FOR BETTER OR FOR WORSE: DYNAMIC LOGICS OF PREFERENCE

Johan van Benthem, Amsterdam & Stanford, http://staff.science.uva.nl/~johan/

February 2008

Abstract

In the last few years, preference logic and in particular, the dynamic logic of preference change, has suddenly become a live topic in my Amsterdam and Stanford environments. At the request of the editors, this article explains how this interest came about, and what is happening. I mainly present a story around some recent dissertations and supporting papers, which are found in the references. There is no pretense at complete coverage of preference logic (for that, see Hanson 2001) or even of preference change (Hanson 1995).

1 Logical dynamics of agency

Agency, information, and preference Human agents acquire and transform information in different ways: they observe, or infer by themselves, and often also, they ask someone else. Traditional philosophical logics describe part of this behaviour, the 'static' properties produced by such actions: in particular, agents' knowledge and belief at some given moment. But rational human activity is goal-driven, and hence we also need to describe agents' evaluation of different states of the world, or of outcomes of their actions. Here is where preference logic have come to describe what agents prefer, while current dynamic logics describe effects of their physical actions. In the limit, all these things have to come together in understanding even such a simple scenario as a *game*, where we need to look at what players want, what they can observe and guess, and which moves and long-term strategies are available to them in order to achieve their goals.

Logical dynamics of information and belief There are two dual aspects to this situation. The static description of what agents know, believe, or prefer at any given moment has long been performed by standard systems of philosophical logic since the 1950s – of course, with continued debate surrounding the merits of particular proposals. But there is also the dynamics of actions and events that produce information and generate attitudes for agents – and gradually, these, too, have been made a subject of logical investigation in the program of 'logical dynamics' (van Benthem 1996, van Benthem 2008). For instance, an observation or an answer to a question are informative events that can be put explicitly inside complete systems of dynamic logic, which describe what agents know before and after such events take place. For purposes of exposition, this paper will use the current methodology of 'dynamic epistemic logic' (cf. van Ditmarsch, van der Hoek & Kooi 2007, Baltag, van Ditmarsch, and Moss 2008), and some concrete systems will be found below. A typical formula of such a system might say the following:

$[!\varphi]K_i\psi$ after receiving the 'hard information' that φ, agent i knows that ψ.

This describes knowledge of individual agents after direct *information update*, and the account can also deal with complex group scenarios where agents have different observational access to the actual event taking place (think of drawing a card in a game). By now, there are also dynamic logics that describe more subtle 'policy-driven' activities, such as absolute or conditional beliefs agents get after an event takes place that triggers a *belief revision* (van Benthem 2007A, Baltag & Smets 2006), with formulas like:

$[\Uparrow\varphi]B_i\psi$ after receiving 'soft information' that φ, agent i believes that ψ.

Preference change, and beyond Once on this road, since rational action is about choosing on the basis of information and preference, it was only a matter of time before dynamic *preference change* and its triggering events became a topic of investigation. This paper will report on some of this. And logical dynamics does not even stop here. In principle, any static aspect of agency or language use studied in the logical tradition can be 'dynamified', including shifts in temporal perspective, group standing, etc. (cf. van Benthem, Muskens, Visser 1997). One issue that arises here is how all these separate dynamifications hang together. Can we really just look at events that produce knowledge, belief, or preference separately, and put them together compositionally? Or is there some deeper conceptual entanglement between these notions calling for more delicate formal constructions? All these issues will be discussed for the case of preference below.

Overview This paper is mainly based on some recent publications in the Amsterdam environment over the last three years. Indeed, 'dynamics' presupposes an account of 'statics', and hence we first give a brief survey of preference logic in a simple modal format using binary comparison relations between possible worlds – on the principle that 'small is beautiful'. We also describe a recent alternative approach, where world preferences are generated from criteria or constraints. We show how to dynamify both views by adding explicit events that trigger preference change in the models, and we sketch how the resulting systems connect. Next, we discuss some entanglements between preference, knowledge and belief, and what this means for combined dynamic logics. On top of this, we also show how more delicate aspects of preference should be incorporated, such as its striking 'ceteris paribus' character, which was already central in Von Wright 1963. Finally, we relate our considerations to social choice theory and game theory.

2 Modal logic of betterness

Preference is a very multi-faceted notion: we can prefer one individual object, or one situation, over another – but preference can also be directed toward kinds of objects or generic types of situation, often defined by propositions. Both perspectives make sense, and a bona fide 'preference logic' should do justice to all of them eventually. We start with a simple scenario on the object/world side, leaving other options for later.

Basic models In this paper, we start with a very simple setting. *Modal models* $\boldsymbol{M} = (W, \leq, V)$ consist of a set of worlds W (but they really stand for any sort of objects that are subject to evaluation and comparison), a *'betterness' relation* $\leq$ between worlds ('at least as good as'), and a valuation V for proposition letters at worlds (or, for unary properties of objects). In principle, the comparison relation may be different for different agents, but in what follows, we will suppress agent subscripts $\leq_i$ whenever possible for greater readability. Also, we use the artificial term 'betterness' to stress that this is an abstract comparison relation, making no claim yet concerning the natural rendering of the intuitive term 'preference', about which some people hold passionate proprietary opinions. Still, this semantics is entirely natural and concrete. Just think of decision theory, where worlds (standing for outcomes of actions) are compared as to utility, or game theory, where end nodes of a game tree (standing for different final histories of the game) are related by preference relations for the different players. In other words, our simple modal models represent a widespread use of the term 'preference' in science.

Digression: plausibility Very similar models have been widely used to model another notion, viz. 'relative plausibility' as judged by an agent. This happens in the semantics of belief and doxastic conditionals, where beliefs are those propositions which are true in all most plausible relevant worlds – and further plausibility models are also crucial to the best-known semantics for belief revision. While preference is not the same as plausibility (except for very wishful thinkers), this formal analogy has proven quite helpful as a source of ideas and transfer of results across the two fields. [1] We will return to the issue of more genuine conceptual 'entanglements' between preference and belief later on.

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. In particular, a modal assertion like $<\leq>\varphi$ will make the following 'local' assertion at a world w:

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

i.e., there is a world v at least as good as w which satisfies φ. In combination with other operators, this simple formalism can express many natural notions concerning rational preference-driven action. For instance, consider finite game trees, which are natural models for a dynamic logic of atomic actions (players' moves) and unary predicates indicating players' turns at intermediate nodes and their utility values at end nodes (van Benthem 2002). Van Benthem, van Otterloo & Roy 2006 show how the *backward induction solution* of a finite game [2] may be defined as the unique binary relation *bi* on the game tree satisfying the following modal preference-action law:

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

---

[1] Cf. the analysis of non-monotonic logic via abstract world preference in Shoham 1988.

[2] A famous 'benchmark example' in the logical analysis of games; cf. Harrenstein 2004. Apt & Zvesper 2008 give a logical take on rationality in solution procedures for strategic form games.

Here *move* is the union of all one-step move relations available to players, and * denotes the reflexive-transitive closure of a relation. The formula then says there is no alternative move to the *BI*-prescription at the current node all of whose outcomes would be better than the *BI*-solution. Thus, modal preference logic seems to go well with games. [3]

But there are more examples. Already Boutilier 1994 observed how such a simple modal language can also define conditional assertions, normally studied per se as a complex new binary modality (Lewis 1973), and how one can then analyze their logic in standard terms. [4] For instance, in modal models with finite pre-orders (see below), the standard truth definition of a conditional $A \Rightarrow B$ reads as 'B is true in all maximal A-worlds' – and this clause can be written as the following modal combination:

$$[](A \to \langle\leq\rangle(A \,\&\, [\leq](A \to B))), \qquad \text{with } [] \text{ some appropriate universal modality.}$$

While this formula may look complex at first, the point is that the inferential behaviour of the conditional, including its well-known non-monotonic features, can now be completely understood via the base logic for the unary modalities, say, as a sub-theory of modal *S4*. Moreover, the modal language easily defines variant notions whose introduction seems a big deal in conditional logic, such as existential versions saying that each A-world sees *at least one* maximal A-world which is B. Of course, explicit separate axiomatizations of these defined notions retain an independent interest: but we now see the whole picture. [5]

Constraints on betterness orders Which properties should a betterness relation have? Many authors like to work with *total orders*, satisfying reflexivity, transitivity, and connectedness. This is also common practice in decision theory and game theory, since

[3] This, and also the following examples are somewhat remarkable, because there has been a widespread prejudice that modal logic is not very suitable to formalizing preference reasoning.

[4] This innovative move is yet to become common knowledge in the logical literature.

[5] There still remains the question of axiomatizing such defined notions per se: and that may be seen as the point of the usual completeness theorems in conditional logic. Also, Halpern 1997 axiomatized a defined notion of preference of this existential sort.

these properties are enforced by the desired numerical representation of agents' utilities. But if we look at the logical literature on preference or plausibility, things are less clear, and properties have been under debate ever since the pioneering study Halldén 1957. E.g., transitivity has been extensively criticized as a constraint on intuitive preference (Hanson 2001). And in conditional logic, Lewis' use of totality is often abandoned in favour of just *pre-orders*, satisfying just the conditions of reflexivity and transitivity, while acknowledging *four* intuitively irreducible basic relations between worlds:

$w \leq v$, $\neg\, v \leq w$ (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.

We feel this pleads for having a large class of models, noting the extra modal principles enforced through frame correspondence *if* we make the relation satisfy extra constraints. [6] The point of a logical analysis is to impose structure where needed, but also, to identify the 'degrees of freedom' where parameters are to be set in an intuitive notion.

Further relations? Finally, we note that there may be a case for having two independent betterness relations in models: a weak order $w \leq v$ for 'at least as good', and a strict order $w < v$ for 'better', defined as $w \leq v \;\&\; \neg\, v \leq w$. Van Benthem, Girard & Roy 2007 axiomatize the logic of this extended language, with two separate modalities for the weak and strict betterness relations, which has some elegant principles about their interplay.

For more on the austere modal framework of this section and its unifying power, cf. the dissertation Girard 2008, who shows, drawing upon much more relevant literature than

[6] Some people feel a relation 'is' only a preference relation when we impose constraints like transitivity. But this seems a category mistake. A formal relation in a model is just a mathematical object, though it may come to *stand for* a preference in a context of modeling, which requires some scenario attaching the formal model to some reality being described. Moreover, given several decades of research on preference relations, it seems highly unlikely that there is any stable base se of constraints: preference might be more of a 'family notion'.

we have discussed here, that our basic 'order logic' is a wide-ranging pilot environment for studying essential patterns in reasoning with preference and belief. [7]

3 Defining global propositional preference

As we have said, a betterness relation need not yet determine what we mean by agents' preferences in some more colloquial sense. Indeed, many authors consider 'preference' really a relation between propositions, with von Wright 1963 as a famous example. These differences seem largely terminological, which is precisely why debates are often bitter. [8]

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 relation lifting, also elsewhere, is the $\forall\exists$ stipulation that

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

As we said, this was axiomatized by Halpern 1997. But alternatives are possible. Van Benthem, Girard & Roy 2008 analyze von Wright's own 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. And still further combinations occur. Liu 2008 provides a brief history of further proposals for relation lifting in various fields (decision theory, philosophy, computer science), but no consensus on one canonical notion of preference seems to have ever emerged. This may be a feature, rather than a bug. Preference as a comparison relation between propositions may turn out different depending on the scenario. For instance, in a *game*, when comparing sets of outcomes that can be reached by selecting available moves, players may have different options. One would indeed say that we prefer a set whose minimum utility value exceeds the maximum of another (this

---

[7] We have not even exhausted all approaches cooking in Amsterdam right now. For another kind of modal preference logic in games, including a 'normality' operator, see Apt & Zvesper 2007.

[8] Compare William James' famous squirrel going 'round' the tree (or not…): cf. James 1907.

is like the $\forall\forall$ reading) – but it would also be quite reasonable to say that the maximum of one set exceeds the maximum of the other, which would be rather like the $\forall\exists$ reading.

Extended modal logics The main insight from the current modal literature on preference is two-fold. First, many different liftings are definable in our modal base logic extended with a universal modality $U\varphi$: 'φ is true in all worlds'. This standard feature from 'hybrid logic' 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:

$$U(\varphi \rightarrow \langle\leq\rangle\psi).$$

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

Of course, eventually, one can also use stronger formalisms for describing preferences, such as first-order logic (cf. Suppes 1957), but this is just the ordinary balance in logic between finding illuminating formalizations of key notions and argument patterns, and the quest for formalisms combining optimal expressivity with computational ease. [9] We have nothing against richer languages, but modal logic is an attractive first level to start.

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 can happen which make us *change* these evaluations of worlds. This dynamic idea has been in the air for quite while now. [10] In particular, van Benthem, van Eijck & Frolova 1993 already proposed a first system for 'changing preferences', as triggered by various actions that can be defined in a dynamic logic. One example was the 'upgrade event' $\#(A)$ which removes all betterness arrows running from A-worlds to $\neg A$-worlds

---

[9] For this Balance between expressive power and computational complexity, cf. the chapter by van Benthem & Blackburn in the *Handbook of Modal Logic*, Elsevier, Amsterdam, 2007.

[10] We only review one strand here: cf. again Hanson 1995 for a different point of entry.

from the current model. In the same period, Boutilier & Goldszmidt 1993 described a dynamic semantics for conditionals $A \Rightarrow B$, in terms of actions which produce a minimal change in a given world comparison relation so as to *make* all 'best' A-worlds in the new pattern B-worlds. This idea was developed much more systematically in Veltman 1996 on the logical dynamics of default reasoning [11], and subsequent publications such as Tan & van der Torre 1999 on deontic reasoning and the dynamics of changing obligations that lies behind it. In particular, the systems to be discussed in this paper may be traced back to Zarnic 1999 on practical reasoning, which analyzed actions '*FIAT* φ' for factual assertions φ as changes in a comparison relation making the φ–worlds 'best'. Next, again in deontic logic, Yamada 2006 proposed analyzing acceptance of 'commands' as relation changers, and provided some complete logics in the dynamic-epistemic style.

Of course, realistic preference change has many more features than those mentioned here, which will come to light on a deeper analysis of agents (cf. Lang & van der Torre 2008). Moreover, various formal proposals already exist (cf. Hanson 1995). But in the remainder of this paper, we concentrate merely on how our basic ideas work, when pursued in a systematic logical methodology of the sort found in the Dutch Lowlands.

5 A basic dynamic preference logic

How does a dynamic logic of preference change work? We present some basic features from van Benthem & Liu 2007, starting with about the simplest scenario.

Dynamic logic of 'suggestions' Betterness models will be as before, and so is the modal base language, with modalities $<\leq>$ and U. But the syntax now adds a feature, borrowed

[11] Veltman insists that the *meaning* of conditionals has this dynamic character, making logical formulas 'implicitly dynamic'. Most work that we are reporting on has 'explicit dynamics', and assumes the traditional static meanings for logical formulas, while using these in explicit triggers for dynamic actions which change models. In other words, one can do 'logical dynamics' without committing to 'update semantics' – and vice versa.

from dynamic logic of programs in computer science. For each formula of the language, we add a model-changing action $\#(\varphi)$ of 'suggestion' [12], defined as follows:

For each model $\boldsymbol{M}, w$, the model $\boldsymbol{M}\#\varphi, w$ is $\boldsymbol{M}, w$ with the
new relation $\leq' = \leq - \{(x, y) \mid \boldsymbol{M}, x \models \varphi \;\&\; \boldsymbol{M}, y \models \neg\varphi\}$.

Note that this model change event is a function, providing unique values for each $\boldsymbol{M}, w$.

Next, we enrich the formal language by adding action modalities interpreted as follows: [13]

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

These allow us to 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 dynamic analysis consists in finding the 'recursion equation' explaining when a preference obtains after an action, in so far as the language can express it. Here is the relevant valid principle for suggestions, whose two cases can be seen to follow the above definition of the above model change:

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

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:

$$[\#(\varphi)]\, p \;\leftrightarrow\; p$$
$$[\#(\varphi)]\, \neg\psi \;\leftrightarrow\; \neg[\#(\varphi)]\psi$$
$$[\#(\varphi)](\psi\&\chi) \;\leftrightarrow\; [\#(\varphi)]\psi \;\&\; [\#(\varphi)]\chi$$
$$[\#(\varphi)]U\psi \;\leftrightarrow\; U[\#(\varphi)]\psi$$
$$[\#(\varphi)]<\leq>\psi \;\leftrightarrow\; (\neg\varphi \;\&\; <\leq>[\#(\varphi)]\psi) \vee ((\varphi \;\&\; <\leq>(\varphi \;\&\; [\#(\varphi)]\psi).$$

---

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

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

Proof These axioms express the following semantic facts, respectively: upgrade does not change atomic facts, upgrade is a function, upgrade is a normal modality, upgrade does not change the domain of worlds of the model, and upgrade follows the definition of suggestion as explained earlier. It is easy to see that, when applied inside out, these laws can reduce any valid formula to an equivalent one not containing any dynamic modalities, for which the given base logic is already complete by assumption. [14] ♣

This logic automatically gives us a dynamic logic of upgraded propositional preferences. For instance, we can compute as follows how $\forall\exists$–type preferences $P\psi\chi$ arise:

$$[\#(\varphi)]P\psi\chi \leftrightarrow [\#(\varphi)]U(\psi \rightarrow <\leq>\chi) \leftrightarrow$$
$$U[\#(\varphi)](\psi \rightarrow <\leq>\chi) \leftrightarrow U([\#(\varphi)]\psi \rightarrow [\#(\varphi)]<\leq>\chi) \leftrightarrow$$
$$U([\#(\varphi)]\psi \rightarrow (\neg\varphi \,\&\, <\leq>[\#(\varphi)]\chi) \vee ((\varphi \,\&\, <\leq>(\varphi \,\&\, [\#(\varphi)]\chi) \leftrightarrow$$
$$P([\#(\varphi)]\psi \,\&\, \neg\varphi)[\#(\varphi)]\chi \;\;\&\;\; P([\#(\varphi)]\psi \,\&\, \varphi)(\varphi \;\&\; [\#(\varphi)]\chi).$$

General relation transformers But this is still just a 'trial run' for one particular kind of preference change. Van Benthem & Liu 2007 also consider other relation transformers.

For instance, let $\Uparrow(\varphi)$ be the relation change which makes all φ–worlds better than all $\neg\varphi$-worlds, while keeping the old order inside these zones. In preference terms, this makes φ the 'most desirable good', while in terms of belief revision (van Benthem 2007), it is a piece of 'soft information' making the φ–worlds the most plausible ones – though still leaving a loop hole for $\neg\varphi$ perhaps being true. Again, we can find a complete recursion axiom for this notion, this time as follows, using an 'existential modality' E: [15]

[14] This reductive analysis shows that the process of preference can be analyzed compositionally. Moreover, it shows that the base language was well-designed, in 'expressive harmony' with the dynamic superstructure. Even so, the real dynamic account of *preference change* is of course in *the recursive procedure itself*, and it lies only hidden implicitly in the base language.

[15] Van Benthem 2007A uses this axiom to analyze agents' *conditional beliefs* after receiving soft information, with a recursion based on the definition of such beliefs in our modal base language.

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

But in principle, there can be many further triggers for betterness change, depending on how people adjust to what others claim, command, etc. Thus, it is hard to specify just a small set of changes, with logic serving as an arbiter of how one should respond to them. The task of a dynamic logic of preference is rather providing the appropriate generality, and spotting where some 'trigger' needs to be provided as input to the update. [16]

Here is one way of achieving parametrization of preference change. The new betterness relations in our examples are *definable* from the old ones in the following straightforward syntactic '*PDL* program format', involving *test*, *sequential composition* and *union*:

$$\#(\varphi)(R) \quad = \quad (?\varphi \,;\, R \,;\, ?\varphi) \cup (?\neg\varphi \,;\, R \,;\, ?\neg\varphi) \cup (?\neg\varphi \,;\, R \,;\, ?\varphi)$$
$$\Uparrow(\varphi)(R) \quad = \quad (?\varphi \,;\, R \,;\, ?\varphi) \cup (?\neg\varphi \,;\, R \,;\, ?\neg\varphi) \cup (?\neg\varphi \,;\, T \,;\, ?\varphi)$$

where *'T'* is the universal relation in the model.

Note that the former definition can only go to a sub-relation of the current one, while the second may add new links as well. Both types fall under the following result:

Theorem Any relation transformer τ with a program definition in the *PDL* format has a complete reduction axiom which can be computed effectively from τ's definition.

The proof is a simple recursive recipe, viewing the definitions basically as 'substitutions' of new relations for old. There are also other ways of achieving generality, e.g., in terms of 'event models' (see Section 10 below), but the program method, too, is powerful. [17]

Constraints on betterness ordering once more While adding a dynamic superstructure to an existing modal logic seems a somewhat 'conservative' enterprise of mere addition,

---

[16] Many people have the mistaken belief that this 'plurality' is reprehensible wantonness, whereas localizing the proper *degrees of freedom* for an agent is a precisely a key task for logical analysis.

[17] Van Eijck's commentary in Apt & van Rooij, eds., 2008 uses this technique for belief revision, linking up with 'factual change' in *DEL* as treated in van Benthem, van Eijck & Kooi 2006.

there are several points where matters can be more interesting. One is that, if a static base language is to have enough power for 'pre-encoding' the effects of dynamic changes, it must have the right expressiveness. A good example are the static conditional beliefs needed to pre-encode effects of belief revision, or the 'conditional common knowledge' of van Benthem, van Eijck & Kooi 2006 needed for pre-encoding group knowledge that arises after public announcement. Indeed, some basic logical notions seem to have just this 'forward-looking' character. Such issues of language design may be relevant to preference logic once we study group preferences, but we have not encountered them yet.

But another issue has been noted in van Benthem & Liu 2007. Suppose that our current betterness order satisfies some relational constraints, what guarantees that its transformed version will still satisfy these same constraints? For instance, it is easy to see that the above suggestions take pre-orders to pre-orders, but they can destroy the *totality* of a betterness order. Liu 2008, Chapter 4, analyzes this further, but we have no general results yet. There is an interesting debate here. Some people see this potential loss of basic order properties as a basic drawback of the relation transformer approach. But we feel that the situation is exactly 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 the case vs. what should become the case It is tempting to read instructions like $\Uparrow(\varphi)$ as '*see to it that* you come to prefer, or believe, that φ'. This is a forward-oriented view of dynamics: one should make some minimal change resulting in the truth of some stated 'postcondition'. But this is not really the spirit of dynamic-epistemic logic, which rather lets events tell us the 'preconditions' of their occurrence. The two views clash, e.g., in deontic logic, when a command says that you must make sure some proposition becomes true without telling you how. In principle, our approach is 'constructive': triggers in the logic must tell us exactly how the model is to be changed. For the other view, temporal logics (Belnap et al. 2001, van Benthem & Pacuit 2006) may be the better format, where the model already gives the possible future histories.

6 Alternative: constraint-based preference

So far, we have followed the beaten modal path, starting from an ordering of worlds, and deriving notions of preference that apply to propositions, definable in our languages. But there is also another approach to preference, conceptually equally attractive, which works from the opposite direction. Object comparisons are often made on the basis of *criteria*, and then derived from the way in which we apply these criteria, and prioritize between them. For instance, cars may be compared as to price, safety, and comfort, in some order of importance. In that case, the criteria are primary, and the object or world order is derived. This framework, too, occurs in many scientific settings, including philosophy and economics, with various connections made between the two fields in Rott 2001. Another example of its descriptive power is 'Optimality Theory' in linguistics and general cognitive science (Prince and Smolensky 1993, Smolensky 2006). [18]

First-order priority logic A recent logical formalization of this approach to preference was given in de Jongh & Liu 2007. In the simplest case, one starts from a finite sequence $\boldsymbol{P}$ of propositions, or properties, and then orders objects as follows:

> $x < y$ 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 method of lexicographic ordering, if we view each property $P \in \boldsymbol{P}$ as inducing the following simple object order: [19]

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

De Jongh and Liu give a first-order toy language for describing these induced preferences between objects. They also prove a representation result for object or world models:

---

[18] By the way, note that a priority order among propositions need not be a preference relation. I do not 'prefer' safety of my vehicle to sleek design, I just consider it more essential.

[19] 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.

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 discusses this situation further, and notes that the literature has many other ways of defining object order from property orders, which can be studied in similar ways. This diversity may be compared with that for 'lifting' object order to world order before.

Dynamics Again, this style of analysis suggests an obvious engine for preference change. This time, it is the priority order and set of relevant properties which can change, thereby inducing a changing in the defined object order. A new criterion may become relevant, or a criterion may lose its former importance. De Jongh & Liu study four main operations: *permuting* properties in a priority sequence, *prefixing* a new property, *postfixing* a new property, and *inserting* a property at some specified position. Together, these allow for any manipulation of finite sequences. Moreover, they lead to complete dynamic logics for the changed derived object-level preferences after such changes have taken place first at the level of the prioritized properties. The format is borrowed from the earlier modal one, and therefore, we do not repeat the precise results here. What all this does show is that the style of dynamification in earlier sections also works for first-order logics, making our modal setting a convenience, rather than a straightjacket.

One interesting thing is that the priority dynamics has its own intuitions, different from the account of 'suggestions' or 'commands' we had before. For instance, Girard 2008 re-interprets it as a sort of *agenda* for investigation, determining what is more important than what. He then links the dynamics of 'agenda change' to issues in the philosophy of science, where 'research programs' serve as agendas that keep changing over time.

Two-level connections The two approaches so far may be viewed as complementary. [20] One either starts from a primitive betterness relation between worlds and then lifts this to

---

[20] There are obvious connections here with the duality in belief revision between working with a basic world order, or a primitive *'entrenchment order'* of propositions; cf. the excellent survey in Gärdenfors & Rott 1995; but we do not pursue this analogy here.

obtain propositional preference orders, or one starts from a primitive ‘importance order’ of propositions, and then derives world order. It is of obvious interest to compare, and perhaps combine the two perspectives, and Liu 2008 has an extensive discussion. To do so, she considers *two-level structures* $(W, \leq, \boldsymbol{P}, <)$ having both worlds with a betterness order $\leq$ and a set of ‘important propositions’ with a primitive priority order $<$:

$$\begin{tikzcd} \boldsymbol{P}, < \arrow[d, bend left, "\text{deriving}"] \\ W, \leq \arrow[u, bend left, "\text{lifting}"] \end{tikzcd}$$

This picture immediately suggests a number of questions, many of them still unresolved. E.g., structurally, what happens when we derive a betterness order form a priority order, and then lift it again? And what happens vice versa? [21] And in terms of languages, what happens when we treat the propositions in $\boldsymbol{P}$ as distinguished propositional constants in a modal language, and try to relate modal betterness logic with modal constraint logic? We have no general answers here, but at least, Liu 2008 does state elegant correspondences between the dynamics at the two levels. In particular, she shows that prefixing of 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*, the following identity holds, making the following diagram commute:

$$\begin{tikzcd} \boldsymbol{P}, < \arrow[r] \arrow[d, dash, "lex"'] & \varphi\, ;\boldsymbol{P}, < \arrow[d, dash, "lex"] \\ W, \leq \arrow[r] & W,\ \Uparrow(\varphi)(\leq) \end{tikzcd}$$

$$lex(\varphi\, ; \boldsymbol{P}) = \Uparrow(\varphi)\ (lex(\boldsymbol{P}))$$

Again, the general theory of inducing dynamics from one level to another seems open. There also seems to be room here for a more general calculus of natural operations on priority sequences, called ‘agenda algebra’ in the dissertation Girard 2008. [22]

---

[21] Nevertheless, as we said before, a priority order is not necessarily a preference order.

[22] For instance, for each set of properties, there is a set of *disjoint* properties generating the

7 Further aspects of preference: ceteris paribus logic

All our logics so far, whether betterness- or priority-based, described pure preferences. But in reality, preferences usually have a defeasible character: they hold only *ceteris paribus*, in von Wright's terminology. Van Benthem, Girard & Roy 2008 discuss this feature, and describe what needs to change in our modal approach to accommodate this in order to become a more realistic account of reasoning with preferences.

Normality versus equality First, the term 'ceteris paribus', though widely used, has no unambiguous meaning. In fact, one can distinguish two main views. In many scenarios, the *normality sense* says that we only make the preference 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 our current model. These worlds are singled out, either by some explicit description N, or just as the most plausible worlds in some doxastic plausibility order. In the former scenario, our earlier logic still suffices. We could express a global preference $P\varphi\psi$ in this normality sense as

$$P(N\&\varphi)(N\&\psi).$$

But this approach by explicit definition of normal worlds will not work in general, and then we must use models with both betterness and plausibility orders, as in Lang, van der Torre & Weydert 2003, with some matching combined logic of preference and belief. We will return to this issue of what may be called 'entanglement' in the next Section.

For now, we note that there is also another *equality sense* of 'ceteris paribus': indeed, the one favoured by von Wright. In this sense, a preference statement is made globally, though under the proviso that certain propositions do not change their truth values. For instance, someone who generally prefers work over vacation, might still be said to prefer night over day with work/vacation 'frozen' in a 'ceteris paribus', even though there are

---

same object order. Finding the latter effectively is a matter of merging Boolean normal form principles with some preference logic. A few first principles are found in the cited references.

vacation days that she would prefer to work nights. More precisely, for von Wright, a ceteris paribus preference for φ over ψ with respect to some proposition A means that

> both (i) among the A–worlds I prefer φ over ψ,
> and (ii) among the $\neg A$-worlds I prefer φ over ψ.

Thus, cross-comparisons between the A and $\neg A$ worlds are irrelevant to the truth of the preference. [23] For the case of more 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'. Von Wright himself proposed a particular set of relevant propositions $\boldsymbol{A}$ to be kept 'constant', viz. all the proposition letters of the language that do not occur in the two formulas φ, ψ being compared in a preference statement $P\varphi\psi$. His preference logic has explicit rules of reasoning expressing this feature (von Wright 1963).

This scenario is interesting because the same relation $\equiv_{\boldsymbol{A}}$ has been studied elsewhere as an account of the intuitive notions of 'dependence' and 'independence' among propositions (Doyle & Wellman 1994). It also occurs in the semantics of questions and answers in natural language (ten Cate & Shan 2002), and in treatments of supervenience and dependence in philosophy. Thus there is some logical interest to formalizing this. [24]

Equality-based ceteris paribus preference logic Van Benthem, Girard & Roy 2008 make equality-based ceteris paribus preferences an explicit part of the language, making reasoners specify explicitly which propositions are to be 'frozen' in their comparisons. They give a modal logic *CPL* extending basic preference logic with operators

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

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

---

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

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

$U(\varphi \rightarrow <\Gamma>^{\leq}\psi)$

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 We do not specify all axioms, but 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 riders. The most important one of these tell us how to change the sets Γ:

$<\Gamma'>^{\leq}\varphi \rightarrow <\Gamma>^{\leq}\varphi$ *if* $\Gamma \subseteq \Gamma'$

$((\neg)\alpha \; \& <\Gamma>^{\leq}((\neg)\alpha \; \& \; \varphi) \rightarrow <\Gamma \cup \{\alpha\}>^{\leq}\varphi$

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 main contribution here is an explicit calculus for reasoning with ceteris paribus propositions. This improves over Von Wright, where the set Γ is implicit in the context, with some tricky features. For instance, Von Wright's account of preference reasoning has no *monotonicity* in the sense that $P\varphi\psi$ implies $P(\varphi\&\alpha)\psi$, even though this inference seems plausible. The reason is that the extended formula $\varphi\&\alpha$ changes the set of relevant ceteris paribus propositions insidiously, a phenomenon explicit in the indexed modalities of the logic *CPL*, which wears the true monotonicity properties upon its sleeves.

Further developments The *CPL* axioms for changing ceteris paribus sets suggest an underlying dynamic process of context change, or in earlier terms, 'agenda change'. Van Benthem, Girard & Roy 2008 also give a *dynamic logic version* of the system, where the 'agenda' is an independent item, which can be extended or simplified – though not all natural operations admit of *DEL*-style recursion axioms. Another source of open problems is the full infinitary version of the system, which is still bisimulation-invariant, but sits somewhere in the landscape of *infinitary modal logics* at some distance from propositional dynamic logic, or other well-behaved calculi. Finally, the connection with *logics of dependence* is intriguing, but not yet understood. For instance, dependence patterns occur typically also in preference reasoning in game theory, our initial example. The authors show that Nash Equilibrium can be defined in their logic, but for this, they use only their local modality looking at worlds (i.e., strategy profiles in the game setting) having the same strategies for the other players as the current world (profile). [25] This seems more like the normality sense of ceteris paribus. The more sweeping equality sense would look at all equivalence classes arising from fixing any strategy profile for the other players, thus moving closer to game-theoretic notions like ‘strictly dominated strategies’.

8 Entanglement: preference, knowledge, and belief

Now we get to an issue which tends to generate heat among academics. So far, we have analyzed preference per se, as a mere matter of betterness comparison across worlds. But to many people, preference is a deeply *epistemic* or *doxastic* notion, manipulable by changes in beliefs, and subject to introspection. Can we do justice to such intuitions? The standard ‘piecewise’ approach here would be to add epistemic or doxastic structure to our models, and then define ‘real preference’ in terms of operator combinations with the earlier modalities for betterness as well as knowledge and belief. Or should the marriage be more intimate? We discuss these issues briefly, following Liu 2008, Chapter 4. It should be noted that these issues come up in different settings, and, e.g., de Jongh & Liu 2008 make belief-based preference their central notion, providing a complete first-order-style axiomatization. In what follows, we explain the same issues in a modal setting.

First degree of entanglement: combine separate operators Van Benthem & Liu 2007 present a combined system with both knowledge and preference, whose models have both epistemic accessibility relations and a preference order. Their formal language has both betterness modalities <≤> as before, the auxiliary universal modality, and epistemic

---

[25] As we have said before, there is a flourishing literature on logics providing definitions for basic game-theoretic notions, so it is the ceteris paribus aspect that is of interest here.

knowledge modalities $K\varphi$ interpreted as usual as truth of φ in all epistemically accessible worlds. This language can interpret delicate nested operator combinations such as

$KP\varphi\psi$ knowing that some global betterness relationship holds,

$PK\varphi K\psi$ preferring to know certain things over others. [26]

The semantics typically allows for comparisons beyond epistemically accessible worlds, however. This gives it the option of expressing a sense of 'regret' in which I prefer marching in the Roman Army to being a peaceful academic, even though I know that, alas, the former alternative cannot be. Of course, more realistic (and less romantic) agents will not use this facility provided by the system, and the logic does not force them to.

A language like this can improve on the earlier definition of global preferences $P\varphi\psi$, now reading the earlier $U(\varphi \rightarrow \langle\leq\rangle\psi)$ with a universal modality in epistemic terms:

$$K(\varphi \rightarrow \langle\leq\rangle\psi).$$

Public announcement logic Next, the dynamics in the system will have two forms. There are the betterness changing events we described before, but there are also purely informative events like a public announcement or public observation $!\varphi$ of some φ true right now in the actual world. These are the simplest forms of learning some new piece of 'hard information'. They are treated in the standard format of dynamic-epistemic logic, as a restriction of the current model $\boldsymbol{M}, s$ to its sub-model $\boldsymbol{M}|\varphi, s$ consisting of the worlds satisfying φ in $\boldsymbol{M}$. Again, we extend the language with modalities, this time as follows:

$$\boldsymbol{M}, s \models [!\varphi]\psi \quad \text{iff} \quad \textit{if } \boldsymbol{M}, s \models \varphi, \textit{ then } \boldsymbol{M}|\varphi, s \models \psi$$

Here the condition is needed because of the precondition that the new information is true, and hence update is only a partial function. The key recursion principle of the resulting

[26] This combination raises some tricky issues of intuitive interpretation, which might work better in an epistemic or doxastic temporal logic that can deal with scenarios of investigation.

public announcement logic (cf. Gerbrandy 1999, van Benthem 2006) is the following law, which describes which knowledge arises from receiving hard information:

$$[!\varphi]K_i\psi \;\leftrightarrow\; (\varphi \rightarrow K_i(\varphi \rightarrow [!\varphi]\psi))$$

This structure is easily combined with the earlier dynamic logics of preference change. For instance, as a special case we have

Theorem The combined logic of public announcement and suggestion consists of all separate principles for these operations plus two 'cross-comparisons' describing betterness after update and knowledge after upgrade:

$$[!\varphi]\langle\leq\rangle\psi \;\leftrightarrow\; (\varphi \rightarrow \langle\leq\rangle(\varphi \;\&\; [!\varphi]\psi))$$

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

This logic can handle scenarios which involve both information and preference changes.

Digression: upgrade versus update Sometimes, it even offers alternative descriptions for one story. Take the example from Liu 2008 about buying a house. I am indifferent about buying one near the park or in town, but now I learn that a freeway will be built near the park, and I come to prefer the house in town. This may be described as a *2-world* model

• 'buy park house', • 'buy town house'

with an indifference relation between them, where a 'suggestion' upgrade leaves both worlds, but removes a ≤-link, leaving a strictly better town house. But alternatively, one could describe the buying scenario in terms of a *4-world* model with extended options

• 'park house, no freeway', • 'park house, freeway',

• 'town house, no freeway', • 'town house, freeway',

with betterness relations between them, where a public announcement 'freeway' removes 2 worlds to get the model we got before by upgrading. We will return to this issue below.

Completely similar points can be made about belief. One can take any complete dynamic logic of belief change as found, say, in van Benthem 2007A or Baltag & Smets 2006, and merge it with any dynamic logic of preference upgrade. This will then deal with combined notions like 'believing that φ is better', or it 'being better to believe φ'.

Second degree of entanglement*: *new modalities for intersections Still, the expressive power of the merged languages described here may not yet be suitable for getting at the real entanglement of preference and knowledge or belief. An epistemized preference formula $K(\varphi \rightarrow \langle\leq\rangle\psi)$ (subject to introspection, and knowledge-dependent) refers to ψ-worlds that are better than epistemically accessible φ–worlds, but there is no guarantee that these ψ–worlds are *themselves* epistemically accessible. But in our intuitive reading, for instance, of the normality sense of ceteris paribus preference, we made the betterness comparison *inside* the set of normal worlds (cf. again Lang, van der Torre & Weydert 2003), and likewise, we may want to make it inside the epistemically accessible worlds. [27]

To describe this, it makes sense to introduce a modal language that can talk about the *intersection* of the epistemic relation ~ and the betterness relation ≤. [28] That is,

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

Now we can define versions of '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 richer logic is no longer bisimulation-invariant, but it is not much more complex than the earlier one. And also, Liu 2008 notes how it supports exactly the same recursive style of dynamic analysis that we had before. In particular, the following law is valid:

---

[27] A similar entanglement, this time of epistemic and doxastic structure, is found in the work on belief revision by Baltag & Smets 2006, and van Eijck's Note in Apt & van Rooij, eds., 2008.

[28] Intersection really played already with the ceteris paribus logic *CPL*, where betterness became intersected with truth-value equivalence for a formula set Γ.

$$<\#(\varphi)><\leq\cap\sim>\psi \quad \leftrightarrow \quad (\neg\varphi \;\&\; <\leq\cap\sim><\#(\varphi)>\psi) \vee$$
$$(\varphi \;\&<\leq\cap\sim> (\varphi \;\&\; <\#(\varphi)>\psi)$$

Again, completely similar points hold for belief instead of knowledge, using intersection modalities with respect to betterness and plausibility relations between worlds. [29] Dynamic informational actions then include both announcements of hard information and various sorts of plausibility-changing 'soft information' that trigger belief revision.

Third degree of entanglement: preference and belief as duals Finally, all this piecemeal modal combination might still be too simple and technically driven. Preference and belief may also be taken to be totally inter-definable notions, and much of the literature on the foundations of decision theory (cf. Pacuit & Roy 2006 and the references therein) suggests that we can learn a person's beliefs from her preferences, as revealed by her actions [30] – and also vice versa, that we can learn her preferences from her beliefs. We leave the pros and cons of this conceptual connection as an open problem, which actually highlights the broader challenge of relating preference logic to decision theory.

9 Multi-agent interaction and group preference

As a final topic which I see as central to preference logic, I want to mention another feature of information dynamics which also makes sense for preference, viz. its *multi-agent interactive* character which also involves an analysis of groups as new agents in their own right. For a start, let us look at the most obvious interactive test-bed for logics of preference and information, making the earlier issues much more concrete, viz. *games*.

Game theory, epistemic preference logic, and backward induction Combined epistemic preference logics have already been applied to a variety of issues in games. Harrenstein 2004 used them to define Nash equilibrium, and van Otterloo 2005 has a further chapter on preferences of players, and how these change when further information becomes

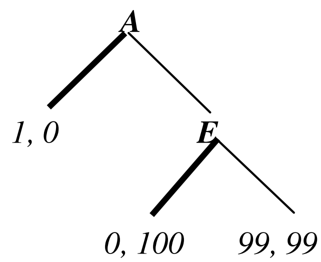
---

[29] Note that all issues discussed so far also arise in the *constraint-based* approach of Section 6.

[30] Cf. Lewis 1988 for a dissenting (though controversial) view on this Humean theme.

available about their ‘intentions’, i.e., the strategies that they will play from now on. Van Benthem 2007B discusses the role of ‘promises’ in games, viewed in a similar way as *public announcements of intentions*, while also discussing related settings where players’ preferences (encoded as betterness relations on nodes in the game tree) are not known. [31]

The entanglement of knowledge and belief with betterness and preference becomes quite concrete and vivid in this setting. Consider the well-known game solution procedure of *Backward Induction*. In the following picture, an equilibrium with outcomes *(1, 0)* will be computed by inductive bottom-up reasoning about players’ ‘rationality’ – incidentally, making both hugely worse off than the cooperative outcome *(99, 99):*



As pointed out in Board 1998, van Benthem 2002, the reasoning behind the standard Nash equilibrium here really rests on deriving *expectations* from the given betterness relations among end nodes, and then choosing moves accordingly. More concretely, there are three worlds, one for each complete history of the game, and the backward induction reasoning creates a plausibility ordering among these, which is actually the same for both players, with the world of *(1, 0)* on top, then that with *(0, 100)* and then that with *(99, 99)*. Thus in games, the plausibility relations that we merely stipulate in models for belief revision arise from an underlying analysis connecting belief with preference.

But we also see that this entanglement between belief and preference is not ‘absolute’. It depends crucially on assumptions that we make about the *type of agent* involved. One can only predict beliefs from people’s preferences by assuming, for instance, that they are

---

[31] Changes in games may improve their equilibria. E.g., in the game that follows, ***E*** might promise that she will not go left, and this public announcement changes the game to one with just the ‘right’ move for her – and a new equilibrium *(99, 99)* results.

rational utility-maximizing agents in the sense of decision theory or game theory. [32] Thus, I am not yet convinced that preference and belief are truly dual notions, as a majority view seems to have it. They rather seem like separate notions to me, though they may be connected tightly through making different assumptions on agents. And it would rather be a task for preference logic to sort out what natural assumptions are, in addition to the ubiquitous 'rationality', and how they may become subject to *explicit reasoning*.

Preferences and intentions Much more sophisticated scenarios are discussed in the dissertation Roy 2008, which is an extensive logic-inspired study of the role of *intentions* and commitments in decision making and game playing. Rational intentions are based on preferences, but they add further aspects of agents' capabilities and their *plans* for achieving goals, which are beyond our simple preference-based logic frameworks so far. While these richer models are definitely worthwhile, they lie beyond our horizon here.

Preference merge and group modalities While games still involve interaction between individual agents by themselves, the next obvious step is to introduce *groups* themselves as new collective agents. Indeed, game theorists study coalitions, while in epistemic and doxastic logic, common knowledge or common belief of groups has become a standard notion in understanding stable behaviour in communication and interaction. The naturally corresponding issue in preference logic would be how group preferences arise out of individual ones. This issue has also come up in belief revision theory, under the name of 'belief merge' for groups of agents who need to merge their plausibility relations.

A highly sophisticated paradigm for relation merge among many agents is that proposed in Andréka, Ryan & Schobbens 2002. It puts the relations to be merged in an ordered *priority graph* $\boldsymbol{G} = (G, <)$ of indices (which may have multiple occurrences), and sets

$$x \leq_{\boldsymbol{G}} y \text{ iff for all indices } i \in \boldsymbol{G}, \text{ either } x \leq_i y, \text{ or there is some } j > i \text{ in } \boldsymbol{G} \text{ with } x <_i y$$ [33]

[32] Incidentally, in this setting, it is crucial to make betterness comparisons with worlds that we believe will not happen: it is precisely those worlds which keep the actual prediction 'in place'.

[33] Thus, either x comes below y, or if not, y 'compensates' for this by doing better on some comparison relation in the set with a higher priority in the graph.

Girard 2008, Liu 2008 show how this elegant set-up generalizes (amongst many other things) the priority sequences of de Jongh & Liu 2007 in Section 6, as well as the 'agendas' we hinted at in connection with ceteris paribus preference logic (Section 7). Andréka, Ryan & Schobbens 2002 prove a number of interesting mathematical results about priority graphs, including their universality as a preference aggregation procedure for hierarchical groups, and a complete algebraic axiomatization. Girard 2008 provides an alternative complete axiomatization in a suitable modal language.

As for *dynamics* in this new two-level perspective, there are some natural operations for changing and combining priority graphs, viz. their *sequential* and *parallel composition*. These lead to an elegant calculus of graph operations and their induced group preference relations. This may be viewed as a compositional logical calculus of group preference, much richer than the simple set-based approaches which have been around in the literature – and it applies equally well to preference formation as belief merge.

Dynamics of social choice All this points at a junction between preference logic including group preferences and *social choice theory*. This is indeed where things seem to be heading these days. Preference logics with group preferences seem to be the natural counterpart to epistemic logics with various forms of group knowledge, and taken together, they provide a rich account of groups that can learn and form new preferences. Of course, much remains to be understood concerning the fine-structure of informative actions for groups, the ways in which they *deliberate*, and the ways in which agents are subject to preference change. These include at least two processes: (a) adjustment of one's initial preferences through social encounters, and (b) even leaving initial individual preferences intact, joining in the formation of new groups with preferences of their own. The empirical reality of voting procedures, and rules for rational discussion and debate would seem to provide excellent challenges for extended preference logic in this sense.

10 Conclusions and further issues

We have given an overview of dynamic logics of preference change as being developed in Amsterdam, first for individual agents, and eventually also for groups of agents. Many topics have been suppressed in this sketch [34], such as the use of *product update* (Baltag & Smets 2006) as a congenial but different methodology, *numerical plausibility and utility change* (dating back to Aucher 2003), and in particular, connections and contrasts with *probability* and decision theory. As to the latter, so far, nothing in our preference logics, 'entangled' or not, matches the role of *expected value* in decision and game theory, where utilities of alternative options are weighed probabilistically. How serious is this limitation? Does it relegate preference logic, no matter how broad and 'dynamic', to the side-lines forever? We do not know, but we do think that the presentation given here links preference logic in its traditional guise to exciting new developments in logic, computation, belief revision, and social choice theory (cf. Endriss & Lang, eds., 2005). And maybe that is quite enough for one paper.

Acknowledgment

I wish to thank Fenrong Liu and Jonathan Zvesper for their useful comments.

References

H. Andréka, M. Ryan, & P-Y Schobbens, 2002, 'Operators and Laws for Combining Preference Relations', *Journal of Logic and Computation* 12(1), 13 – 53.

K. Apt & R. van Rooij, eds., 2007, *Proceedings KNAW Symposium on Games and Interaction*, Texts in Logic and Games, Amsterdam University Press.

K. Apt and J. Zvesper, 2007, 'Common Beliefs and Public Announcements in Strategic Games with Arbitrary Strategy Sets'', CWI & ILLC Amsterdam.

G. Aucher, 2003, 'A Combined System for Update Logic and Belief Revision', Master's Thesis, ILLC, University of Amsterdam.

A. Baltag, H. van Ditmarsch & L. Moss, 2008, 'Epistemic Logic and Information Update', to appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam.

A. Baltag & S. Smets, 2006, 'Dynamic Belief Revision over Multi-Agent Plausibility Models', *Proceedings LOFT 2006*, Department of Computing, University of Liverpool. To appear in *Texts in Logic and Games*, Amsterdam, 2008.

[34] As well as ILLC Gloriclass fellows Andreas Witzel, Joel Uckelmans, and Cédric Dégrémont.

N. Belnap, M. Perloff & M. Xu, 2001, *Facing the Future*, Oxford University Press, Oxford.

J. van Benthem, 1996, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

J. van Benthem, 2002, 'Extensive Games as Process Models', *Journal of Logic, Language and Information* 11, 289–313.

J. van Benthem, 2006, 'One is a Lonely Number: on the Logic of Communication', in Z. Chatzidakis, P. Koepke & W. Pohlers, eds., *Logic Colloquium '02,* ASL & A.K. Peters, Wellesley MA, 96 – 129.

J. van Benthem, 2007A, 'Dynamic Logic of Belief Revision', *Journal of Applied Non-Classical Logics* 17, 129 – 155.

J. van Benthem 2007B, 'Rationalizations and Promises in Games', *Philosophical Trends*, 'Supplement 2006' on logic, Chinese Academy of Social Sciences, Beijing, 1–6.

J. van Benthem, 2008, 'Logic, Rational Agency, and Intelligent Interaction', Research Report ILLC Amsterdam. To appear in D. Westerståhl et al. eds., *Proceedings 14th Congress of Logic, Methodology and Philosophy of Science Beijing 2007*, College Publications, London.

J. van Benthem, J. van Eijck & A. Frolova, 1993, 'Changing Preferences', Report CS-93-10, Centre for Mathematics & Computer Science, Amsterdam.

J. van Benthem, J. van Eijck & B. Kooi, 2006, 'Logics of Communication and Change', *Information and Computation* 204(11), 1620 – 1662.

J. van Benthem, P. Girard & O. Roy, 2007, 'Everything Else Being Equal. A Modal Logic Approach to Ceteris Paribus Preferences', Institute for Logic, Language and Computation, University of Amsterdam. To appear in the *Journal of Philosophical Logic*.

J. van Benthem & F. Liu, 2007, 'Dynamic Logics of Preference Upgrade', *Journal of Applied Non-Classical Logics* 17, 157 – 182.

J. van Benthem, R. Muskens & A. Visser, 1997, 'Dynamics', a chapter in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier Science Publishers, Amsterdam, 587 – 648.

J. van Benthem, van Otterloo & Roy, 2006, 'Preference Logic, Conditionals, and Solution Concepts in Games', in H. Lagerlund, S. Lindström & R. Sliwinski, eds., *Modality Matters*, University of Uppsala, 61 – 76.

J. van Benthem & E. Pacuit, 2006, 'The Tree of Knowledge in Action', *Proceedings Advances in Modal Logic,* ANU Melbourne.

O. Board, 1998, 'Belief Revision and Rationalizability', *Proceedings TARK 1998*, 201 – 213.

C. Boutilier & M. Goldszmidt, 1993, 'Revision by Conditional Beliefs', *Proceedings AAAI 11*, Morgan Kaufmann, Washington D.C., 649 – 654.

C. Boutilier, 1994, 'Conditional Logics of Normality; A Modal Approach', *Artificial Intelligence* 68, 87 – 154.

B. ten Cate & Ch-ch Shan, 2002, 'The Partition Semantics of Questions, Syntactically', In *Proceedings of the ESSLLI-2002 student session*, Malvina Nissim, ed., 255 – 269. 14th European Summer School in Logic, Language and Information.

H. van Ditmarsch, W. van der Hoek & B. Kooi, 2007, *Dynamic-Epistemic Logic,* Springer Publishers, Berlin.

J. Doyle & M. Wellman, 1994, 'Representing Preferences as Ceteris Paribus Comparatives', in *Decision-Theoretic Planning: Papers from the 1994 Spring {AAAI} Symposium*, AAAI Press, Menlo Park, California, 69 – 75.

U. Endriss & J. Lang, eds., 2006, *Proceedings of the 1st International Workshop on Computational Social Choice (COMSOC-2006),* ILLC, University of Amsterdam.

P. Gärdenfors & H. Rott, 1995, 'Belief Revision', in D. M. Gabbay, C. J. Hogger & J. A. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming* 4, Oxford University Press, Oxford.

J. Gerbrandy, 1999, *Bisimulations on Planet Kripke*, Dissertation, Institute for Logic, Language, and Computation, University of Amsterdam.

P. Girard, 2008, *Modal Logics for Belief and Preference Change*, Dissertation, ILLC Amsterdam & Department of Philosophy, Stanford University.

S. Halldén, 1957, *On the Logic of "Better",* Gleerup, Lund.

J. Halpern, 1997, 'Defining Relative Likelihood in Partially-Ordered Preferential

Structure', *Journal of Artificial Intelligence Research* 7, 1 – 24.

S-O Hanson, 1995, 'Changes in Preference', *Theory and Decision* 38, 1 – 28.

S-O Hanson, 2001, 'Preference Logic', in D, Gabbay & F. Guenthner, eds., *Handbook of Philosophical Logic IV*, 319 – 393, Kluwer, Dordrecht.

P. Harrenstein, 2004, *Logic in Conflict*, Dissertation, Institute of Computer Science, University of Utrecht.

W. James, 1907, *Pragmatism: A New Name for Some Old Ways of Thinking*, David McKay, New York.

D. de Jongh & F. Liu, 2006, 'Optimality, Belief, and Preference', in S. Artemov & R. Parikh, eds., *Proceedings of the Workshop on Rationality and Knowledge*, ESSLLI Summer School, Malaga.

D. de Jongh & F. Liu, 2008, 'Preference, Priorities and Belief', paper under submission, ILLC, University of Amsterdam.

J. Lang, L. van der Torre & E. Weydert, 2003, 'Hidden Uncertainty in the Logical Representation of Desires', *Proceedings IJCAI XVIII*, 685 – 690.

J. Lang & L. van der Torre, 2008, 'From Belief Change to Preference Change', IRIT Toulouse & University of Luxemburg.

D. Lewis, 1973, *Counterfactuals*, Blackwell, Oxford.

D. Lewis, 1988, 'Desire as Belief, *Mind* 97, 323–332.

F. Liu, 2008, Changing for the Better: preference Dynamics and Agent Diversity, Dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

S. van Otterloo, 2005, *A Strategic Analysis of Multi-Agent Protocols*, Dissertation DS-2005-05, ILLC, University of Amsterdam & University of Liverpool.

E. Pacuit and O. Roy, 2006, 'Preference Based Belief Dynamics', *Proceedings of The 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2006),* Computer Science Department, University of Liverpool.

A. Prince & P. Smolensky, 1993, 'Optimality Theory: Constraint Interaction in Generative Grammar', Rutgers University Center for Cognitive Science.

H. Rott, 2001, *Change, Choice and Inference*, Oxford University Press, Oxford.

O. Roy, 2008, *Thinking before Acting: Intentions, Logic, and Rational Choice*, Dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

Y. Shoham, 1988, *Reasoning about Change*, The MIT Press, Cambridge, Mass.

P. Smolensky, 2006, *The Harmonic Mind*, The MIT Press, Cambridge, Mass.

P. Suppes, 1957, *Introduction to Logic*, Van Nostrand, Princeton and New York.

Y-H Tan & L. van der Torre, 1999, 'An Update Semantics for Deontic Reasoning', in P. McNamara & H. Prakken, eds., *Norms, Logics and Information Systems*, IOS Press, 73 – 90.

J. Väänänen, 2007, *Dependence Logic*, Cambridge University Press, Cambridge.

F. Veltman, 1996, 'Defaults in Update Semantics', *Journal of Philosophical Logic*, 25, 221 – 261.

G. H. von Wright, 1963, *The Logic of Preference*, Edinburgh University Press, Edinburgh.

T. Yamada, 2006, 'Acts of Command and Changing Obligations', in K. Inoue, K. Satoh & F. Toni, eds., *Proceedings CLIMA VII*. Also in *Lecture Notes in AI,* 4371 (2007), 1 – 19, Springer Verlag, Berlin.

B. Zarnic, 1999, 'Validity of Practical Inference', ILLC Research report PP-1999-23, University of Amsterdam.

J. Zvesper, 2008, 'What You Really Want', working paper Gloriclass Center, ILLC, University of Amsterdam.