Chapter 9 PREFERENCE STATICS AND DYNAMICS

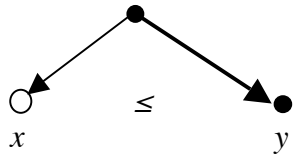
So far, we have shown how logical dynamics deals with agents' knowledge and beliefs, and informational events that change these. But as we noted in Chapter 1, agency also involves a second major system, not of information but of *evaluation*. It is values mixed with information that provide the driving force for rational action – and the colour of life. The barest record of evaluation are agents' *preferences* between worlds or actions. Thus, the next task in this book is dealing with preferences, and how they change under triggers like suggestions or commands. While this topic seems different in flavour from earlier ones, properly viewed, it yields to the same techniques as in Chapters 3, 7. Therefore, we will present our dynamic logics with a lighter touch, while emphasizing further interesting features of preference that make it special from a logical perspective.

9.1 Logical dynamics of practical reasoning

Here are a few examples showing how preferences function in scenarios of agency.

Example Decision problems.

A standard decision problem looks at actions available to agents, and then asks what they will, or should, do based on their preference between the outcomes:



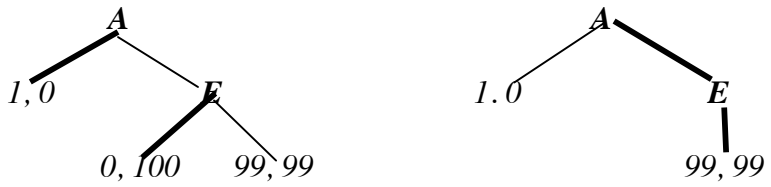
If an agent can choose between x and y, and she prefers y, then she should choose y. ■

Information and evaluation also come together in *games*, looking at what players want, observe, and guess, and which moves are available to achieve their goals. In this setting, multi-agent interaction is essential, as my actions depend on what I think about yours.

Example Reasoning about interaction.

Here is an example that was already discussed in Chapter 1. In the following two game trees, preferences at end nodes are encoded in utility pairs *(value of **A**, value for **E**)*. The solution method of *Backward Induction* (cf. Chapter 10) is a typical piece of multi-agent reasoning about preference and belief. In the game to the right, essentially a single-agent decision problem, it tells player ***A*** to go right, where ***E*** takes both to the desirable outcome *(99, 99)*. But interaction with more agents can be tricky. In the game on the left, Backward

Induction tells ***E*** to turn left when she can, and then ***A*** (who realizes what ***E*** will do) will turn left at the start – where both suffer, since their pay-off is much lower than *(99, 99)*:



Why should players act this way? The reasoning is a mixture of all notions so far. ***A*** turns left since she believes that ***E*** will turn left, and then her preference is for grabbing the value *1*. Thus, practical reasoning intertwines action, preference, and belief. ■

We will return to this scenario in Chapter 10, but for now, it shows the sort of preference ordering we want to study. In this chapter, we focus on preference alone, bringing in belief and action later. One area where this makes sense is *deontic logic,* the study of reasoning about agents' obligations. The latter are usually expressed in propositions that are true in the best of all worlds according to some moral authority. Moreover, the relevant 'better' order may change as commands come in, or new laws are enacted. Thus, deontic logic is a multi-agent preference logic, involving myself and one or more moral authorities. [189]

In these examples, we see an order of 'better', and we see maximizing along it in the notion 'best'. Thus, logics of preference can be designed in the style of plausibility models for belief (Chapter 7), while the methods of Chapters 3, 4 can deal with preference change. For a start, here are some ingredients from the literature:

Preference logics Von Wright 1963 proposed a logic with formulas $P\varphi\psi$ saying that every φ–situation is preferred over every ψ–situation 'ceteris paribus': a phrase to which we will return. Note that preference here is 'generic', running between propositions, i.e., sets of situations. Von Wright's calculus for reasoning with preference contains laws such as

$$P\varphi\psi \leftrightarrow P(\varphi \wedge \neg\psi)(\psi \wedge \neg\varphi)$$

This has led to current preference logic (cf. Hanson 2001), for which we will use a modal language below. Beyond this, there is a recent interest in *preference change* and its triggers (Gruene & Hanson eds. 2008). Our chapter will develop this evaluation dynamics in tandem with information dynamics, as we want to understand the entanglement.

---

[189] The ideal situation might make my personal preference order coincide with the moral one, as in Kant's dictum that one should make duty coincide with inclination.

9.2 Modal logic of betterness

Preference is multi-faceted: we can prefer one individual object, or one situation, over another – but preference can also be directed toward kinds of objects or generic situations, defined by propositions. A bona fide preference logic should do justice to both views. We start with a simple scenario on the object/world side, moving to generic notions later.

Basic models In this chapter, we start with a very simple semantic setting:

Definition Modal betterness models.
Modal betterness models $\boldsymbol{M} = (W, \leq, V)$ have a set of worlds W, [190] a reflexive and transitive *betterness relation* $x \leq y$ ('world y is at least as good as world x'), and a valuation V for proposition letters at worlds (or equivalently, for unary properties of objects). ■

In practice, the relation may vary among agents, but we suppress subscripts $\leq_i$ for greater readability. We use the artificial term 'betterness' to stress that this is an abstract comparison, making no claim about the intuitive term 'preference', whose uses are diverse. These models occur in decision theory, where worlds are outcomes of actions, and game theory, where worlds are complete histories, with preferences for different players. The same orders were also used in Chapter 7 for relative plausibility as judged by an agent. While preference is not the same as plausibility, the analogy is helpful.

Modal languages Over our base models, we can interpret a standard modal language, and see which natural notions and patterns of reasoning can be defined in it.

Definition Preference modality.
A modal assertion $<\leq>\varphi$ makes the following local assertion at a world w:

$$\boldsymbol{M}, w \models <\leq>\varphi \quad \text{iff} \quad \text{there exists a } v \geq w \text{ with } \boldsymbol{M}, v \models \varphi$$

that is: there is a world v at least as good as w that satisfies φ. ■

In combination with other modal operators, this sparse-looking betterness formalism can express many natural notions concerning preference-driven action.

Example Defining backward induction in preference action logic.
Finite game trees are models for a dynamic logic of atomic actions (players' moves) and unary predicates indicating players' turns at intermediate nodes and their utility values at

[190] These really stand for any sort of objects that are subject to evaluation and comparison.

end nodes (van Benthem 2002). In Chapter 10, we present a result from van Benthem, van Otterloo & Roy 2006 showing how the *backward induction solution* of a finite game [191] is the unique binary relation *bi* on the tree satisfying this modal preference-action law:

$$<bi>[bi^*](end \rightarrow \varphi) \rightarrow [move]<bi^*>(end \wedge <\leq>\varphi)$$

Here *move* is the union of all move relations available to players, and * is reflexive-transitive closure. The formula says there is no alternative to the *BI*-move at the current node all of whose outcomes would be better than the *BI*-solution. ■

Thus, modal preference logic goes well with games. Boutilier 1994 showed how it can also define conditionals (Lewis 1973), analyzing conditional logic in standard terms. We used this in Chapter 7. On finite reflexive and transitive orders, the following formula defines a conditional $A \Rightarrow B$ in the sense of '*B* is true in all maximal *A*-worlds':

$$U(A \rightarrow <\leq>(A \wedge [\leq](A \rightarrow B))), \qquad \text{with } U \text{ the universal modality.}$$ [192]

The same expressive power will be relevant for preference.

Constraints on betterness Which properties should betterness have? *Total orders* satisfying reflexivity, transitivity, and connectedness are the norm in decision theory and game theory, as these properties relate to numerical utilities. But in the logical literature on preference or plausibility, even transitivity has been criticized (Hanson 2001). And in conditional logic, Lewis' totality is often abandoned in favour of *pre-orders* satisfying just reflexivity and transitivity, while acknowledging *four* irreducible basic relations:

$w \leq v, \neg\, v \leq w$ (often written as $w < v$)	*w strictly precedes v*
$v \leq w, \neg\, w \leq v$ (often written as $v < w$)	*v strictly precedes w*
$w \leq v, v \leq w$ (sometimes written as $w \sim v$)	*w, v are indifferent*
$\neg\, w \leq v, \neg\, v \leq w$ (sometimes written as $w \# v$)	*w, v are incomparable*.

[191] A famous benchmark example in the logical analysis of games; cf. Harrenstein 2004.

[192] The modal language also easily defines variants, such as the existential 'each *A*-world sees *at least one* maximal *A*-world that is *B*'. Axiomatizing inference with these and other defined notions *per se* is the point of completeness theorems in conditional logic. For preference, Halpern 1997 explicitly axiomatized a defined notion of preference of our later universal-existential type $\forall\exists$.

We prefer such a large class of models, with extra modal axioms if we want the relation to satisfy further constraints. The point of a logical analysis is to impose structure where needed, but also, to respect the right 'degrees of freedom' in an intuitive notion.

Further relations, further modalities? Given this, one can start with two relations: a weak order $w \leq v$ ('at least as good') and a strict order $w < v$ ('better'; $w \leq v \wedge \neg v \leq w$). Van Benthem, Girard & Roy 2007 axiomatize this language using separate modalities.

Example Frame correspondence for weak/strict betterness modalities.
By a standard modal argument, the axiom $(\psi \wedge \langle\leq\rangle\varphi) \rightarrow (\langle<\rangle\varphi \vee \langle\leq\rangle(\varphi \wedge \langle\leq\rangle\psi)$ corresponds to the first-order frame property that $\forall x \forall y\, (x \leq y \rightarrow (x < y \vee y \leq x))$. ■

For much more on modal preference logic, see the dissertation Girard 2008.

9.3 Defining global propositional preference

As we have said, a betterness relation need not yet determine what we mean by agents' preferences in a more colloquial sense. Many authors consider preference a generic relation between propositions, with von Wright 1963 as a famous example. [193]

Varieties of set lifting Technically, defining preferences between propositions calls for a comparison of sets of worlds. For a given relation $\leq$ among worlds, this may be achieved by *lifting*. One ubiquitous proposal in betterness lifting is the $\forall\exists$ stipulation that

a set Y is preferred to a set X if $\forall x \in X\ \exists y \in Y: x \leq y$.

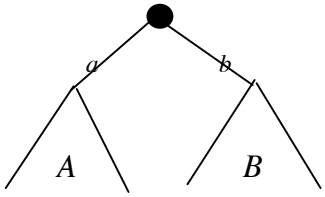
Van Benthem, Girard & Roy 2008 analyze von Wright's view as the $\forall\forall$ stipulation that

a set Y is preferred to a set X if $\forall x \in X\ \forall y \in Y: x \leq y$,

and provide a complete logic. Liu 2008 provides a history of proposals for relation lifting in various fields (decision theory, philosophy, computer science), but no consensus has emerged. This may be a feature, rather than a bug. Preference between propositions may be genuinely different depending on the scenario, and then logic should not choose:

Example Different set preferences in games.
Comparing sets of outcomes A, B reached by available moves, players have options:

[193] These differences are largely terminological – which is why debates are often bitter.



One might prefer a set whose minimum utility value exceeds the maximum of another,

this is the $\forall\forall$ reading $\quad max(A) < min\ (B)$

but it is also reasonable to have the maximum of one set exceed that of the other,

the $\forall\exists$ reading $\quad max(A) < max(B)$

where each value in A has at least one higher value in B. Or, a pessimist might have the minimum of the preferred set higher than that of the other. There is no best choice. ■

Extended modal logics Many different liftings are definable in a modal base logic extended with a universal modality $U\varphi$: 'φ is true in all worlds' (cf. Chapter 2). This standard feature gives some additional expressive power without great cost in the modal model theory and the computational complexity of valid consequence. For instance, the $\forall\exists$ reading of preference is expressed as follows, with formulas for definable sets of worlds:

$U(\varphi \rightarrow <\leq>\psi)$.

In what follows, we will use the notation $P\varphi\psi$ for lifted propositional preferences. [194]

9.4 Dynamics of evaluation change

But now for preference change. A modal model describes a current evaluation pattern for worlds, as seen by one or more agents. But the reality is that these patterns are not stable. Things happen that make us *change* our evaluations. This dynamics has been in the air, witness our later references, leading up to mechanisms of relation change much like those discussed for plausibility in Chapter 7. [195] Realistic preference change has further features, seen with a deeper analysis of agents (Hanson 1995, Lang & van der Torre 2008). In this chapter, we show how the simpler version fits with the perspective of this book.

---

[194] One can also use stronger (first-order) logics to describe preferences. This is the balance in logic between illuminating definitions of key notions and argument patterns and computational complexity (cf. Chapter 2). Richer languages are fine, but modal logic is a good place to start.

[195] We only discuss one logic strand here: cf. again Hanson 1995 for a different point of entry.

9.5 A basic dynamic preference logic

We start with a very simple scenario from van Benthem & Liu 2007.

Dynamic logic of suggestions Betterness models are as before, and so is the modal base language with modalities $<\leq>$ and U. But the syntax now adds, for each formula of the language, a model-changing action $\#\varphi$ of 'suggestion' [196], defined as follows:

Definition Ordering change by suggestion.

For each model $\boldsymbol{M}$, w and formula φ, the *suggestion function* $\#\varphi$ returns the model $\boldsymbol{M}\#\varphi$, w equal to $\boldsymbol{M}$, w, but for the new relation $\leq' = \leq - \{(x, y) \mid \boldsymbol{M}, x \models \varphi \ \& \ \boldsymbol{M}, y \models \neg\varphi\}$. [197] ■

Next, we enrich the language with action modalities interpreted as follows: [198]

$$\boldsymbol{M}, w \models [\#\varphi]\psi \quad \text{iff} \quad \boldsymbol{M}\#\varphi, w \models \psi$$

These talk about what agents will prefer after their comparison relation has changed. For instance, if you tell me to drink beer rather than wine, and I accept this, then I now come to prefer beer over wine, even if I did not do so before.

Now, as in dynamic epistemic logic, the heart of the analysis is the recursion equation explaining when a preference obtains after an action. Here is the valid principle for suggestions, whose two cases follow the definition of the above model change:

$$<\#\varphi><\leq>\psi \ \leftrightarrow \ (\neg\varphi \wedge <\leq><\#\varphi>\psi) \vee (\varphi \wedge <\leq>(\varphi \wedge <\#\varphi>\psi)$$

Theorem The dynamic logic of preference change under suggestions is axiomatized completely by the static modal logic of the underlying model class plus the following equivalences for the dynamic modality:

$$\begin{array}{lcl}
[\#\varphi]\, p & \leftrightarrow & p \\
[\#\varphi]\, \neg\psi & \leftrightarrow & \neg[\#\varphi]\psi \\
[\#\varphi](\psi\wedge\chi) & \leftrightarrow & [\#\varphi]\psi \wedge [\#\varphi]\chi \\
[\#\varphi]U\psi & \leftrightarrow & U[\#\varphi]\psi \\
[\#\varphi]<\leq>\psi & \leftrightarrow & (\neg\varphi \wedge <\leq>[\#\varphi]\psi) \vee ((\varphi \wedge <\leq>(\varphi \wedge [\#\varphi]\psi).
\end{array}$$

---

[196] This is of course just an informal reading, not a full-fledged social analysis of suggestion.

[197] In this chapter '$\#\varphi$' stands for an act of suggesting that φ. Please do not confuse this with the notation '$\#\varphi$' for an act of *promotion* for φ in the syntactic awareness dynamics of Chapter 5.

[198] Here the syntax is recursive: the formula φ may itself contain dynamic modalities.

Proof These axioms say that (i) upgrade for a suggestion does not change atomic facts, (ii) upgrade is a function, (iii) its modality is a normal one, (iv) upgrade does not change the set of worlds, and crucially, (v) the upgrade modality encodes the betterness effect of a suggestion. Applied inside out, these principles reduce any valid formula to an equivalent one without dynamic modalities, for which the base logic is complete by assumption. ■

This logic automatically gives us a dynamic logic of upgraded propositional preferences.

Example Recursion laws for generic preferences.

Using the axioms, one computes how $\forall\exists$–type preferences $P\psi\chi$ change along:

$$[\#\varphi]P\psi\chi \leftrightarrow [\#\varphi]U(\psi \rightarrow \langle\leq\rangle\chi) \leftrightarrow$$
$$U[\#\varphi](\psi \rightarrow \langle\leq\rangle\chi) \leftrightarrow U([\#\varphi]\psi \rightarrow [\#\varphi]\langle\leq\rangle\chi) \leftrightarrow$$
$$U([\#\varphi]\psi \rightarrow (\neg\varphi \wedge \langle\leq\rangle[\#\varphi]\chi) \vee ((\varphi \wedge \langle\leq\rangle(\varphi \wedge [\#\varphi]\chi) \leftrightarrow$$
$$P([\#\varphi]\psi \wedge \neg\varphi)[\#\varphi]\chi \wedge P([\#\varphi]\psi \wedge \varphi)(\varphi \wedge [\#\varphi]\chi).$$ ■

General relation transformers This is just a trial run. Other relation transformers for betterness act on other triggers, and we aim for the same generality as in Chapter 7:

Example Drastic commands.

Let $\Uparrow\varphi$ be the radical trigger that makes all φ–worlds better than all $\neg\varphi$-ones, keeping the old order otherwise. This is stronger than a suggestion, making φ most desirable. Then we can use the axiom in Chapter 7 for safe belief, now using an existential modality E:

$$[\Uparrow\varphi]\langle\leq\rangle\psi \leftrightarrow (\neg\varphi \wedge E(\varphi \wedge [\Uparrow\varphi]\psi)) \vee (\neg\varphi \wedge \langle\leq\rangle(\neg\varphi \wedge [\Uparrow\varphi]\psi))$$
$$\vee (\varphi \wedge \langle\leq\rangle(\varphi \wedge [\Uparrow\varphi]\psi))$$

The three clauses follow the three cases in the definition of radical upgrade. ■

These are just technical examples. Further betterness changes encode how people respond to what others claim or command, with a variety as with policies for belief revision – but we leave that to applications. A dynamic logic of preference should provide the right generality in triggers for upgrade. One general format are the *PDL* programs of Chapters 4, 7, involving *test*, *sequential composition* and *union*, as in this earlier example: [199]

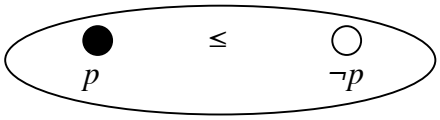
Fact Suggestion is the map $\#\varphi(R) = (?\varphi ; R ; ?\varphi) \cup (?\neg\varphi ; R ; ?\neg\varphi) \cup (?\neg\varphi ; R ; ?\varphi)$.

---

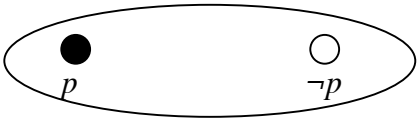
[199] Cf. also van Eijck's commentary on Sandu's chapter in Apt & van Rooij, eds., 2008.

Constraints on betterness order once more Suppose that betterness satisfies constraints, will its transformed version still satisfy these? Indeed, the above suggestions take pre-orders to pre-orders, but they can destroy the *totality* of a betterness order:

Example Losing connectedness.
Suppose the agent prefers $\neg P$ over P as in the following connected model:



A suggestion $\#P$ will turn this into the following non-connected model:



■

Some people see this loss of basic properties as a drawback of relation transformers. But we feel that the situation is the other way around. The fact that some natural relation transformers break certain relational constraints on preference shows how fragile these constraints really are, and they provide natural scenarios for counter-examples.

Coda: what was versus what should become the case Especially in deontic logic, it is tempting to read upgrade commands $\#\varphi$ as 'come to prefer that φ', or 'your new duty will become to achieve φ'. Duties are often about what you should make the world look like. This is the *forward-oriented* view discussed in Chapters 3, 4, 11: one must produce a change making some postcondition on betterness true. Our approach is *backward-oriented*, defining upgrades in terms of truth in the initial model – but the two perspectives co-exist peacefully for purely factual assertions. The contrast also comes up in related practical settings (cf. the notion of '*FIAT* in Zarnic 1999), and *DEL* extended with factual change by stipulating postconditions (cf. Chapter 4) would also make sense here.

9.6 An alternative semantics: constraint-based preference

Now we come to a major feature of preference that was not part of our study of belief and plausibility order in Chapter 7. So far, we started from a betterness ordering of worlds, and then defined lifted notions of preference between propositions, i.e., properties of worlds. But another approach works in the opposite direction. Object comparisons are often made on the basis of *criteria*, and derived from how we apply these criteria, and prioritize them. Cars may be compared as to price, safety, and comfort, in some order of importance. On that view, criteria are primary, object order is derived. In our setting, criteria would be

properties of worlds, expressed in propositions. This idea occurs in philosophy, economics (Rott 2001), linguistics and cognitive science. We will now develop this alternative:

First-order priority logic A recent logic for this view of preference is found in de Jongh & Liu 2007. Take any finite linear *priority sequence* $\boldsymbol{P}$ of propositions, expressing the importance an agent attaches to the corresponding properties:

Definition Object preference from priority.
Given a priority sequence $\boldsymbol{P}$, the *derived object order* $x < y$ holds iff x, y differ in at least one property in $\boldsymbol{P}$, and the first $P \in \boldsymbol{P}$ where this happens is one with $Py, \neg Px$. ■

This is really a special case of the well-known notion of lexicographic ordering, if we view each property $P \in \boldsymbol{P}$ as inducing the following simple object order: [200]

$$x \leq^{P} y \text{ iff } (Px \rightarrow Py).$$

De Jongh and Liu give a complete first-order logic for induced preferences between objects. It hinges on the following representation result for object or world models:

Theorem The orders produced via linear priority sequences are precisely the total ones with *reflexivity*, *transitivity*, and *quasi-linearity*: $\forall xyz: x \leq y \rightarrow (x \leq z \vee z \leq y)$.

Liu 2008 extends this to betterness *pre-orders* induced, not by linear sequences but by the *priority graphs* of Andréka, Ryan & Schobbens 2002 (cf. Chapter 12). These can model more realistic situations where criteria may be incomparable. She also notes that there are many ways of defining object order from property order, that can be studied similarly. This diversity may be compared with that for lifting object order to world order.

This view makes sense much more generally than only for preference. One can also take a priority approach to belief, deriving plausibility order of worlds from an *entrenchment order* of propositions (cf. Gärdenfors & Rott 1995). Further interpretations will follow.

Dynamics This framework, too, facilitates preference change. This time, the priority order and set of relevant properties can change: a new criterion may come up, or a criterion may lose importance. Four main operations are *permuting* properties in a sequence, *prefixing* a new property, *postfixing* a new property, and *inserting* a property. Together, these allow

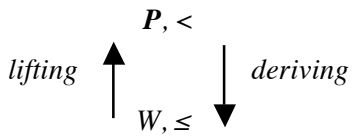
[200] We will be free-wheeling in what follows between weak versions $\leq$ and strict ones $<$;
but everything we say applies equally well to both versions and their modal axiomatizations.

for any manipulation of finite sequences. De Jongh & Liu 2007 give a complete dynamic logic in our style, taking our methods to first-order logics. This can be generalized again to non-linear priority graphs, leading to the algebra of graph operations in Andréka, Ryan & Schobbens 2002 (cf. Chapter 12), in particular, sequential and parallel composition.

Again, there are many interpretations for this dynamics. For instance, in a deontic setting, the priority graph may be viewed as a system of laws or norms, and changes in this structure model the adoption of new laws, or the appearance (or disappearance) of norms (Grossi & Liu 2010). Girard 2008 interprets priority graphs as agendas for investigation, and links agenda change to evolving research programs in the philosophy of science.

Two-level connections The two views of preference are not rivals, but complementary. One either starts from a betterness relation between worlds and lifts this to propositional preference, or one starts from a importance order of propositions, and derives world order. One can combine these perspectives in interesting ways (Liu 2008):

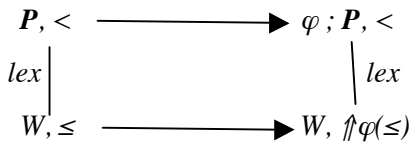
Definition Two-level structures.
Preferential *two-level structures* $(W, \leq, \boldsymbol{P}, <)$ have both worlds with a betterness order $\leq$ and a set of 'important propositions' with a primitive priority pre-order $<$:



■

This picture suggests various interpretations, and many new questions. [201] We just state an elegant correspondence between the dynamics at the two levels. Prefixing propositions φ to a current priority sequence $\boldsymbol{P}$ has the same effect as the earlier relation transformer $\Uparrow\varphi$. More precisely, writing the lexicographic derivation of object order as a function *lex*:

Fact The identity $lex(\varphi ; \boldsymbol{P}) = \Uparrow\varphi\,(lex(\boldsymbol{P}))$ holds, making this diagram commute:



[201] What happens when we derive betterness order from priority, and then lift it again – and vice versa? What happens when we treat the propositions in $\boldsymbol{P}$ as distinguished propositional constants in a modal language, and relate modal betterness logic with a modal constraint logic?

Proof The definition of radical upgrade $\Uparrow\varphi$ really says that being φ is the first over-riding priority, while after that, the old order is followed. [202] ■

The general theory of inducing dynamics from one level to another seems open.

9.7 Further aspects of preference: ceteris paribus logic

We have now discussed a basic modal approach to preference, as well as an alternative criteria-based view. Next comes one more major feature that presents a challenge to logic. Preferences usually hold only *ceteris paribus*, that is, under relevant circumstances. Van Benthem, Girard & Roy 2009 describe how to accommodate this in our logics:

Normality versus equality The term 'ceteris paribus' has two meanings. The *normality sense* says that we only make a comparison under normal circumstances. I prefer beer over wine, but not when dining at the Paris Ritz. This may be modeled by the 'normal' or most plausible worlds of a current model, as in Chapter 7.

Definition Preference under normal circumstances.

A global preference $P\varphi\psi$ holds in the *normality sense*, in any of the earlier lifted senses, if the latter holds when restricted to the doxastically most plausible worlds of the model. ■

If the set of normal worlds is definable by some proposition N, we can state this in our base logic as $P(N\wedge\varphi)(N\wedge\psi)$, using any relevant P. But in general, we need both betterness and plausibility orders, as in Lang, van der Torre & Weydert 2003, with a matching combined logic of preference and belief. We will return to this in our next section.

In the *equality sense* of ceteris paribus, preference holds under the proviso that some propositions do not change truth values. You may prefer night over day, but only with 'work versus vacation' fixed (there may be vacation days that you prefer to work nights).

Definition Equality-based ceteris paribus preference.

A *ceteris paribus preference* for φ over ψ *in the equality sense* with respect to proposition A means that (i) among the A–worlds, the agent prefers φ over ψ, and (ii) among the $\neg A$-worlds, the agent prefers φ over ψ. Here, preference can be any of our lifted notions. ■

---

[202] This is an instance of the general algebraic calculus of priority graphs discussed in Chapter 12 – in particular, its law for sequential composition. For the special case of property graphs, there are also further laws. For instance, each such graph has a graph of *disjoint* properties generating the same object order. Rules for finding the latter merge Boolean algebra with preference logic.

On this second account, cross-comparisons between the A and $\neg A$ worlds are irrelevant to a preference. [203] With a set of relevant propositions $\boldsymbol{A}$, one looks at the equivalence classes of worlds under the relation $\equiv_{\boldsymbol{A}}$ of sharing the same truth values on the $\boldsymbol{A}$'s. [204] This relation has also been studied as an account of *dependence* and independence of propositions (Doyle & Wellman 1994). It also occurs in the semantics of questions in natural language (ten Cate & Shan 2002), and with supervenience in philosophy. [205]

Equality-based ceteris paribus preference logic Van Benthem, Girard & Roy 2008 make equality-based ceteris paribus preferences an explicit part of the modal language.

Definition Ceteris paribus modal preference logic.
The modal logic *CPL* extends basic model preference logic with operators defined as

$$\boldsymbol{M}, s \models [\Gamma]\varphi \quad \text{iff} \quad \boldsymbol{M}, t \models \varphi \text{ for all } t \text{ with } s \equiv_\Gamma t,$$
$$\boldsymbol{M}, s \models [\Gamma]^{\leq}\varphi \quad \text{iff} \quad \boldsymbol{M}, t \models \varphi \text{ for all } t \text{ with } s \equiv_\Gamma t \text{ and } s \leq t,$$
$$\boldsymbol{M}, s \models [\Gamma]^{<}\varphi \quad \text{iff} \quad \boldsymbol{M}, t \models \varphi \text{ for all } t \text{ with } s \equiv_\Gamma t \text{ and } s < t. \qquad \blacksquare$$

Then an Γ-equality-based ceteris paribus preference $P\varphi\psi$ can be defined as follows:

$$U(\varphi \rightarrow \langle\Gamma\rangle^{\leq}\psi)$$

In practice, the sets Γ are often finite, but the system also allows infinite sets, with even recursion in the definition of the ceteris paribus formulas. For the finite case, we have:

Theorem The static logic of *CPL* is completely axiomatizable.

Proof The idea is this. All formulas in the new language have an equivalent formula in the base language, thanks to the basic laws for manipulating ceteris paribus clauses. The most important ones tell us how to change the sets Γ:

$$\langle\Gamma'\rangle^{\leq}\varphi \rightarrow \langle\Gamma\rangle^{\leq}\varphi \qquad \text{if } \Gamma \subseteq \Gamma'$$
$$(\alpha \wedge \langle\Gamma\rangle^{\leq}(\alpha \wedge \varphi) \rightarrow \langle\Gamma \cup \{\alpha\}\rangle^{\leq}\varphi$$
$$(\neg\alpha \wedge \langle\Gamma\rangle^{\leq}(\neg\alpha \wedge \varphi) \rightarrow \langle\Gamma \cup \{\alpha\}\rangle^{\leq}\varphi$$

---

[203] This is a conjunction of two normality readings: one with $N = A$, and one with $N = \neg A$.

[204] Von Wright 1963 kept one particular set $\boldsymbol{A}$ constant, viz. all proposition letters that do not occur in the φ, ψ in a statement $P\varphi\psi$. His preference logic has explicit rules expressing this feature.

[205] For more general logics of dependence, cf. van Benthem 1996, Väänänen 2007.

Applying these laws iteratively inside out will remove all ceteris paribus modalities until only cases $<\varnothing>^{\leq}$ remain, i.e., ordinary preference modalities from the base system. ■

The result is a practical calculus for reasoning with ceteris paribus propositions. [206]

9.8 Entanglement: preference, knowledge, and belief

Finally, we take up an issue that has come up at several places. We have analyzed preference per se, but often it also has epistemic or doxastic aspects, being sensitive to changes in beliefs, and subject to introspection. A standard approach would add epistemic structure to our models, and define richer preferences by combining earlier modalities for betterness, knowledge, and belief. Or should the marriage be more intimate? [207]

First degree: combining modalities Van Benthem & Liu 2007 give a logic of knowledge and preference with epistemic accessibility and betterness. Their language has modalities $<\leq>$, a universal modality, and modalities $K\varphi$. This can state things like

$KP\varphi\psi$	knowing that some generic preference holds,
$PK\varphi K\psi$	preferring to know φ over knowing ψ. [208]

The semantics allows for betterness comparisons beyond epistemically accessible worlds. [209] A language like this can change the earlier definition of lifted preferences $P\varphi\psi$ to

$$K(\varphi \rightarrow <\leq>\psi).$$

Public announcements Preference dynamics now comes in two forms. There are direct betterness changing events as before, but preference may also change through *PAL*-style informative events $!\varphi$ as in Chapter 3. This is easily combined into one system:

Theorem The combined logic of public announcement and suggestion consists of all separate principles for these operations plus two recursion axioms

---

[206] This improves on logics like von Wright's where the set Γ is left implicit in context, that have tricky features of non-monotonicity and other surprising shifts in reasoning.

[207] Cf. Liu 2008. De Jongh & Liu 2008 make belief-based preference their central notion.

[208] Intuitively, this kind of statement raises tricky issues of future *learning*, that might work better in an epistemic or doxastic temporal logic with scenarios of investigation (cf. Chapter 11).

[209] This can express a sense in which I prefer marching in the Roman Army to being an academic, even though I know that the former can never be.

that govern betterness after update and knowledge after upgrade:

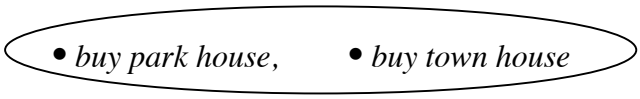
$[!\varphi]{<}{\leq}{>}\psi \leftrightarrow (\varphi \rightarrow {<}{\leq}{>}(\varphi \wedge [!\varphi]\psi))$

$[\#\varphi]K_i\psi \leftrightarrow K_i[\#\varphi]\psi$

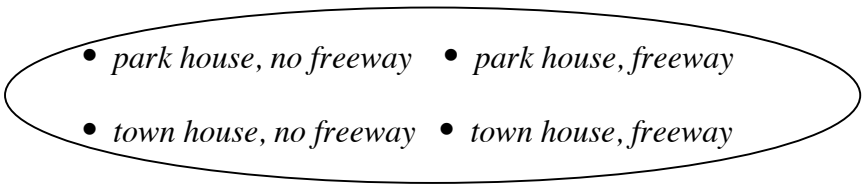
Digression: upgrade versus update The *Art of Modeling* in Chapters 3, 4 returns here in the choice of initial models for dynamics to start. Our system offers alternative models for the same scenario, trading preference change for information update (Liu 2008):

Example Buying a house.

I am indifferent between a house near the park or downtown. Now I learn that a freeway will be built near the park, and I come to prefer the town house. This may be described in our earlier terms as an initial *two-world* model



with a betterness indifference between the worlds. Taking a suggestion 'Town House' leaves both worlds, but removes a ≤-link, leaving a strictly better town house. But one can also describe the scenario in terms of a *four-world* model with extended options



with obvious betterness relations between them. An announcement of 'Freeway' now removes the two worlds to the left to get the model we got before by upgrading. [210] ■

Instead of knowledge, one can also merge the logics of belief of Chapter 7 with preference upgrade. Very similar points will apply as before, but the definable notions get more interesting. For instance, Lang & van der Torre 2008 discuss preference $P\varphi\psi$ as lifted betterness between only the *most plausible worlds* satisfying the propositions φ, ψ.

Second degree: ***intersections*** The logics so far may still miss a yet closer entanglement of preference and knowledge. An epistemic preference formula $K(\varphi \rightarrow {<}{\leq}{>}\psi)$, though subject to introspection, refers to ψ-worlds that are better than epistemically accessible φ–worlds. But there is no guarantee that these ψ–worlds *themselves* are accessible. But in our

[210] This example raises complex issues of the language in models, and pre-encoding future events.

intuitive reading of the normality sense of ceteris paribus preference, we made the comparison *inside* the normal worlds, and likewise, we may want to make it inside the epistemically accessible worlds. [211] To do this, we need a modal language that can talk about the *intersection* of the epistemic relation ~ and the betterness relation ≤:

Definition Modal preference logic with epistemic intersection.
The *epistemic-preferential intersection modality* is interpreted as follows:

$\boldsymbol{M}, s \models \langle \leq \cap \sim \rangle \varphi$ iff there is a t with $s \sim t$ & $s \leq t$ such that $\boldsymbol{M}, t \models \varphi$ ■

Now we can define internally epistemized preference, say, claiming that each epistemically accessible φ-world sees an accessible ψ-world that is at least as good:

$K(\varphi \rightarrow \langle \leq \cap \sim \rangle \psi)$

This new generic epistemic preference is no longer bisimulation-invariant (cf. Chapter 2), but it still allows for recursive analysis:

Theorem The dynamic logic of epistemic-preferential intersection is completely axiomatizable, and its key recursion axiom is the following equivalence:

$\langle \# \varphi \rangle \langle \leq \cap \sim \rangle \psi \leftrightarrow (\neg \varphi \wedge \langle \leq \cap \sim \rangle \langle \# \varphi \rangle \psi) \vee (\varphi \wedge \langle \leq \cap \sim \rangle (\varphi \wedge \langle \# \varphi \rangle \psi)$

Similar results hold with belief instead of knowledge. Dynamic events will produce both hard information and plausibility-changing soft information affecting preference.

Third degree entanglement Still more intimately, preference and belief may even be taken to be interdefinable. Some literature on decision theory (cf. the survey in Pacuit & Roy 2006) suggests that we learn a person's beliefs from her preferences as revealed by her actions – or even, that we learn preferences from beliefs (cf. Lewis 1988). In this book, entangled belief, preference, and action return in our logical study of games in Chapter 10. A rational player keeps all three aligned in a particular way, and different forms of entanglement defines different agent types that must reach equilibrium in a game.

9.9 Conclusion

Agents' preferences can be described in modal languages, just as knowledge and beliefs. These systems admit of dynamic logics that describe preference change triggered by events of accepting a command, suggestion, or more drastic changes in normative structure. But

[211] A similar entanglement of epistemic and doxastic structure is found in Baltag & Smets 2006.

preference is not just pure order of worlds, and we found interesting new phenomena that can be incorporated into logical dynamics, such as syntactic priority structure, management of ceteris paribus propositions, and entanglement with informational attitudes.

9.10 Further directions and open problems

Preference and actions Preferences hold between worlds or propositions, but also *actions*. On an action-oriented view of ethics, if helping my neighbour is better than doing nothing, I must do it, whatever the consequences. How can we incorporate action-oriented preference? A formalism might merge preference logic with *propositional dynamic logic*. Since *PDL* has formulas as properties of states and programs as interstate relations, we can put preference structure at two levels. One kind runs between states, the other between state transitions ('moves', 'events'), as in van der Meijden 1996, van Benthem 1999A.

Obligations and deontic logic This chapter obviously invites a junction with deontic logic and legal reasoning, including conditional obligations and global norms. There is a vast literature on deontic logic that we cannot reference here (cf. P. MacNamara's 2006 entry in the *Stanford On-Line Encyclopedia of Philosophy*, or Tan & van der Torre 1999 on how deontic logic enters computer science, a trend pioneered by J-J Meyer in the 1980s). Hansson 1969 is still relevant as a semantic precursor of the models used in this chapter. For a state-of-the-art study, see Boella, Pigozzi & van der Torre 2009 on the logic of obligations, norms, and how these change. Liu 2008, Grossi 2009 are attempts at creating links with dynamic-epistemic logic on topics like deontic paradoxes and norm change.

Art of modeling and thickness of worlds The example of Buying a House raised general issues of how much language to represent explicitly in the valuation of our models – and also, which possible future informational or evaluative events to encode in the description of worlds. On the whole, our dynamic logics are geared toward 'thin' worlds and light models, creating 'thickness' by further events that transform models. But there is always the option of making worlds in the initial model thicker from the start. Speaking generally, there is a *trade-off* here: the thicker the initial model, the simpler the subsequent events. I believe that every complex dynamic event, say, the *DEL*-style epistemic or doxastic updates of Chapters 4, 7, can be reduced to *PAL*-style public announcement by making worlds thicker. But I have never seen a precise technical formulation for this remodeling. Several technical facts in Chapter 10 on games and Chapter 11 on temporal trees seem relevant here, but I leave clarification of this issue to the reader.

Two-level preference logic The two-level view of object betterness plus priorities among properties, raised many questions, such as harmony in dynamic operations and merging logics. The same issues arise if we would take the two-level view back to Chapter 7, and study relational belief revision in tandem with entrenchment dynamics for propositions. Grossi 2009 shows how joint betterness/priority models for deontic logic with a matching modal language throw new light on the classical paradoxes of deontic reasoning.

Still further entanglement Entangled beliefs and preferences are the engine of decision theory (Hanson 1995, Bradley 2007). Can we add a qualitative logical counterpart to the fundamental notion of *expected value* to our logics, say based on degrees of plausibility? (A related probabilistic question occurred in Chapter 8.) Entanglement gets even richer when we add agents' *intentions*, as in Roy 2008 on the philosophy of action, or in *BDI* logics of agency in computer science (cf. Shoham & Leyton-Brown 2009).

Groups, social choice, and merge Preference logics with *groups* are a natural counterpart to epistemic and doxastic logics with groups (Chapters 2, 7, 12). Their dynamic versions describe group learning and preference formation through fact-finding and deliberation. Coalitional game theory and social choice theory (Endriss & Lang, eds., 2005) provide examples. A combined logic for actions and preferences for *coalitions* in games is given in Kurzen 2007. Group extensions of the logics in this chapter remain to be made, but see Grossi 2007 for a logical theory of institutions with normative aspects. Chapter 12 defines preference merge in structured groups, using Andréka, Ryan & Schobbens 2002.

Quantitative scoring There are also more quantitative versions of our preference logics, with *scoring rules* for worlds. Liu 2005 has models with numerical point values for worlds, and she gives a complete dynamic logic for a numerical *DEL*-style product update rule for worlds (s, e) using the separate values for s, e. Quantitative preference dynamics might be developed for most themes in this chapter, like we did in Chapter 8 for probability.

Dependence logic As we noted, ceteris paribus preference is related to general *logics of dependence* (van Benthem 1996, Väänänen 2007). Given the importance of the notion of dependence in games (Nash equilibrium involves the equality sense of ceteris paribus: cf. van Benthem, Girard & Roy 2009; players' behaviour depends on that of other players in extensive games) and many other disciplines, these links are worth pursuing.

9.11 Literature

The literature on preference logic is surveyed in the Handbook chapter Hanson 2001. Close to this chapter, Van Benthem, van Eijck & Frolova 1993 proposed a dynamic logic for changing preferences triggered by actions including our 'suggestions'. Boutilier & Goldszmidt 1993 gave a semantics for conditionals in terms of actions that minimally change a given order so as to make all best antecedent worlds ones where the consequent is true. This upgrade idea was taken much further in Veltman 1996 on dynamic semantics of default reasoning, and van der Torre & Tan 1999, 2001 on deontic reasoning and changing obligations. Zarnic 1999 studied practical reasoning with actions *FIAT* φ as changes in an ordering making the φ–worlds best. Yamada 2006 analyzed acceptance of deontic commands as relation changers, and gave the first complete dynamic-epistemic logics in our style. This chapter is based largely on Liu 2008, Girard 2008, van Benthem & Liu 2007, van Benthem, Girard & Roy 2009, whose themes have been explained in the text.