
(3.) $\neg B^a \varphi \rightarrow B^a \neg B^a \varphi$,
(4.) $\underline{Pref^a}(d_i, d_j) \leftrightarrow B^a \underline{Pref^a}(d_i, d_j)$.

As usual, it also includes *Modus ponens*(MP), as well as *Generalization* for the operator B^a. It is easy to see that the above principles are valid for $\underline{Pref^a}$ extracted from a priority sequence.

3.6.5. Theorem. *The preference logic for many agents* **KD45-P^G** *is complete.*

Proof. The canonical model of this logic **KD45-P^G** has the required properties: The belief accessibility relation R_a is euclidean and serial. This means that with regard to R_a the model falls apart into a-euclidean classes. Again, in each node $Pref^a$ is a quasi-linear order of the constants and within an a-euclidean class the a-preference order is constant. This quasi-linearity and constancy are of course the required properties for the preference relation. Same for the other agents. This shows completeness of the logic. □

3.6.6. Theorem. *The logic* **KD45-P^G** *has the finite model property.*

Proof. By standard methods. □

Similarly, a representation theorem can be obtained by showing that the model could have been obtained from priority sequences $C_1 \gg_a C_2 \cdots \gg_a C_m (m \in \mathbb{N})$ for all the agents.

3.6.7. Theorem. (*representation theorem*). $\vdash_{\mathbf{KD45-P^G}} \varphi$ *iff* φ *is valid in all models with each* $\underline{Pref^a}$ *obtained from a priority sequence.*

Proof. Let there be k agents $a_0, \ldots, a_{k-1}$ and suppose $\varphi(d_1, \ldots, d_n)$. We provide each agent a_j with her own priority sequence $P_{n\times j+1} \gg_{a_j} P_{n\times j+2} \gg_{a_j} \ldots \gg_{a_j} P_{n\times(j+1)}$. It is sufficient to show that any model for **KD45-P^G** for the reduced language can be extended by valuations for the $P_j(d_i)$'s in such a way that the preference relations are preserved. For each a_i-euclidean class, we follow the same procedure for $d_1, \ldots, d_n$ w.r.t. $P_{n\times j+1}, P_{n\times j+2}, ..., P_{n\times(j+1)}$ as in Theorem 3.3.2 w.r.t $P_1, \ldots, P_n$. The preference orders obtained in this manner are exactly the $Pref^{a_j}$ relations in the model. □

In the above case, the priority sequences for different agents are separate, and thus very different. Still stronger representation theorems can be obtained by requiring that the priority sequences for different agents are related, e.g. in the case of *cooperative agents* that they are equal. We will consider the two agent case in the following.

3.6.8. THEOREM. (*for two cooperative agents*). $\vdash_{\mathbf{KD45-P^G}} \varphi$ *iff* φ *is valid in all models obtained from priority sequences shared by two cooperative agents.*

Proof. The 2 agents are a and b. We now have the priority sequence $P_1 \gg_a P_2 \gg_a ... \gg_a P_n$, same for b. It is sufficient to show that any model $\mathcal{M}$ with worlds W for **KD45-P$^{\mathbf{G}}$** for the reduced language can be extended by valuations for the $P_j(d_i)$'s in such a way that the preference relations are preserved. We start by making all $P_j(d_i)$'s true everywhere in the model. Next we extend the model as follows. For each a-euclidean class E in the model carry out the following procedure. Extend $\mathcal{M}$ with a complete copy $\mathcal{M}_E$ of $\mathcal{M}$ for all of the reduced language i.e. without the predicates P_j. Add R_a relations from any of the w in E to the copies v_E such that $w\, R_a\, v$. Now carry out the same procedure as in the proof of Theorem 3.3.2 in E's copy E_E. What we do in the rest of $\mathcal{M}_E$ is irrelevant. Now, in w, a will believe in $P_j(d_i)$ exactly as in the model in the previous proof, the overall truth of $P_j(d_i)$ in the a-euclidean class E in the original model has been made irrelevant. The preference orders obtained in this manner are exactly the $Pref^a$ relations in the model. All formulas in the reduced language keep their original valuation because the model $\mathcal{M}_E$ is bisimilar for the reduced language to the old model $\mathcal{M}$ as is the union of $\mathcal{M}$ and $\mathcal{M}_E$.

Finally do the same thing for b: add for each b-euclidean class in $\mathcal{M}$ a whole new copy, and repeat the procedure followed for a. Both a and b will have preferences with regard to the same priority sequence. □

For *competitive agents* we assume that if agent a has a priority sequence $D_1 \gg_a D_2 \gg \cdots \gg_a D_m (m \in \mathbb{N})$, then the opponent b has priority sequence $\neg D_m \gg_b \neg D_{m-1} \gg \cdots \gg_b \neg D_1$.

3.6.9. THEOREM. (*for two competitive agents*). $\vdash_{\mathbf{KD45-P^G}} \varphi$ *iff* φ *is valid in all models obtained from priority sequences for competitive agents.*

Proof. Let's assume two agents a and b. For a we take a priority sequence $P_1 \gg_a P_2 \gg_a \cdots \gg_a P_n \gg_a P_{n+1} \gg_a \cdots \gg_a P_{2n}$, and for b, we take $\neg P_{2n} \gg_b \neg P_{2n-1} \gg_b \cdots \gg_b \neg P_n \gg_b \neg P_{n-1} \gg_b \cdots \gg_b \neg P_1$. It is sufficient to show that any model $\mathcal{M}$ with worlds W for **KD45-P$^{\mathbf{G}}$** for the reduced language can be extended by valuations for the $P_j(d_i)$'s in such a way that the preference relations are preserved. We start by making all $P_1(d_i) \dots P_n(d_i)$ true everywhere in the model and $P_{n+1}(d_i) \dots P_{2n}(d_i)$ all false everywhere in the model. Next we extend the model as follows.

For each a-euclidean class E in the model carry out the following procedure. Extend $\mathcal{M}$ with a complete copy $\mathcal{M}_E$ of $\mathcal{M}$ for all of the reduced language i.e. without the predicates P_j. Add R_a relations from any of the w in E to the copies v_E such that $w\, R_a\, v$. Now define the values of the $P_1(d_i) \dots P_n(d_i)$ in E_E as in the previous proof and make all $P_m(d_i)$ true everywhere for $m > n$. The preference orders obtained in this manner are exactly the $Pref^a$ relations in the model.

For each b-euclidean class E in the model carry out the following procedure. Extend $\mathcal{M}$ with a complete copy $\mathcal{M}_E$ of $\mathcal{M}$ for all of the reduced language i.e. without the predicates P_j. Add R_b relations from any of the w in E to the copies v_E such that $w\, R_b\, v$. Now define the values of the $\neg P_{2n}(d_i) \dots \neg P_{n+1}(d_i)$ in E_E as for $P_1(d_i) \dots P_n(d_i)$ in the previous proof and make all $P_m(d_i)$ true everywhere for $m \leq n$. The preference orders obtained in this manner are exactly the $Pref^b$ relations in the model.

All formulas in the reduced language keep their original valuation because the model $\mathcal{M}_E$ is bisimilar for the reduced language to the old model $\mathcal{M}$ as is the union of $\mathcal{M}$ and all the $\mathcal{M}_E$. □

3.6.10. Remark. These last representation theorems show that they are as is to be expected not only a strength but also a weakness. The weakness here is that they show that cooperation and competition cannot be differentiated in this language. On the other hand, the theorems are not trivial, one might think for example that if a and b cooperate, $B_a Pref_b(c, d)$ would imply $Pref_a(c, d)$. This is of course completely false, a and b can even when they have the same priorities have quite different beliefs about how the priorities apply to the constants. But the theorems show that no principles can be found that are valid only for cooperating agents. Moreover they show that if one wants to prove that $B_a Pref_b(c, d) \rightarrow Pref_a(c, d)$ is not valid for cooperating agents a counterexample to it in which the agents do not cooperate suffices.

3.7 Preference over propositions

Most other authors on preference have discussed preference over propositions rather than objects. Our approach can be applied to preference over propositions as well. We are going to develop this ideas further in this section. As we know, preference is always intertwined with beliefs. In the following, we will propose a system combining them. And we specially take the line that preference is a state of mind and that therefore one prefers one alternative over another if and only if one believes one does. If we take this line, the most obvious way would be to go to second order logic and consider priority sequence $A_1(\varphi) \gg A_2(\varphi) \gg \dots, \gg A_n(\varphi)$, where the A_i are properties of propositions. However, we find it close to our intuitions to stay first order as much as possible. With that in mind, we define the new priority sequence for the propositional case as follows.

3.7.1. Definition. A *propositional priority sequence* is a finite ordered sequence of formulas written as follows

$$\varphi_1(x) \gg \varphi_2(x) \gg \cdots \gg \varphi_n(x) \quad (n \in \mathbb{N})$$

where each of $\varphi_m(x)$ is a propositional formula with an additional propositional variable, x, which is a common one to each $\varphi_m(x)$.

Formulas $\varphi(x)$ can express properties of propositions, for instance, applied to ψ, $x \to p_1$ expresses that ψ implies p_1, "ψ has the property" p_1.

We apply our approach in previous sections to define preference in terms of beliefs. As we have seen in Section 3.4, there are various ways to do it. We are guided by the definition of decisive preference in formulating the following:

3.7.2. Definition. Given a propositional priority sequence of length n, we define preference over propositions ψ and θ as follows:

$$\begin{aligned} Pref(\psi, \theta) \quad \text{iff} \quad & \text{for some } i\ (B(\varphi_1(\psi) \leftrightarrow B(\varphi_1(\theta)) \wedge \cdots \wedge (B(\varphi_{i-1}(\psi)) \leftrightarrow \\ & B(\varphi_{i-1}(\theta))) \wedge (B(\varphi_i(\psi) \wedge \neg B(\varphi_i(\theta)) \end{aligned}$$

Note that preference between propositions is in this case almost a preference between mutually exclusive alternatives: in the general case one can conclude beyond the quasi-linear order that derives directly from our method only that if $B(\psi \leftrightarrow \theta)$, then ψ and θ are equally preferable. Otherwise, any proposition can be preferable over any other.

For some purposes (this will get clearer in the proof of the representation theorem below), we need a further generalization, hence here we give a slightly more complex definition.

3.7.3. Definition. A *propositional priority sequence* is a finite ordered sequence of sets of formulas written as follows

$$\Phi_1 \gg \Phi_2 \gg \cdots \gg \Phi_n$$

where each set Φ_i consists of propositional formulas that have an additional propositional variable, x, which is a common one to each Φ_i.

A new definition of preference is given by:

3.7.4. Definition. Given a propositional priority sequence of length n, we define preference over propositions ψ and θ as follows:

$$\begin{aligned} Pref(\psi, \theta) \quad \text{iff} \quad & \exists i(\forall j < i(\exists \varphi \in \Phi_j(B\varphi(\psi)) \leftrightarrow \exists \varphi \in \Phi_j(B\varphi(\theta) \wedge \\ & \exists \varphi \in \Phi_i(B\varphi(\psi)) \wedge \forall \varphi \in \Phi_i \neg B(\varphi\theta))) \end{aligned}$$

3.7.5. Remark. In fact, the priority set Φ_m could be expressed by one formula

$$\bigvee_{\varphi\in\Phi_m} B\varphi.$$

But then we would have to use B in the formulas of the priority sequence, which we prefer not to.

The axiom system **BP** that arises from these considerations combines preference and beliefs in the following manner:

(a) $\underline{Pref}(\varphi,\varphi)$

(b) $\underline{Pref}(\varphi,\psi)\wedge\underline{Pref}(\psi,\theta)\rightarrow\underline{Pref}(\varphi,\theta)$

(c) $\underline{Pref}(\varphi,\psi)\vee\underline{Pref}(\psi,\varphi)$

(d) $B\underline{Pref}(\varphi,\psi)\leftrightarrow\underline{Pref}(\varphi,\psi)$

(e) $B(\varphi\leftrightarrow\psi)\rightarrow\underline{Pref}(\varphi,\psi)\wedge\underline{Pref}(\psi,\varphi)$.

As usual, it also includes *Modus ponens (MP)*, as well as the Generalization Rule for the operator B. The first three are standard for preference, and we have seen the analogue of (d) in Section 3.4. (e) is new, as a connection between beliefs and preference. It expresses that if two propositions are indistinguishable on the plausible worlds they should be equally preferable. It is easy to see that the above axioms are valid in the models defined as follows.

3.7.6. Definition. A model of **BP** is a tuple $\langle W, R, \{\preceq_w\}_{w\in W}, V\rangle$, where W is a set of worlds, R is a euclidean and serial accessibility relation on W. Namely, it satisfies $\forall xyz((Rxy\wedge Rxz)\rightarrow Ryz)$ and $\forall x\exists yRxy$. Moreover, for each w, $\preceq_w$ is a quasi-linear order on propositions (subsets of W), which is constant throughout each euclidean class and which is determined by the part of the propositions that lies within the 'plausibility part' of the euclidean class. V is an evaluation function in an ordinary manner.

3.7.7. Theorem. *The* **BP** *system is complete w.r.t the above models.*

Proof. Assume $\nvdash_{\mathbf{BP}}\theta$. Take the canonical model $\mathcal{M}=(W,R,V)$ for the formulas using only the propositional variables of θ. To each world of W a quasi-linear order of all formulas is associated, and it only depends on the extension of the formula (the set of nodes where the formula is true) in the plausible part of the model. This order is constant throughout the euclidean class defined by R. $\neg\theta$ can be extended to a maximal consistent set Γ. We consider the submodel generated by Γ, $\mathcal{M}'=(W',R,V)$, which naturally is an euclidean class. Since each world in W' has access to the same worlds, each world that satisfies the same atoms satisfies the same formulas. In fact, each formula φ in this model is equivalent to a purely propositional formula, a formula without B or $Pref$. To see this, one just has to realize that $B\psi$ is in the model either equivalent to $\top$ or $\bot$, and the same holds for $Pref(\psi, \theta)$. (Note that this argument only applies because we have just one euclidean class.) Now apply a p-morphism to $\mathcal{M}'$ which identifies worlds that satisfy the same formula. This gives a finite model consisting of one euclidean class with a constant order that still falsifies θ. Moreover, each world is characterized by a formula $\pm p_1 \wedge \cdots \wedge \pm p_k$ that expresses which atoms are true in it. In consequence, each subset of the model (proposition) is also definable by a purely propositional formula, a disjunction of the formulas $\pm p_1, \wedge \cdots \wedge \pm p_k$ describing its elements. □

Similarly, we can prove the representation result.

3.7.8. Theorem. (*representation theorem*) $\nvdash_{\mathbf{BP}} \varphi$ *iff* φ *is valid in all models obtained from priority sequences.*

Proof. The order of the finitely many formulas defining all the subsets of the models can be represented as a sequence

$$\Phi_1, \ldots, \Phi_k$$

where Φ_1 are the best propositions ($\varphi, \psi \in \Phi_1$ implies $\varphi \trianglelefteq \psi$ and $\psi \trianglelefteq \varphi$, Φ_i are the next best propositions, etc. Then the following is the priority sequence which results in the given order:

$$\{x \leftrightarrow \varphi \mid \varphi \in \Phi_1\} \gg \cdots \gg \{x \leftrightarrow \varphi \mid \varphi \in \Phi_k\}.$$

□

So far our discussions on the preference relation over propositions are rather general. We do not presuppose any restriction on such a relation. However, if we think that the preference relation over propositions is a result of lifting a preference relation over possible worlds (as discussed before), we specify its meaning in a more precise way, following the obvious option of choosing different combinations of quantifiers. For example, we can take $\forall\exists$ preference relations over the propositions, i.e. preference relations over propositions lifted from preference relations over worlds in the $\forall\exists$ manner. Regarding the axiomatization, we will then have to add the following two axioms to the above **BP** system, the new system will denoted as $\mathbf{BP}^{\forall\exists}$. It has two more axioms:

- $B(\varphi \to \psi) \to \underline{Pref}(\psi, \varphi)$.
- $\underline{Pref}(\varphi, \varphi_1) \wedge \underline{Pref}(\varphi, \varphi_2) \to \underline{Pref}(\varphi, \varphi_1 \vee \varphi_2)$

3.7.9. Theorem. *$BP^{\forall\exists}$ is complete.*

Proof. By an adaption of the proof by [Hal97]. The difference is: [Hal97] uses a combination of preference and universal modality. Instead, our system is a combination of belief and preference. This means what is preferred in our system is decided by the plausible part of the model. However, this will not affect the completeness proof much. □

3.7.10. Remark. In fact, $[pref]\varphi$ in Chapter 2 can be defined now as $\underline{Pref}(\varphi, \top)$. Then the preference used in the system $\mathbf{BP}^{\forall\exists}$ is simply the following:

$$\underline{Pref}(\varphi, \psi) \leftrightarrow B(\psi \to \langle pref \rangle \varphi)$$

We will come back to this point in Chapter 4.

Similarly, we get the representation result for this restricted case:

3.7.11. Theorem. (*representation theorem*) *$\vdash_{\mathbf{BP}^{\forall\exists}} \varphi$ iff φ is valid in all $\forall\exists$-models obtained from priority sequences.*

The proof is same as for the basic system.

Finally, to conclude this subsection, recall that we had a logic system to discuss preference over objects when beliefs are involved. With our new system just presented, we can talk about preference over propositions. But what is the relation between these two systems? The following theorem provides an answer.

3.7.12. Theorem. *$\vdash_{\mathbf{KD45-P}} \varphi(d_1, \ldots, d_n)$ iff $\vdash_{\mathbf{BP}} \varphi(p_1, \ldots, p_n)$ where the propositional variables $p_1, \ldots, p_n$ do not occur in $\varphi(d_1, \ldots, d_n)$.*

Proof. In order to prove this theorem, we need to prove the following lemma:

3.7.13. Lemma. *If $\nvdash_{\mathbf{KD45-P}} \varphi(d_1, \ldots, d_n)$, then for each n there is a model $\mathcal{M} \models \neg\varphi$ with at least n elements.*

Proof. Assume that we only have a model $\mathcal{M} = (W, R, V)$ in which W has m elements, where $m < n$. Take one element of W, say w, and make copies of it, say, w_1, w_2,..., w_k, till we get at least n elements. If wRv, then we make w_iRv, and if vRw, then vRw_i. In this way we get a new model with at least n elements. It is bisimilar to the original model. □

Now we are ready to prove the theorem.
($\Rightarrow$). It is easy to see that all the **KD45-P** axioms and rules are valid in **BP** if one replaces each d_i by p_i.

($\Leftarrow$). It is sufficient to transform any finite **KD45-P** model $\mathcal{M}$ with only one euclidean class into a **BP** model $\mathcal{M}'$ with at least n possible worlds in which for each w and each ψ, $\mathcal{M}', w \models \psi(p_1, \ldots, p_n)$ iff $\mathcal{M}, w \models \psi(d_1, \ldots, d_n)$. Let $\mathcal{M} = (W, R, \preceq, V)$, then $\mathcal{M}' = (W', R, \trianglelefteq, V')$, where V' is like V except that for the $p_1, \ldots, p_n$, we assign $V'(p_i) = V'(p_j)$ if $d_i \preceq d_j \wedge d_j \preceq d_i$, otherwise, $V'(p_i) \neq V'(p_j)$. [5] According to Lemma 3.7.13, there are enough subsets to do this. Finally, we set $V'(p_i) \lhd V'(p_j)$ iff $d_i \prec d_j$ and extend $\lhd$ to other sets in an arbitrary manner. □

If one thinks of propositional variables as representing basic propositions, then this theorem says that reasoning about preference over objects is the same as reasoning about preference over basic propositions. This is not surprising if one thinks of basic propositions as exclusive alternatives as are objects. Of course, the logic of preference over propositions in general is more expressive. One can look at this latter fact in two different ways: (a) one may think the logic over preference over all propositions as essentially richer than the logic of the basic propositions or objects, or (b) one may think that the essence of the logic of propositions is contained in the basic propositions (represented by the propositional variables) and the rest needs to be carried along in the theory to obtain a good logical system but is of little value by itself.

By applying the method of [Hal97] we can adapt the above proof to obtain the following:

3.7.14. THEOREM. $\vdash_{\mathbf{KD45-P}} \varphi(d_1, \ldots, d_n)$ *iff* $\vdash_{\mathbf{BP}^{\forall\exists}} \varphi(p_1, \ldots, p_n)$ *where the propositional variables* $p_1, \ldots, p_n$ *do not occur in* $\varphi(d_1, \ldots, d_n)$.

Up to now we have used decisive preference. Another option is to use deliberate preference. Let us look at this in a rather general manner. Assume that $Supe(\varphi, \psi)$ has the property in a model that for each φ, ψ,

$$\models (\varphi \leftrightarrow \varphi') \wedge (\psi \leftrightarrow \psi') \rightarrow (Supe(\varphi, \psi) \leftrightarrow Supe(\varphi', \psi')),$$

we then say 'superior' is a *local property* in that model. We can now state the following propositions.

3.7.15. THEOREM. *If we define* $Pref(\varphi, \psi)$ *as* $B(Supe(\varphi, \psi))$ *in any model where* $Supe(\varphi, \psi)$ *is a local partial order, then* $Pref(\varphi, \psi)$ *satisfies the principles of* **BP**, *except possibly connectedness.*

It is to be noted that

$$\varphi \rightarrow \langle pref \rangle \psi$$

[5] Note that the $V'(p_i)$ are only relevant for the ordering $\trianglelefteq$ because the p_i's only occur directly under the $Pref$ in $\varphi(p_1, \ldots, p_n)$.

is not a local property even if $\preceq$ is a subrelation of R. Nevertheless, in case $\preceq$ is a subrelation of R, $B(\varphi \to \langle pref \rangle \psi)$ does satisfy the principles of **BP** minus connectedness, and the additional $\mathbf{BP}^{\forall\exists}$ axioms, as we commented in Remark 3.7.10. For this purpose the following weakening of locality is sufficient:

$$\models (\varphi \leftrightarrow \varphi') \wedge B(\varphi \leftrightarrow \varphi') \wedge (\psi \leftrightarrow \psi') \wedge B(\psi \leftrightarrow \psi') \to (Supe(\varphi, \psi) \leftrightarrow Supe(\varphi', \psi')).$$

3.8 Discussion and conclusion

Partially ordered priority sequence

A new situation occurs when there are several priorities of incomparable strength. Take the Example 3.1.1 again, however, instead of considering three properties, Alice also takes the 'transportation convenience' into her account. But for her neighborhood and transportation convenience are really incomparable. Abstractly speaking, this means that the priority sequence is now *partially ordered.* We show in the following how to define preference based on a partially ordered priority sequence. In other words, we consider a set of priorities $C_1..., C_n$ with the relation $\gg$ between them a partial order.

3.8.1. Definition. We define $Pref_n(x, y)$ by induction, where $\{n_1, ..., n_k\}$ is the set of immediate predecessors of n.

$$\underline{Pref_n(x,y)} ::= \underline{Pref_{n_1}}(x,y) \wedge ... \wedge \underline{Pref_{n_k}}(x,y) \wedge ((C_n(y) \to C_n(x)) \vee (Pref_{n_1}(x,y) \vee ... \vee Pref_{n_k}(x,y)))$$

where as always $Pref_m(x, y) \leftrightarrow \underline{Pref_m}(x, y) \wedge \neg \underline{Pref_m}(y, x)$

This definition is, for finite partial orders, equivalent to the one in [Gro91] and [ARS02]. More discussion on the relation between partially ordered priorities and G-spheres, see [Lew81]. When the set of priorities is unordered, again, we refer to [Kra81]. We come back to this issue in Chapter 4.

Conclusion

In this chapter we considered preference over objects. We showed how this preference can be derived from priorities, properties of these objects. We did this both in the case when an agent has complete information and in the case when an agent only has beliefs about properties. We considered both the single and multi-agent case. In all cases, we constructed preference logics, some of them extending the standard logic of belief. This leads to interesting connections between preference and beliefs. We strengthened the usual completeness results for logics of this kind to representation theorems. The representation theorems describe the reasoning that is valid for preference relations that have been obtained from priorities. In the multi-agent case, these representation theorems are strengthened to special cases of cooperative and competitive agents. We studied preference change with regard to changes of the priority sequence, and change of beliefs. We applied the dynamic epistemic logic approach, and in consequence reduction axioms were presented. We proposed a new system combining preference and beliefs, talking about preference over propositions. We concluded by some discussion on generalizing the linear orders in this chapter to partial orders.