Dynamic logic of preference upgrade

Johan van Benthem* — **Fenrong Liu****

*Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Netherlands. Department of Philosophy, Stanford University, USA.

E-mail: johan@science.uva.nl

**Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Netherlands. Institute of Philosophy, Chinese Academy of Social Sciences, China.

E-mail: fenrong@science.uav.nl

ABSTRACT. *Statements not only update our current knowledge, but also have other dynamic effects. In particular, suggestions or commands 'upgrade' our preferences by changing the current order among worlds. We present a complete logic of knowledge update plus preference upgrade that works with dynamic-epistemic-style reduction axioms. This system can model changing obligations, conflicting commands, or 'regret'. We then show how to derive reduction axioms from arbitrary definable relation changes. This style of analysis also has a product update version with preferences between actions, as well as worlds. Some illustrations are presented involving defaults and obligations. We conclude that our dynamic framework is viable, while admitting a further extension to more numerical 'utility update'.*

KEYWORDS: *preference upgrade, information update, dynamic logic*

1. Introduction: changing preferences

The notion of preference occurs across many areas, such as philosophy of action, decision theory, optimality theory, and game theory. Individual preferences between worlds or actions can be used to predict behavior by rational agents. More abstract notions of preference also occur in conditional logic, non-monotonic logic and belief revision theory, whose semantics order worlds by relative similarity or plausibility.

Preference logics Preference logics in the literature describe different comparative structures by means of various devices ([HAN 90]). Agents' preferences can run between worlds or between actions, preference statements can be weaker or stronger in what they say about worlds or actions being compared – and also, they may be more 'objective' or more 'epistemic'. A statement like "I prefer sunsets to sunrises" can be cast merely in terms of 'what is better for me', or as a more complex propositional attitude involving my beliefs about the relevant events. In this paper, we take an objective approach, where a binary preference relation supports a unary modality "true in some world which is at least as good as the current one" ([BOU 94], [HAL 97]). [BEN 06c] show how such a language, when extended with a few operators from hybrid languages, can define several conditionals, Nash equilibrium, and backward induction solutions to games. The language also expresses various kinds of preference that agents may have between propositions, i.e., types of events. Moreover, we add explicit epistemic operators, allowing us to express agents' attitudes toward what is good or better for them.

Preference dynamics Our main concern in this paper, however, is one of *dynamics*. Preferences are not static, but they change through commands of moral authorities, suggestions from friends who give good advice, or just changes in our own evaluation of worlds and actions. Such changes can have various triggers. For instance, intuitively, a command

"See to it that φ!"

makes worlds where φ holds preferred over those where it does not - at least, if we accept the preference induced by the issuer of the command. But also a process of planning, with just our own goals in mind, may gradually introduce preferences over actions as ways toward reaching the goal, as we learn more about the actual world. These and other dynamic aspects of preference have been noted by many authors, including [BEN 93], [HAN 95], [ZAR 03], [TOR 99], and [YAM 06].

Related ideas all play in the dynamic semantics for conditional logics ([SPO 88], [VEL 96]). In its static Lewis-style semantics, a conditional $\varphi \Rightarrow \psi$ says roughly that

ψ is true in all most-preferred φ-worlds ($\natural$)

But one plausible way of accepting a conditional is, not as a true/false description of a current preference, but rather as an instruction for *adjusting* that preference so as to make ($\natural$) the case. Even more simply, consider a so-called default assertion like

"*Normally* φ"

As [VEL 96] points out, this does not eliminate $\neg\varphi$-worlds from our current model, in the usual dynamic sense of information update. Accommodating this assertion rather makes the $\neg\varphi$-worlds doxastically less preferred than φ-worlds.

Trigger 1: suggestions There are many triggers for preference change, and dynamic preference logics should provide a format for studying these in an appropriate generality. To find such formats, in this paper, we start from a simple test scenario that may be called a 'suggestion'. Consider someone who is indifferent between taking a trip (p) and staying at home ($\neg p$). Now his friend comes along and says

"Let's take a trip!"

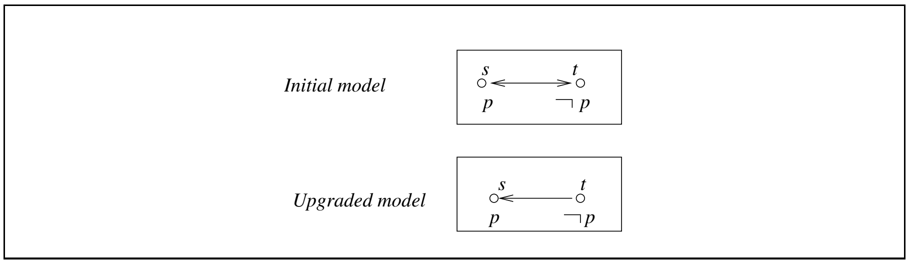

Figure 1.

'Taking' this suggestion means that any preference we might have had for staying at home is removed from the current model. Figure 1 shows we have in mind:

Thus, in our scenario, a suggestion removes already existing preference links: but it does not add new ones. Note that, in addition to arrows drawn, our preference relations always have reflexive loops. This mechanism will be studied in greater detail later on, as an entry into more general kinds of preference upgrade. Even so, by way of contrast, here is one alternative, which does not remove links, but rather adds them.

Trigger 2: commands In the above picture, the agent now prefers the trip, so this has become her priority, or in a deontic reading of the preference relation, her duty. But in general, suggestions are weaker than commands. Taking the suggestion does not mean that the person will now prefer all p-worlds to the $\neg p$-ones. It all depends on the preference structure already in place. If the agent was indifferent between p and $\neg p$ with arrows both ways, the suggestion induces a preference. But the agent may be unable to compare the two situations, as in this model with two unrelated worlds:

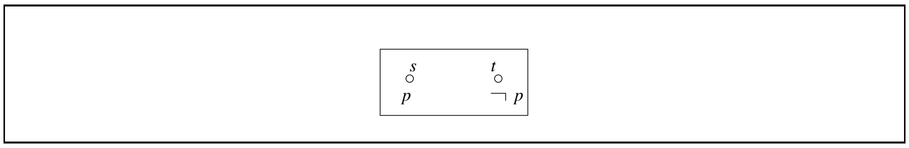

Figure 2.

A suggestion in the relation-decreasing sense does not make the worlds comparable. With real commands "Take that trip!", however, we want to make sure the agent now prefers p. Then, we need to *add* preference links to the picture, making the world with $\neg p$ less preferred. Our proposals also deal with upgrades that add links.

Dynamic logics of upgrade Whether eliminative or additive, preference change is reminiscent of existing systems for information *update* in dynamic-epistemic logic ([GER 99], [BAL 98], [BEN 06a], [DIT 06]). In the latter paradigm, incoming assertions or observations change the domain of the current model and/or its accessibility relations. In our scenario, current preference relations are changed by incoming suggestions or commands. Thus, we will speak henceforth of preference *upgrade* as a counterpart to the better-known term update. The main point of this paper is that preference upgrade is a viable phenomenon, just as susceptible to systematic modification as information, temporal perspective, or other parameters of 'logical dynamics' ([BEN 93], [BEN 96], or in the setting of conditional logic, [SPO 88], [VEL 96]). We will show how this dynamics can be implemented by the very same methodology that has been developed for information update in dynamic-epistemic logic.

This paper is structured as follows. First we present a new joint epistemic preference logic (Section 2). Its semantics is based on preferences between worlds. This allows us to talk about knowing or not knowing one's preferences, or regretting that the best scenario is not going to happen. Next, in Section 3, we provide formal definitions for preference upgrade, with an emphasis on the above 'suggestions' increasing our preference for one proposition over its negation. Interestingly, this also suggests alternative formulations for information update. Section 4 defines a dynamic version of the static epistemic preference language, where information update lives together with preference upgrade. There is a completeness theorem in terms of the usual style of reduction axioms recursively analyzing postconditions of actions. This is our first 'existence proof' for a compositional dynamics of upgrade, in tandem with update of information. In Section 5, we consider more general upgrade scenarios: first with general schemes of link elimination, and then, with the full strength of 'product update' for information using 'event (action) models'. This requires enriching the action models of dynamic-epistemic logic with agents' preferences between events. Section 6 then outlines some applications of our dynamic upgrade logics, to default reasoning, deontic logic, and logics of commands. Section 7 is a brief survey of related work, and Section 8 contains our conclusions and further directions.

This paper proposes a certain style of thinking about preference upgrade, and an existence proof for a logical methodology in doing so. We do not address all intuitive senses of preference, or all logical issues arising in the areas where it plays a role. A more extensive discussion of upgrade mechanisms with various triggers, various senses of preference, and further applications, is found in [JON 06] and [LIU 06b].

2. Epistemic preference logic

2.1. *Language and semantics*

The main language used in this paper has two components: a preference modality as in [BEN 06c], and the standard knowledge operators from epistemic logic.

DEFINITION 1. — *Take a set of propositional variables P and a set of agents I, with p ranging over P and i over I. The* ***epistemic preference language*** *is given by:*

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid [pref]_i\varphi \mid U\varphi.$$

Intuitively, $K_i\varphi$ stands for 'agent i knows that φ', while $[pref]_i\varphi$ says that all worlds which the agent considers as least as good as the current one satisfy φ. U is an auxiliary universal modality. [1]

How is this formal language connected to 'preference' as it occurs in natural discourse? One may be inclined to read $\langle pref\rangle_i\varphi$ as 'agent i prefers φ'. But as with other logical systems, there is a gap between the formalism and common usage. E.g., just saying that the agent sees some better world where φ holds seems too weak, while the universal modality $[pref]_i\varphi$ 'in all better worlds' seems much too strong. Cf. [HAN 01] for a thorough discussion of senses of preference, and ways in which formal languages do or do not match up. Here we just point out the following facts. First, our formal language can also express intermediate senses of 'betterness' for preference, using *combinations* of modalities. E.g., $[pref]_i\langle pref\rangle_i\varphi$ will express, at least on finite connected models, that some *best* world has φ. And one can also express that *all* best worlds satisfy φ: cf. [BEN 06c]. Moreover, our approach emphasizes comparisons of worlds, i.e., objects, rather than propositions, whereas common notions of preference often play between propositions, or semantically, sets of worlds. Such preferences between propositions can be *defined* on our approach (see again [BEN 06c]). For instance,

$$U(\psi \to \langle pref\rangle_i\varphi)$$

expresses one strong sense of 'agent *i* prefers φ to ψ', viz. each ψ-world s has at least one epistemic alternative which is φ and which is at least as good as s according to the agent. But one can also define the original notion of preference in [WRI 63] which says that the agent prefers all φ-worlds to all ψ-worlds (cf. [BEN 06c]; [BEN 06b] also deals with Von Wright's 'ceteris paribus' clause in the relevant comparisons between worlds). For the moment, we take this expressive power of our simple-looking modal language for granted. The virtue of our simple base modalities is that these 'decompose' more complex preference statements in a perspicuous manner, while allowing for a simple dynamic approach later on.

DEFINITION 2. — *An* ***epistemic preference model*** *is a tuple $\mathcal{M}$=(S, $\{\sim_i \mid i \in I\}$, $\{\preceq_i \mid i \in I\}$, V), with S a set of possible worlds, $\sim_i$ the usual equivalence relation of epistemic accessibility for agent i,*[2] *and V a valuation for proposition letters. Moreover, $\preceq_i$ is a reflexive and transitive relation over the worlds.*

We read $s \preceq_i t$ as 't is at least as good for agent i as s', or 't is weakly preferred to s'. If $s \preceq_i t$ but not $t \preceq_i s$, then t is *strictly preferred* to s, written as $s \prec_i t$. If

1. For technical convenience, we often shift to the corresponding existential modalities $\langle K\rangle_i$, $\langle pref\rangle_i$, and $E\varphi$. These seem more difficult to read in terms of intuitive linguistic expressions. But they help in finding and checking valid principles, and in semantic arguments generally.
2. Interpreting the knowledge operator with the equivalence relation is optional in an approach. There are many philosophical discussions about its justification. Various alternatives haven been proposed in terms of model classes. For complete epistemic logics over equivalence relations or other model classes, see the standard references, e.g. [FAG 95] or [BLA 01].

$s \preceq_i t$ and $t \preceq_i s$, then agent i is *indifferent* between s and t. Models can also have a distinguished actual world, but we rarely use this feature here.

Note that we do not require that our preference relations be *connected* in the sense of the Lewis sphere models for conditional logic. In general, we want to allow for genuinely incomparable worlds where an agent has no preference either way, not because she is indifferent, but because she has no means of comparing the worlds at all. This is just as in the semantics for the minimal conditional logic. Of course, in special settings, such as the standard utility-based preference orderings of outcomes in a game, connectedness may be quite appropriate.

DEFINITION 3. — *Given an epistemic preference model $\mathcal{M} = (S, \{\sim_i \mid i \in I\}, \{\preceq_i \mid i \in I\}, V)$, and a world $s \in S$, we define $\mathcal{M}, s \models \varphi$ (formula φ **is true in** $\mathcal{M}$ **at** s) by induction on φ:*

1. $\mathcal{M}, s \models p$ iff $s \in V(p)$
2. $\mathcal{M}, s \models \neg\varphi$ iff not $\mathcal{M}, s \models \varphi$
3. $\mathcal{M}, s \models \varphi \wedge \psi$ iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \psi$
4. $\mathcal{M}, s \models \langle K \rangle_i \varphi$ iff for some $t : s \sim_i t$ and $\mathcal{M}, t \models \varphi$
5. $\mathcal{M}, s \models \langle pref \rangle_i \varphi$ iff for some $t : s \preceq_i t$ and $\mathcal{M}, t \models \varphi$
6. $\mathcal{M}, s \models E\varphi$ iff for some t: $\mathcal{M}, t \models \varphi$.

Expressive power As we noted, [BEN 06c] have shown that the pure modal preference part of this language, with the help of the universal modality, can express a variety of natural notions of preference between propositions, including the original one proposed by Von Wright, as well as other natural options. Moreover, following [BOU 94], they show that this language can faithfully embed non-iterated conditionals $\varphi \Rightarrow \psi$ using the above preference operator $\langle pref \rangle_i$, as follows:

$$U(\varphi \rightarrow \langle pref \rangle_i(\varphi \wedge [pref]_i(\varphi \rightarrow \psi)).$$

But with our additional epistemic operators, we can also express the interplay of preference and knowledge. The following examples represent (a) an intuition of self-reflection of 'preference', and (b) an unfortunate but ubiquitous phenomenon:

– $\langle pref \rangle_i \varphi \rightarrow K_i \langle pref \rangle_i \varphi$: Preference Positive Introspection
– $\langle pref \rangle_i \varphi \wedge K_i \neg\varphi$: Regret.

We will return to mixed epistemic-preference principles later on.

2.2. *Proof system and completeness*

Our epistemic preference logic can be axiomatized completely in a standard modal style, given our choice of epistemic preference models (cf. [BLA 01]).

THEOREM 4. — *Epistemic preference logic is completely axiomatizable w.r.t epistemic-preference-models.*

PROOF. — The proof is entirely by standard techniques. ■

Additional axioms in our language impose further frame conditions on models. Here are two examples, based on standard modal frame-correspondence techniques:

FACT 5. —

– A preference frame $\mathcal{F} = (S, \{\sim_i | \ i \in I\}, \{\preceq_i | \ i \in I\})$ satisfies *connectedness*, i.e., $\forall x \forall y : x \preceq_i y \vee y \preceq_i x$, iff the following formula is true in the frame:

$(\varphi \wedge E\psi) \rightarrow \langle pref \rangle_i \psi \vee E(\psi \wedge \langle pref \rangle_i \varphi)$.

– An epistemic preference frame $\mathcal{F}$ makes the *Preference Introspection Axiom* $\langle pref \rangle_i \varphi \rightarrow K_i \langle pref \rangle_i \varphi$ true iff it satisfies the following first-order condition:

$\forall s \forall t \forall u : (s \preceq_i t \wedge s \sim_i u \rightarrow u \preceq_i t)$

Nevertheless, we will work with the minimal system described above in this paper, leaving such extras to asides.

3. Modelling preference upgrade

3.1. *Brief review of epistemic information update*

The basic paradigm for epistemic update is public announcement. Suppose that an agent does not know if p is the case, but learns this fact through an announcement $!p$. Then we get the following sort of model change, where the dotted line in the initial static model indicates the agent's uncertainty in the initial situation:

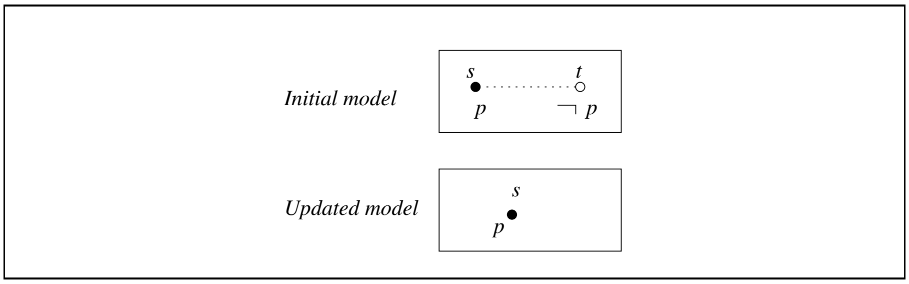

Figure 3.

The announcement eliminates the $\neg p$-world from the epistemic model, and afterwards, the agent knows that p. There is an extensive literature on dynamic epistemic logics for public announcements and more sophisticated epistemic events, that can modify information in different ways for different agents. See [BAL 98], [BEN 06a], and Section 5.4 below.

These logics all work essentially on the same design principle. First, a class of models is chosen representing the relevant information structures, together with some appropriate static language for describing these. Usually, these are models for some version of standard epistemic logic. Next, an update mechanism is proposed which transforms given models under some chosen set of epistemic actions. For public announcement, this simply eliminates worlds, yielding a definable submodel:

> *A public announcement* $!\varphi$ *of a true proposition* φ *turns the current model* $(\mathcal{M}, s)$ *with actual world* s *into the model* $(\mathcal{M}_{!\varphi}, s)$ *whose worlds are just the set* $\{w \in S \mid \mathcal{M}, w \models \varphi\}$*. And accessibility relations and valuations are retained on the restricted domain.*

More complex actions update to *products* $\mathcal{M} \times \mathcal{E}$ of the current epistemic model $\mathcal{M}$ with some 'event model' $\mathcal{E}$ containing all relevant events or actions.

Next, the static language gets a dynamic extension where the informative events themselves are displayed and manipulated. For public announcement, a typical static-dynamic assertion of this sort is

> $[!\varphi]K_i\psi$*: after a truthful public announcement of* φ*,*
> *the agent* i *knows that* ψ*.*

Here the semantic clause for the dynamic modality is simply as follows:

$$\mathcal{M}, s \models [!\varphi]\psi \textit{ iff (if } \mathcal{M}, s \models \varphi\textit{, then } \mathcal{M}_{!\varphi}, s \models \psi\textit{)}$$

Usually, the effects of events can then be described completely in a recursive manner, leading to a compositional analysis of communication and other cognitive processes. As a crucial illustration, here is the key *reduction axiom* in current logics of public announcement for a true assertion resulting in an epistemic possibility for agent i:

$$\langle !\varphi\rangle\langle K\rangle_i\psi \leftrightarrow \varphi \wedge \langle K\rangle_i\langle !\varphi\rangle\psi$$

As discussed in the literature, semantically, this reflects a sort of perfect recall for updating agents. Computationally, axioms like this help drive a reduction algorithm for dynamic epistemic statements to static epistemic statements, allowing us to borrow known decision procedures for the base language.

3.2. *Upgrade as relation change*

With the paradigm of public announcement in mind, we now define the mechanism of preference change described informally in the above. Our static models are of course the epistemic preference structures of Section 2:

$$\mathcal{M} = (S, \{\sim_i \mid i \in I\}, \{\preceq_i \mid i \in I\}, V)$$

Our triggers are events of publicly suggesting φ, written as follows:

$$\sharp\varphi$$

These lead to the following model change, removing preferences for $\neg\varphi$ over φ:

DEFINITION 6. — *Given any epistemic preference model* $(\mathcal{M}, s)$*, the* ***upgraded model*** $(\mathcal{M}_{\sharp\varphi}, s)$ *is defined as follows.*

(a) $(\mathcal{M}_{\sharp\varphi}, s)$ *has the same domain, valuation, epistemic relations, and actual world as* $(\mathcal{M}, s)$*, but*

(b) the new preference relations are now

$\preceq_i^* = \preceq_i - \{(s,t) \mid \mathcal{M}, s \models \varphi$ *and* $\mathcal{M}, t \models \neg\varphi\}$.[3]

We suppress agent subscripts henceforth whenever convenient.

Upgrade for suggestion events replaces a preference relation by a definable subrelation. This may be written as follows in the standard notation of dynamic logic ([HAR 00]):

$$R := R - (?\varphi; R; ?\neg\varphi)$$

We will consider more general relation-changing operations in Section 5. For instance, if one wanted to add links, rather than just subtract them, the format would still work. E.g., the relation-extending stipulation

$$R := R \cup (?\neg\varphi; \top; ?\varphi)$$

where $\top$ is the universal relation, would make every φ-world preferable to every $\neg\varphi$-world. With our upgrade defined, we are in a position to define a dynamic language for preference upgrade. But before doing so in Section 4, we consider some features of the mechanism just defined.

Preservation properties of upgrade Perhaps the most pressing issue is whether a proposed model changing operation stays inside the class of intended static models. For the update associated with public announcements $!\varphi$, this was so - and the reason is the general logical fact that submodels preserve *universally defined* relational properties like reflexivity, transitivity, and symmetry. For our notion of upgrade, the properties to be preserved are reflexivity and transitivity of preference relations (epistemic relations remain unchanged). This time, no general result comes to the rescue, since we only have the following counterpart to the preservation result for submodels:

FACT 7. — The first-order properties preserved under taking subrelations are precisely those definable using *negated atoms*, $\wedge$, $\vee$, $\exists$, $\forall$.

But neither reflexivity nor transitivity is of this particular syntactic form. Nevertheless, using some special properties of our proposal, we can prove

FACT 8. — The operation $\mathcal{M}_{\sharp\varphi}$ preserves reflexivity and transitivity.

PROOF. — Reflexivity is preserved since we never delete loops (s, s). As for transitivity, suppose that $s \preceq^* t \preceq^* u$, while not $s \preceq^* u$. By the definition of $\sharp\varphi$, we

3. [HAR 04] analyzes new defined preference relations in a set-theoretic format.

must then have $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, u \models \neg\varphi$. Consider the intermediate point t. Case 1: $\mathcal{M}, t \models \varphi$. Then the link (t, u) should have been removed from $\preceq$. Case 2: $\mathcal{M}, t \models \neg\varphi$. In this case, the link (s, t) should have been removed. Either way, we have a contradiction. ■

On the other hand, our upgrades $\sharp\varphi$ can lead to loss of connectedness of the preference order. Our earlier example already showed this in Section 1 (Figure 2). Likewise, our upgrades can lead to a loss of positive introspection, see the following scenario:

EXAMPLE 9. —

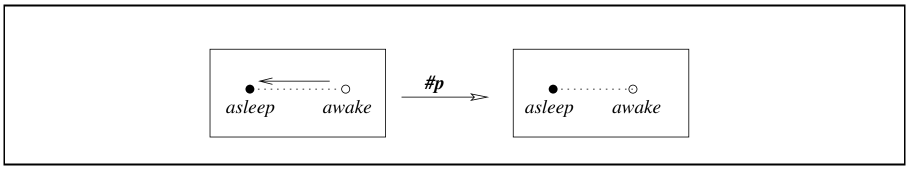

Figure 4.

There are two worlds 'asleep' and 'awake'. In both models, we do not know if we are sleeping or awake. Initially, we prefer being asleep, and we know our preference. Now an upgrade happens, suggesting that real waking life is not so bad after all. Then we still do not know if we are sleeping or awake, but at the 'awake' world we prefer being awake (thought not the case at the 'asleep' world). Focusing on the 'asleep' world in the new model, we still prefer being asleep there. But we no longer know that we prefer it – since we might be in the 'awake world'. Introspection fails! □

In some settings, preference introspection seems plausible, and a desirable property of models to be preserved. We can then change the above notion of upgrade to deal with this, e.g., by making sure that similar links are removed at epistemically indistinguishable worlds, or study which special sorts of upgrade in our language have the property of always preserving preference introspection. The latter would then be the 'reasonable' or 'sensible' series of suggestions.

Update by link cutting Update and upgrade do not lead wholly separate lives in our setting. For instance, if we want to model the earlier phenomenon of 'regret' about worlds that are no longer viable options, epistemic updates for $!\varphi$ should not remove the $\neg\varphi$-worlds, since we might still want to refer to them, and perhaps even mourn their absence. One way of doing this is by redefining the update for public announcement as a relation-changing operation of 'link cutting'. This time, instead of the above $!\varphi$, we write the relevant update action as follows, note that the notation '!' is now behind φ:

$$\varphi!$$

and we write the updated model as $\mathcal{M}_{\varphi!}$ in order to distinguish it from that we have by eliminating worlds. We should really change notations to reflect the two kinds of exclamation mark – but we trust the reader can disambiguate in context. The correct semantic operation for $\varphi!$ on models is this:

DEFINITION 10. — *The **modified public update model** $\mathcal{M}_{\varphi!}$ is the original model $\mathcal{M}$ with its worlds and valuation unchanged, but with accessibility relations $\sim_i$ replaced by a version without any crossing between the φ- and $\neg\varphi$-zones of $\mathcal{M}$:*

$$(?\varphi; \sim_i; ?\varphi) \cup (?\neg\varphi; \sim_i; ?\neg\varphi)$$

FACT 11. — The pure epistemic logic of public announcement is the same with $!\varphi$ and with $\varphi!$.

Nevertheless, the second update stipulation has some advantages. It was first proposed, in [SNY 04] (cf. [BEN 04]) for modelling the behavior of *memory-free* agents, whose epistemic accessibility relations are quite different from those for the idealized update agents of standard dynamic epistemic logic. Moreover, in the present setting, in stating regrets, we need the consistency of a formula like

$$K_i p \wedge \langle pref \rangle_i \neg p.$$

Yes, I know that p, but it would be better if it weren't... Modified update allows us to have this consistently.

Link cutting has some curious features, too. E.g., link cutting in the current model is the same for announcements $\varphi!$ and $(\neg\varphi)!$: both remove links between φ-worlds and $\neg\varphi$-ones. The only difference is that the former can only take place at a current world which satisfies φ, and the latter in one satisfying $\neg\varphi$. This is reflected in valid principles of the logic, but we do not pursue this issue here.

Discussion: update and upgrade Distinguishing the two versions of information update also leads to a subtle distinction in a combined update-upgrade logic. If processing $!\varphi$ eliminates all worlds we know to be non-actual, our preference statements adjust automatically to what we know about the facts. This is the behavior of realists, who never cry over spilt milk. For those realists i, the following combined announcement/preference principle will be valid, at least for atomic statements p which do not change their truth values by being announced

$$[!p][pref]_i p.$$

But this principle is not valid for more nostalgic souls, who still deplore the way things turned out to be. For them, update amounts to link-cutting $\varphi!$, they stick to their preferences between all worlds, and the new fact may even introduce regrets:

$$\langle pref \rangle_i \neg p \rightarrow [p!](\langle pref \rangle_i \neg p \wedge K_i p).$$

4. Dynamic Epistemic Upgrade Logic

4.1. *Language and semantics*

Now we introduce an enriched dynamic language for update and upgrade. Its static part is the earlier language of Section 2, but its action vocabulary contains both public announcements $\varphi!$ and suggestions $\sharp\varphi$. Adding the original world-eliminating announcements $!\varphi$ is a routine matter, so we highlight the latter less standard variant.

DEFINITION 12. — *Let P be a set of proposition letters and I a set of agents, with p ranging over P, i over I. The **dynamic epistemic preference language** is given by:*

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid [pref]_i\varphi \mid U\varphi \mid [\pi]\varphi$$
$$\pi ::= \varphi! \mid \sharp\varphi.$$

We could also add the usual program operations of composition, choice, and iteration from propositional dynamic logic to the action vocabulary - but we have no special use for these. The new language can be interpreted on epistemic preference models as follows, where we choose the 'regret' variant of update for the novelty:

DEFINITION 13. — *Given an epistemic preference model $\mathcal{M}$, the **truth definition** for formulas is as before, but with two new key clauses for the action modalities:*

$(\mathcal{M}, s) \models [\varphi!]\psi$ *iff if* $\mathcal{M}, s \models \varphi$, *then* $\mathcal{M}_{\varphi!}, s \models \psi$
$(\mathcal{M}, s) \models [\sharp\varphi]\psi$ *iff* $\mathcal{M}_{\sharp\varphi}, s \models \psi$.

4.2. *Preference upgrade logic*

On epistemic preference models, all valid principles of the static language of Section 2 still hold. Moreover, the usual axioms for public announcement hold, be it with one twist. As we saw, the usual updates $!\varphi$ eliminate all $\neg\varphi$-worlds, but updates $\varphi!$ leave all worlds in the model, cutting links instead. This makes no difference with purely epistemic dynamic axioms, but it does with global existential modalities over the whole domain of the model. The usual reduction axiom is this:

$$\langle !\varphi\rangle E\psi \leftrightarrow \varphi \wedge E\langle !\varphi\rangle\psi$$

But the axiom below is different, as $E\varphi$ can still refer to worlds after the update which used to be $\neg\varphi$. Further comments will be found below. We focus on what is new here: upgrade, and its interplay with modified update. It is easy to see the soundness of the following principles, stated with existential modalities for convenience:

THEOREM 14. — *The following formulas are valid:*

1. $\langle\varphi!\rangle p \leftrightarrow (\varphi \wedge p)$
2. $\langle\varphi!\rangle\neg\psi \leftrightarrow (\varphi \wedge \neg\langle\varphi!\rangle\psi)$

3. $\langle\varphi!\rangle(\psi \wedge \chi) \leftrightarrow (\langle\varphi!\rangle\psi \wedge \langle\varphi!\rangle\chi)$
4. $\langle\varphi!\rangle\langle K\rangle_i\psi \leftrightarrow (\varphi \wedge \langle K\rangle_i\langle\varphi!\rangle\psi)$
5. $\langle\varphi!\rangle\langle pref\rangle_i\psi \leftrightarrow (\varphi \wedge \langle pref\rangle_i\langle\varphi!\rangle\psi)$
6. $\langle\varphi!\rangle E\psi \leftrightarrow (\varphi \wedge E(\langle\varphi!\rangle\psi \vee \langle\neg\varphi!\rangle\psi))$

7. $\langle\sharp\varphi\rangle p \leftrightarrow p$
8. $\langle\sharp\varphi\rangle\neg\psi \leftrightarrow \neg\langle\sharp\varphi\rangle\psi$
9. $\langle\sharp\varphi\rangle(\psi \wedge \chi) \leftrightarrow (\langle\sharp\varphi\rangle\psi \wedge \langle\sharp\varphi\rangle\chi)$
10. $\langle\sharp\varphi\rangle\langle K\rangle_i\psi \leftrightarrow \langle K\rangle_i\langle\sharp\varphi\rangle\psi$
11. $\langle\sharp\varphi\rangle\langle pref\rangle_i\psi \leftrightarrow (\neg\varphi \wedge \langle pref\rangle_i\langle\sharp\varphi\rangle\psi) \vee (\langle pref\rangle_i(\varphi \wedge \langle\sharp\varphi\rangle\psi))$
12. $\langle\sharp\varphi\rangle E\psi \leftrightarrow E\langle\sharp\varphi\rangle\psi$

PROOF. — The first four formulas are the well-known valid reduction axioms for public announcement. The fifth formula, about commutation of $\langle\varphi!\rangle$ and $\langle pref\rangle_i$, expresses the fact that epistemic update does not change any preference relations. The special case of $E\varphi$ has been commented on above.

Next comes a similar set of reduction principles for upgrade. Axiom 7 is like Axiom 1, but simpler - as there is no precondition for $\sharp\varphi$: this operation can always be performed. Given that, we just state that atomic facts do not change under upgrade. The next two axioms express that upgrade is a function. Then comes a commutation principle for preference and knowledge which reflects the fact that upgrade does not change any epistemic relations.

Axiom 11 is crucial, as it encodes precisely how we changed the preference relation. It says essentially this. After an upgrade for φ, a preference link leads from the current world to a φ-world iff this same link existed before. This means that it has not been removed, ruling out the case where it led from an actual world verifying φ to some other one verifying $\neg\varphi$. The three cases where the link does persist are described succinctly in the two disjuncts on the right-hand side. Finally, as the upgrade may have changed truth values of formulas, we must be careful, and say that, before the upgrade, the link went to a world satisfying $\langle\sharp\varphi\rangle$ rather than φ.

The last axiom in the list is simply a commutativity principle for preference and existential modalities. ■

This dynamic epistemic upgrade logic (henceforth, *DEUL*) can explain general effects of changes in information and preference. In particular, we can think of our upgrade system as transforming underlying *world*- or *object*-comparison relations, but then, in the matching logic, recording also what changes take place because of this at the level of *propositions*. Thus, given the earlier-noted expressive power of the modal language for notions of preference between propositions, we can derive principles telling us what new propositional preferences obtain after an upgrade action, and relate these to the propositional preferences that we had before. As an illustration, consider the ‘For all There exists’ notion of preference stated earlier:

$$P^{\forall\exists}(\varphi, \psi) \quad \text{iff} \quad U(\psi \to \langle pref \rangle_i \varphi)$$

FACT 15. — The following equivalence holds

$$\langle \sharp A \rangle P^{\forall\exists}(\varphi, \psi) \quad \text{iff} \quad P^{\forall\exists}(\langle \sharp A \rangle \varphi, \langle \sharp A \rangle \psi) \wedge P^{\forall\exists}((\langle \sharp A \rangle \varphi \wedge A), (\langle \sharp A \rangle \psi \wedge A)).$$

PROOF 16. — This is a simple calculation showing how the $DEUL$ axiom system works in practice:

$$\begin{aligned}
&\langle \sharp A \rangle P^{\forall\exists}(\varphi, \psi) \leftrightarrow \langle \sharp A \rangle U(\psi \to \langle pref \rangle_i \varphi) \\
&\leftrightarrow U(\langle \sharp A \rangle (\psi \to \langle pref \rangle_i \varphi)) \\
&\leftrightarrow U(\langle \sharp A \rangle \psi \to \langle \sharp A \rangle \langle pref \rangle_i \varphi) \\
&\leftrightarrow U(\langle \sharp A \rangle \psi \to (\neg A \wedge \langle pref \rangle_i \langle \sharp A \rangle \varphi) \vee (\langle pref \rangle_i (A \wedge \langle \sharp A \rangle \varphi))) \\
&\leftrightarrow U(\langle \sharp A \rangle \psi \wedge \neg A \to \langle pref \rangle_i \langle \sharp A \rangle \varphi) \wedge U(\langle \sharp A \rangle \psi \wedge A \to \langle pref \rangle_i (\langle \sharp A \rangle \varphi \wedge A)) \\
&\leftrightarrow P^{\forall\exists}(\langle \sharp A \rangle \varphi, \langle \sharp A \rangle \psi) \wedge P^{\forall\exists}((\langle \sharp A \rangle \varphi \wedge A), (\langle \sharp A \rangle \psi \wedge A)).
\end{aligned}$$

■

A similar analysis applies Von Wright's 'All All' notion of preference between propositions, relating new preferences in this sense to earlier ones – but we leave this calculation to the reader. [4]

In addition, as noted earlier, our epistemic upgrade logic can deal with combined scenarios like introducing 'regret'. Say, a sequence of instructions

$$\sharp p; \neg p! \quad \text{for atomic } p$$

will first make p attractive, and afterwards, unobtainable. The logic records this as the (derivable) validity of regret principles like that at the end of Section 3:

$$\langle pref \rangle_i p \to [\sharp p][\neg p!](\langle pref \rangle_i p \wedge K_i \neg p)$$

DEUL can analyze the basic propositional scenarios of obeying successive commands or reasoning toward achieving practical goals proposed in [ZAR 03] and [YAM 06].

THEOREM 17. — *DEUL is completely axiomatized by the above reduction axioms.*

PROOF. — The reduction axioms, whose soundness we have already seen, are clearly sufficient for eventually turning every formula of our language into a static one without announcement or suggestion modalities. Then we can use the completeness theorem for our static language. ■

The same reduction method also shows that *DEUL* is decidable.

We have reached the first major conclusion of this paper:

Preference upgrade has a complete compositional logic-
just like, and even jointly with, knowledge update.

4. One might want to be more radical here, and insist on dynamic preference-changing actions directly *at the level of propositions*, without any dependence on an underlying world-level. This is in line with versions of belief revision theory where one is instructed to come to believe certain propositions. We have some thoughts on this alternative; but it would involve both entrenchment and preference relations on sets of propositions, a more syntactic perspective which raises as many design issues as the world-based semantic framework used in this paper.

4.3. *New issues of interest: coherence*

Despite the technical analogies between information update and preference upgrade, there are also intuitive differences. One typical illustration is the intuitive notion of 'coherence'. In pure public announcement logics, the only relevant aspects of coherence for a sequence of assertions seem to be these:

> (a) Do not make *inconsistent* and false assertions at the actual world; and, not to waste anyone's time: (b) Do not make assertions which are *common knowledge* in the whole group, and which do not change the model.

But in combination with upgrade, we can make other distinctions. E.g., the effect of a sequence with two conflicting suggestions

$$\sharp p; \sharp\neg p$$

is not inconsistency, but it still has some strange aspects. Generally speaking, such a sequence makes the ordering non-connected, as it removes arrows either way between p-worlds and $\neg p$-worlds. It is an interesting issue which sequences of upgrades are coherent, in that they leave connected preference relations connected.

In reality, one often resolves conflicts in suggestions by means of some authority ranking among the issuers of those suggestions. This is somewhat like the reality of information update. We often get contradictory information from different sources, and we need some notion of *reliability* differentiating between these to get to any sensible total update. Both issues go beyond the ambitions of this paper, as they involve the gap between actual informational events and their translation into the idealized model changes offered by dynamic epistemic logics, whether for update or upgrade.

5. Relation change and product upgrade

5.1. *Reduction axioms reflect definable operations*

To a logician, standard epistemic update $!\varphi$ essentially relativizes a model $\mathcal{M}$ to a definable submodel $\mathcal{M}_{!\varphi}$. The relation between evaluation at both sides is expressed in the following standard result:

FACT 18. — Assertions φ hold in the relativized model iff their syntactically relativized versions were true in the old model:

$$\mathcal{M}_{!\varphi} \models \psi \text{ iff } \mathcal{M} \models (\psi)^{\varphi}.$$

In this light, the reduction axioms for public announcement merely express the inductive facts about the modal assertion $\langle !\varphi \rangle$ referring to the left-hand side, relating these on the right to relativization instructions creating $(\psi)^{\varphi}$.

This same idea applies to preference upgrade $\sharp\varphi$. This time, the relevant semantic operation on models is *redefinition of base relations*. The same is true for the new link-cutting update operation $\varphi!$. [BEN 06a] notes how relativization and redefinition make up the standard notion of *relative interpretation* between theories in logic when objects are kept fixed - while product update relates to more complex reductions forming new objects as tuples of old objects. In this light, the reduction axioms for *DEUL* reflect a simple inductive definition, this time for what may be called *syntactic re-interpretation* of formulas. This operation leaves all logical operators unchanged, but it changes occurrences of the redefined relation symbol by its definition. There is one slight difference though. Relation symbols for preference only occur implicitly in our modal language, through the modalities. This is why the key reduction axiom in the above reflects a format of the following abstract recursive sort:

$$\langle R := def(R)\rangle\langle R\rangle\varphi \leftrightarrow \langle def(R)\rangle\langle R := def(R)\rangle\varphi.$$

5.2. *Dynamic logic of relation changers*

Further relation-changing operations can be defined, and make sense in our dynamic logics. We already mentioned the case of

$$R := R \cup (?\neg\varphi; \top; ?\varphi).$$

Here again, reduction axioms would be immediate, because of the following straightforward validities from propositional dynamic logic:

$$\langle R \cup (?\neg\varphi; \top; ?\varphi)\rangle\psi \leftrightarrow \langle R\rangle\psi \vee \langle ?\neg\varphi; \top; ?\varphi\rangle\psi$$
$$\leftrightarrow \langle R\rangle\psi \vee (\neg\varphi \wedge E(\varphi \wedge \psi)).$$

The example suggests a much more general observation, which we state informally:

FACT 19. — Every relation-changing operation that is definable in *PDL* without iteration has a complete set of reduction axioms in dynamic epistemic logic.

PROOF. — Clearly, every definition for a new relation $R^\sharp$ in this format is equivalent to a finite union of finite compositions of

(a) atomic relations R_i, (b) test relations $?\varphi$ for formulas of the base language.

The standard *PDL* axioms for union, composition, and tests in *PDL* then rewrite all statements $\langle R^\sharp\rangle\varphi$ to compounds in terms of just basic modalities $\langle R_i\rangle\varphi$. ■

This *PDL*-style analysis can even derive reduction axioms automatically:

EXAMPLE 20. — Our upgrade operation $\sharp\varphi$ is really the relation-changer:

$$R := (?\neg\varphi; R) \cup (R; ?\varphi).$$

Thus, the key reduction axiom can be derived as follows:

$$\begin{aligned}&\langle\sharp\varphi\rangle\langle R\rangle\psi\\ &\leftrightarrow \langle(?\neg\varphi;R)\cup(R;?\varphi)\rangle\langle\sharp\varphi\rangle\psi\\ &\leftrightarrow \langle ?\neg\varphi;R\rangle\langle\sharp\varphi\rangle\psi \vee \langle R;?\varphi\rangle\langle\sharp\varphi\rangle\psi\\ &\leftrightarrow (\neg\varphi\wedge\langle R\rangle\langle\sharp\varphi\rangle\psi)\vee\langle R\rangle(\varphi\wedge\langle\sharp\varphi\rangle\psi).\end{aligned}$$

The latter is just the version that we found 'by hand' in the above. □

But we can do still better than this, and achieve the same generality as dynamic epistemic logics for information update – as will be shown briefly now.

5.3. *Product update*

The usual generalization of eliminative public announcement is product update ([GER 99], [BAL 98], [DIT 06]). We briefly recall the basics.

DEFINITION 21. — *An* ***event model*** *is a tuple $\mathcal{E} = (E, \sim_i, PRE)$ such that E is a non-empty set of events, $\sim_i$ is a binary epistemic relation on E, PRE is a function from E to the collection of all epistemic propositions.*

The intuition behind the function PRE is that it gives the *preconditions* for an action: an event a can be performed at world s only if the world a fulfills the precondition $PRE(a)$.

DEFINITION 22. — *Given an epistemic model $\mathcal{M}$, an event model $\mathcal{E}$, the* ***product update model*** *$\mathcal{M}\times\mathcal{E}$ is defined as follows:*

– *The domain is {(s, a) | s a world in $\mathcal{M}$, a an event in $\mathcal{E}$, $(\mathcal{M}, s) \models PRE(a)$}.*

– *The new uncertainties satisfy $(s, a) \sim_i (t, b)$ iff both $s \sim_i t$ and $a \sim_i b$.*

– *A world (s, a) satisfies a propositional atom p iff s already did in $\mathcal{M}$.*

REMARK 23. — For a version leaving all old worlds in place, as with the above new announcement operator $\mathcal{M}_{\varphi!}$, we need to cut relational links again so as to 'isolate' the pairs (s, a) where $(\mathcal{M}, s)$ fails to satisfy the precondition for action a. □

DEFINITION 24. — *The language has new dynamic modalities $\langle\mathcal{E}, a\rangle$ referring to complex epistemic actions, and these are interpreted as follows:*

$$\mathcal{M}, s \models \langle\mathcal{E}, a\rangle\varphi \textit{ iff } \mathcal{M}\times\mathcal{E}, (s, a) \models \varphi.$$

This is the most powerful epistemic update calculus to date. As with public announcement, it yields a complete and decidable logic via a set of reduction axioms for all possible forms of postcondition (cf. [BAL 98], [BEN 93], [BEN 05b]).

5.4. *Product upgrade*

Next, we enrich epistemic event models with preference relations, indicating which events agents prefer over which others. These preferences may come from pay-offs or other benefits, but they may also be abstract relative plausibilities again, as in models of conditional logic.

DEFINITION 25. — *The output for* ***product upgrade on epistemic preference models*** *are again the above epistemic models $\mathcal{M} \times \mathcal{E}$. But this time, we keep all world/action pairs (s, a) represented, as these are the non-realized options that we can still have regrets about. Then it remains to set the new preferences, and here, we can just follow the above direct product rule for relations:*

$(s,t) \preceq_i (u,v)$ *iff* $s \preceq_i u$ *and* $t \preceq_i v$

This product upgrade covers at least the earlier upgrade instruction $\sharp p$ for suggestions. To see this, consider the event model of Figure 4:

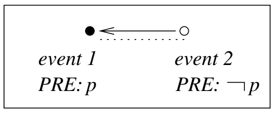

Figure 5.

Here the two events cannot be distinguished epistemically by the agent. Recall that the reflexive loops of preference relations are omitted.

FACT 26. — $\mathcal{M}_{\sharp\varphi} \cong \mathcal{M} \times \mathcal{E}^{\sharp\varphi}$, where the event model $\mathcal{E}^{\sharp\varphi}$ has two events "*seeing that* φ" (*event 1*), "*seeing that not-*φ" (*event 2*), with *event 2* $\preceq$ *event 1*.

PROOF. — From an epistemic viewpoint, the accessible part of $\mathcal{M} \times \mathcal{E}^{\sharp\varphi}$ merely *copies* the old model $\mathcal{M}$, as only one event can take place at each world. The old epistemic accessibilities just get copied with the product rule, since accessibility holds between all pairs of events. As for the new preference structure, consider any pair (s,t) in $\mathcal{M}$ where $\neg\varphi$ holds at s. Then the product model $\mathcal{M} \times \mathcal{E}^{\sharp\varphi}$ contains a unique corresponding pair

((*s, event 2*), (*t, event 1*)).

Our product upgrade rule gives a preference here from left to right. The only case where this copying from $\mathcal{M}$ fails is when the old preference and the event preference do not match up. But this only happens in those cases where $\sharp\varphi$ would reject an existing link, namely, when $s \preceq t$, while $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \neg\varphi$. ■

Thus, as with public announcement and epistemic product update, one simple event model suffices to mimick our base mechanism for update or upgrade.

Much more generally, every upgrade rule which takes a current preference relation to a *PDL*-definable subrelation can be dealt with in the same style as above, by putting in enough events and preconditions. There are of course much more complex event models still, with many more worlds and complex preference relations for agents. These represent more refined scenarios for joint update and upgrade.

Given the technical similarity of our product upgrade rule for preference to that for epistemic accessibility, the following is easy to see:

THEOREM 27. — *The dynamic logic of product update plus upgrade can be axiomatized completely by means of dynamic-epistemic-style reduction axioms.*

We do not spell out here what these axioms look like, but it is a routine exercise.

Our second main conclusion in this paper is this:

> *Preference upgrade can be combined naturally with the richest knowledge update mechanisms known so far.*

Virtues of the combination We think the above setting has independent interest. In philosophy, there is a well-known distinction between *preferences between states-of-affairs*, associated with 'consequentialist ethics', and *preferences between actions* in 'voluntarist ethics' (cf. [SCH 97]). Our product update system models both kinds, and is able to study their interplay. Moreover, there is a computational angle, viz. 'dynamic deontic' versions of *PDL* itself, starting from preferences between worlds, but moving on to preferences between actions ([MEY 88], [MEY 96]). [PUC 04] follows up on the latter, and propose relation change as a way of 'changing policies'. [ROH 05] provides a general background for this in so-called 'sabotage modal logic', where arbitrary links can be cut from models.

Thus, we see our product upgrade system also as one principled 'preferentialized' version of propositional dynamic logic.

6. Illustrations: defaults and obligations

We have presented an upgrade mechanism for incoming triggers that change preferences. We now illustrate this framework in two concrete settings. Our aim is not some full-fledged application to existing systems. We merely show how the logical issues in this paper correspond to real questions of independent interest.

6.1. *Default reasoning*

Consider practical reasoning with default rules of the form "*if* φ*, then* ψ":

> "If I take the train right now, I will be home tonight".

These are defeasible conditionals, which recommend concluding ψ from φ, but without excluding the possibility of $\varphi \wedge \neg\psi$-worlds, be it that the latter are now considered exceptional circumstances. Intuitively, the latter are not 'ruled out' from our current model, but only 'downgraded' when a default rule is adopted. [VEL 96] is an influential dynamic treatment, making a default an instruction for changing the current preference order between worlds. The simplest case has just one assertion φ which is being 'recommended' - in Veltman's terms, there is an instruction "*Normally,* φ". From our perspective, one can go this way, using a scenario of *relation change for defaults*, as in our earlier Section 3. Suppose that we want to give an incoming default rule "*Normally,* φ" 'priority', in that after its processing, all best worlds are indeed φ-worlds. Here is a more drastic procedure, which will validate the preceding intuition:

DEFINITION 28. — *We make all φ-worlds better than all $\neg\varphi$-worlds, and within the φ- and $\neg\varphi$-areas, we leave the old preferences in place.*[5] *Formally, this is one of our earlier PDL-style relation-changes: the old preference relation R becomes*

$$(?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi).$$

Interestingly, this is the union of the earlier link cutting version of public announcements $\varphi!$ plus the upgrade operation with relation extension considered in Section 4.

FACT 29. — Relational default processing can be axiomatized completely.

PROOF. — By the method of Section 5.2, the key reduction axiom follows automatically from the given PDL-form, yielding

$$\langle upgr(\varphi)\rangle\langle pref\rangle\psi \leftrightarrow (\varphi\wedge\langle pref\rangle(\varphi\wedge\langle upgr(\varphi)\rangle\psi)\vee(\neg\varphi\wedge\langle pref\rangle(\neg\varphi\wedge \langle upgr(\varphi)\rangle\psi)) \vee (\neg\varphi \wedge E(\varphi \wedge \langle upgr(\varphi)\rangle\psi)).$$

■

Thus, we have a plausible version of default logic in our upgrade setting. Moreover, their validities are axiomatizable in a systematic style via reduction axioms, rather than more ad-hoc default logics found in the literature.

Things need not stop here. E.g., the relation-changing version puts heavy emphasis on the last suggestion made, giving it the force of a command. This seems too strong in many cases, as it gears everything toward the last thing heard. A more reasonable scenario is this. We are given a sequence of instructions inducing preference changes, but they need not all be equally urgent. We need to find out our total commitments eventually. But the way we integrate these instructions may be partly left up to the *policy* that we choose, partly also to another parameter of the scenario: viz. the relative force or *authority* of the issuers of the instructions. One particular setting where this happens is again Optimality Theory. Ranked constraints determine the order of authority, but within that, one counts numbers of violations. Cf. [PRI 93] for a good exposition, and [JON 06] for a logical exploration.

From default logic to belief revision Default logic is naturally connected with *belief revision*, since new facts may change earlier conclusions. More generally, an

5. This is known as the 'lexicographic' change in the belief revision community. The idea was first suggested in [NAY 94].

analysis of preference change seems very congenial to analyzing belief revision, with world ordering by relative plausibility (cf. [GRO 88], [ROT 06]). Indeed, the companion paper ([BEN 06a]) in this volume shows that the techniques for handling relation change developed in this paper can be used to analyze various belief revision policies, and axiomatize their properties completely.

6.2. *Deontic logic and commands*

Similar considerations apply to deontic logic ([AQV 87]). Originally, this was the study of assertions of obligation

$O\varphi$: 'it ought to be the case that φ',

as well as statements of conditional obligation $O(\varphi|\psi)$, say, emanating from some moral authority. The sum total of all true O-statements represents all the obligations an agent has at the current stage.

In the standard semantics of deontic logic, $O\varphi$ is treated as a universal modality over some deontic accessibility relation. But the intuition is that those φ ought to be case which are true in *all best possible worlds*, as seen from the current one. Again, this suggests a preference order among worlds. And then, once more, we can think of this setting dynamically, using our upgrade scenario.

Initially, there are no preferences between worlds. Then some moral authority starts 'moralizing': introducing evaluative distinctions between worlds. If this process works well, we get a new ordering of worlds from which our current obligations may be computed, as those assertions which are true in all best worlds. Whether a sequence of commands makes sense in this way may depend on more than consistency: and the issue of 'coherence' in Section 3 comes back with greater force now.

Looking backward, or forward in upgrade Deontic logic also raises new issues. One semantic intuition is that, after a command (say, 'Thou shalt not kill'), the core proposition becomes true in all best possible worlds. Thus, in commands, there is a future-oriented aspect:

'See to it that φ' should result in a new situation where $O\varphi$ is true.

But as we have seen in Section 4, not every upgrade $\sharp\varphi$ has the effect that φ becomes true in the new most preferred worlds. Indeed, there is a general difficulty with specifications of the form 'See to it that φ'. Dynamic epistemic logic is mainly about events with their preconditions. Thus, the information one gets from an event is *past-oriented*, describing what was the case at the time the event happened. But, even a simple epistemic event can change the truth value of assertions at worlds - witness public announcements turning ignorance into knowledge.

But it is not so easy to just define an action as achieving the truth of some proposition. This works for simple factual effects of actions like opening a door ([BEN 06c]), but it is not clear what this should even mean with more complex stipulations. E.g., there is no obvious 'seeing to it that' arbitrary mixtures of knowledge and ignorance in groups arise, and the same seems true of complex deontic commands. Whether deontic reasoning needs some sort of future-oriented update and upgrade seems an interesting question. For temporal logics of such $STIT$ operators, cf. [BEL 01].

7. Related work

The ideas in this paper have a long history, and there are many proposals in the literature having to do with 'dynamification' of preferences, defaults, and obligations. We just mention a few related approaches here, though we do not make any detailed comparisons. [MEY 88] was probably the first to look at deontic logic from a dynamic point of view, with the result that deontic logics are reduced to suitable versions of dynamic logics. This connection has become a high-light in computer science since, witness the regular *DEON* conference series. In a line that goes back to [SPO 88], [VEL 96] presents an update semantics for default rules, locating their meaning in the way in which they modify expectation patterns. This is part of the general program of 'update semantics' for conditionals and other key expressions in natural language. [TOR 99] use ideas from update semantics to formalize deontic reasoning about obligations, but with motivations from computer science. In their view, the meaning of a normative sentence resides in the changes it brings about in the 'ideality relations' of agents to whom the norm applies. [MEY 96] takes the deontic logic/dynamic logic interface a step further, distinguishing two notions of permission, one of which, 'free choice permission' requires a new 'dynamic logic of permission', where preferences can hold between actions. Completeness theorems with respect to this enriched semantics are given for several systems. Taking belief change as its starting point, [HAN 95] identified four types of changes in preference, namely revision, contraction, addition and substraction, and showed that they satisfy plausible postulates for rational changes in preferences. [PUC 04] provide a dynamified version of the dynamic logic of permission, in order to deal with building up of agents' policies by adding or deleting transitions. [DEM 05] reduces an extension of van der Meyden's logic to propositional dynamic logic, yielding an EXPTIME decision procedure, and showing how dynamic logic can deal with agents' policies. Following van Benthem's 'sabotage games', [ROH 05] studies general modal logics with operators that describe effects of deleting arbitrary transitions - without a fixed upgrade definition as in our analysis. Model checking for such logics becomes Pspace-complete, and satisfiable is undecidable. [PAC 06] observe that an agent's obligations are often dependent on what she knows, and introduce a close relative of our epistemic preference language, but over temporal tree models. They provide distinctions, like knowing one's duty versus having a duty to know, whose dynamics invites a merge with our system. Our own approach goes back to [BEN 93], which discusses general formats for upgrading preference relations. [ZAR 03] uses similar ideas, combined with a simple update logic to formalize natural language imperatives of the form *FIAT* φ, which can be used in describing the search for solutions of given planning problems. More generally, [YAM 06] takes the update paradigm to logics of commands and obligations, modelling changes brought about by various acts of commanding. It combines a multi-agent variant of the language of monadic deontic logic with a dynamic language for updates and commands. This is closest to what we do. Yamada's command operator for propositions A can be modelled *exactly* as an upgrade sending R to $R; ?A$ in our system. But this paper provides a much more general treatment of possible upgrade instructions. Finally, [ROT 06] presents a format for relation change which can handle all major current policies for belief revision. [BEN 06a] shows how one can axiomatize such policies completely using the methods in Section 5.2 of this paper.

A full-fledged comparison doing justice to all these approaches is unfortunately beyond the scope of this paper.

8. Conclusion

Preference upgrade seems a natural and crucial part of logical dynamics. We have shown it can be modelled as relation change in a standard dynamic format, up to the expressive level of the best available system, that of epistemic product update.

Still, our approach leaves things to be desired. In particular, many settings call for more finely-grained distinctions as to *intensity* of preferences, as happens in quantitative versions of social choice theory. The extended version of this paper, available on-line as [BEN 05a], defines a mechanism of *utility update*, inspired by [SPO 88], [AUC 03] and [LIU 04], which combines utilities of old worlds and of events to compute utilities of new worlds. With such a system, we can upgrade defaults, duties, or preferences in games in a more controlled local fashion, by adding or subtracting 'points'. The relationship between our relational upgrade and utility update also poses some interesting technical issues, for which we refer to [BEN 05a] and the more extensive exploration in [LIU 06a].

Acknowledgements

We thank Hans van Ditmarsch and Andreas Herzig for the opportunity to present this work at the ESSLLI 2005 Summer School in Edinburgh, whose audience provided useful feedback. Special thanks go to Hans van Ditmarsch, Ulle Endriss, Paul Harrenstein, and two anonymous referees, for their helpful comments.

9. References

[AQV 87] AQVIST L., *Introduction to Deontic Logic and the Theory of Normative Systems*, Naples: Bibliopolis, 1987.

[AUC 03] AUCHER G., "A Combination System for Update Logic and Belief Revision", Master's thesis, ILLC, University of Amsterdam, 2003.

[BAL 98] BALTAG A., MOSS L., SOLECKI S., "The Logic of Common Knowledge, Public Announcements, and Private Suspicions", GILBOA I., Ed., *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, 1998, p. 43-56.

[BEL 01] BELNAP N., PERLOFF M., XU M., *Facing the Future*, Oxford University Press, Oxford, 2001.

[BEN 93] VAN BENTHEM J., VAN EIJCK J., FROLOVA A., "Changing Preferences", Tech Report, CS-93-10, Center for Mathematics and Computer Science, Amsterdam, 1993.

[BEN 96] VAN BENTHEM J., *Exploring Logical Dynamics*, CSLI Publication, Stanford, 1996.

[BEN 04] VAN BENTHEM J., LIU F., "Diversity of Logical Agents in Games", *Philosophia Scientiae*, vol. 8(2), 2004, p. 163-178.

[BEN 05a] VAN BENTHEM J., LIU F., "Dynamic Logic of Preference Upgrade", Tech Report, PP-2005-29, ILLC, University of Amsterdam. The extended version of the present paper, 2005.

[BEN 05b] VAN BENTHEM J., VAN EIJCK J., KOOI B., "Common Knowledge in Update Logics", *TARK*, 2005, p. 253-261, extended version to appear in *Journal of Information and Computation* as 'Logic of Communication and Change'.

[BEN 06a] VAN BENTHEM J., "Dynamic Logic for Belief Change", Tech Report, PP-2006-11, ILLC, University of Amsterdam. To appear in the *Journal of Applied Non-Classical Logic*, 2006.

[BEN 06b] VAN BENTHEM J., GIRARD P., ROY O., "Logics for Preference Ceteris Paribus", manuscript. Stanford University, 2006.

[BEN 06c] VAN BENTHEM J., VAN OTTERLOO S., ROY O., "Preference Logic, Conditionals and Solution Concepts in Games", LAGERLUND H., LINDSTRöM S., SLIWINSKI R., Eds., *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, p. 61-77, Uppsala Philosophical Studies 53, 2006.

[BLA 01] BLACKBURN P., DE RIJKE M., VENEMA Y., *Modal Logic*, Cambridge University Press, 2001.

[BOU 94] BOUTILIER C., "Conditional Logics of Normality: a Modal Approach", *Artificial Intelligence*, vol. 68, 1994, p. 87-154.

[DEM 05] DEMRI S., "A Reduction from DLP to PDL", *Journal of Logic and Computation*, vol. 15, 2005, p. 767-785, Oxford University Press.

[DIT 06] VAN DITMARSCH H., VAN DER HOEK W., KOOI B., *Dynamic Epistemic Logic*, Springer, Berlin, 2006, to appear.

[FAG 95] FAGIN R., HALPERN J., MOSES Y., VARDI M., *Reasoning about Knowledge*, MIT Press, Cambridge (Mass.), 1995.

[GER 99] GERBRANDY J., "Bisimulations on Planet Kripke", PhD thesis, ILLC Amsterdam, 1999.

[GRO 88] GROVE A., "Two Modelings for Theory Change", *Journal of Philosophical Logic*, vol. 17, 1988, p. 157-170.

[HAL 97] HALPERN J. Y., "Defining Relative Likelihood in Partially-ordered Preferential Structure", *Journal of Artificial Intelligence Research*, vol. 7, 1997, p. 1-24.

[HAN 90] HANSON S., "Preference-based Deontic Logic", *Journal of Philosophical Logic*, vol. 19, 1990, p. 75-93.

[HAN 95] HANSSON S. O., "Changes in Preference", *Theory and Decision*, vol. 38, 1995, p. 1-28.

[HAN 01] HANSSON S. O., *The Structure of Values and Norms*, Cambridge University Press, 2001.

[HAR 00] HAREL D., KOZEN D., TIURYN J., *Dynamic Logic*, MIT Press, 2000.

[HAR 04] HARRENSTEIN P., "Logic in Conflict. Logical Explorations in Strategic Equilibrium", PhD thesis, Utrecht University, 2004.

[JON 06] JONGH D. D., LIU F., "Optimality, Belief and Preference", Tech Report, PP-2006-38, ILLC, University of Amsterdam. To appear in: S. Artemov and R. Parikh eds, *Proceedings of the Workshop on Rationality and Knowledge,* ESSLLI, Malaga, 2006.

[LIU 04] LIU F., "Dynamic Variations: Update and Revision for Diverse Agents", Master's thesis, ILLC, University of Amsterdam, 2004.

[LIU 06a] LIU F., "Preference Change and Information Processing", Tech Report, PP-2006-41, ILLC, University of Amsterdam. To appear in *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 06)*, Liverpool, 2006.

[LIU 06b] LIU F., "Preference Changes in Games", working paper, ILLC, University of Amsterdam, 2006.

[MEY 88] MEYER J.-J., "A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic", *Notre Dame Journal of Formal Logic*, vol. 29, 1988, p. 109-136.

[MEY 96] VAN DER MEYDEN R., "The Dynamic Logic of Permission", *Journal of Logic and Computation*, vol. 6:3, 1996, p. 465-479.

[NAY 94] NAYAK A., "Iterated Belief Change Based on Epistemic Entrenchment", *Erkentnis*, vol. 41, 1994, p. 353-390.

[PAC 06] PACUIT E., PARIKH R., COGAN E., "The Logic of Knowledge Based on Obligation", to appear in *Knowledge, Rationality and Action (Synthese)*, 2006.

[PRI 93] PRINCE A., SMOLENSKY P., *Optimality Theory: Constraint Interaction in Generative Grammar*, Malden, Ma: Blackwell, 1993.

[PUC 04] PUCELLA R., WEISSMANN V., "Reasoning about Dynamic Policies", *Proceedings FoSSaCS-7*, Springer Lecture Notes in Computer Science 2987, 2004, p. 453-467.

[ROH 05] ROHDE P., "On Games and Logics over Dynamically Changing Structures", PhD thesis, Department of Informatics, Technische Hochschule Aachen (RWTH), 2005.

[ROT 06] ROTT H., "Shifting Priorities: Simple Representations for 27 Iterated Theory Change Operators", LAGERLUND H., LINDSTRÖM S., SLIWINSKI R., Eds., *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, p. 359-384, Uppsala Philosophical Studies 53, 2006.

[SCH 97] SCHEFFLER S., "Relationships and Responsibilities", *Philosophy and Public Affairs*, vol. 26, 1997, p. 189-209.

[SNY 04] SNYDER J., "Product Update for Agents with Bounded Memory", manuscript, Department of Philosophy, Stanford University, 2004.

[SPO 88] SPOHN W., "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States", HARPER W. L. *et al.*., Ed., *Causation in Decision, Belief Change and Statistics II*, p. 105-134, Kluwer, Dordrecht, 1988.

[TOR 99] VAN DER TORRE L., TAN Y., "An Update Semantics for Deontic Reasoning", MCNAMARA P., PRAKKEN H., Eds., *Norms, Logics and Information Systems*, p. 73-90, IOS Press, 1999.

[VEL 96] VELTMAN F., "Defaults in Update Semantics", *Journal of Philosophical Logic*, vol. 25, 1996, p. 221-261.

[WRI 63] VON WRIGHT G. H., *The Logic of Preference*, Edinburgh, 1963.

[YAM 06] YAMADA T., "Commands and Changing Obligations", *Proceedings of the Seventh International Workshop on Computational Logic in Multi-Agent Systems (CLIMA VII)*, 2006.

[ZAR 03] ZARNIC B., "Imperative Change and Obligation to Do", SEGERBERG K., SLIWINSKI R., Eds., *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Aqvist*, p. 79-95, Uppsala philosophical studies 51. Uppsala: Department of Philosophy, Uppsala University, 2003.