Dynamic doxastic logic for defeasible belief revision

Hans van Ditmarsch and Willem Labuschagne *

July 5, 2003

Abstract

An agent may update and revise not merely his knowledge and belief about *facts* but also those about the *beliefs* and *knowledge* of other agents. We define a language wherein both knowledge changes, as the result of definite updates, and belief changes, as the result of changing default rules, can be expressed in the object-level logical language.

1 Introduction

Both belief revision and knowledge change have independently been on the research agenda for quite a while [AGM85, KLM90, vB96, FHMV95].

Belief revision has been studied from the perspective of structural properties of reasoning about changing beliefs, from the perspective of changing, growing and shrinking knowledge bases, and from the perspective of models and other structures of belief change, wherein such knowledge bases may be interpreted, and / or that respect assumed properties of reasoning about beliefs. Such models are the starting point of our investigations. A typical approach involves preferential orders to express increasing or decreasing degrees of belief [KLM90, FL02, MLH00a, Mey01].

Knowledge change has been extensively studied in what is by now called dynamic epistemic logic, with seminal publications of [Pla89, GG97, Ger99, BMS98, vD00, vD02, Bal02, tC02]. This research has been mainly driven by the attempt to model higher-order belief change phenomena, and was initially motivated by the attempt to model so-called 'unsuccessful updates', as in the well-known muddy children problem: from a public update with 'nobody knows whether he/she is muddy', the muddy children may learn that they are muddy. Such research still has to be carried to the stage of integrated frameworks for factual, knowledge and belief change. Our work is a proposal for integrated knowledge and belief change.

*Computer Science, University of Otago, PO Box 56, Dunedin 9015, New Zealand; {hans,willem}@cs.otago.ac.nz

So far, interactions between these two different research directions appear to have been somewhat limited, notwithstanding outstanding expositions that relate them [Seg99]. This may be due to a different choice for accessibility relations in models for belief and, respectively, knowledge: namely templated (preferential) orders versus simple binary relations. Another reason for the lack of interaction however, appears to be the assumed complexity of higher-order belief change: apart from revising or updating one's beliefs about facts, one may also want to revise one's beliefs about other agent's beliefs, including his/her different degrees of belief. Although this has been incorporated in dynamic epistemics, analogues have not been standardly accepted in the belief revision community. We propose relational structures where each agent has a set of accessibility relations associated with him/her that correspond to such preferences, and define a language for integrated and higher-order change of both knowledge and belief to be interpreted on such structures.

We start with a motivating example.

An operator monitors a control system that consists of a light and a fan. Both can be on and off. Neither is known. The light is normally on. Both light and fan are in fact off. The operator uses factual knowledge and defeasible beliefs to reason about the system. Defeasible beliefs are based on default rules, such as 'the light is normally on'. Degrees of belief can be modelled with a *sphere semantics*, where each of some set of factual states (possible worlds) is associated with a templated order, that corresponds to the default rule for that state [KLM90, MLH00b]. If a distinguished outermost sphere is incorporated which corresponds to 'definitely excluded states', knowledge can be modelled as well [FL02, FHLvdW03]. For interpreted and other multiagent systems ('$S5$-like') the system of spheres reduces to the following: the agent's knowledge induces a partition on the set of factual states, and default rules induce a templated order ('concentric circles') on each class of the partition. Our control system can thus be represented by an information state (pointed Kripke model) $(M, 00)$ below, where, e.g., 10 stands for the state where the light is on and the fan off, and point 00 is underlined. Interdependent belief operators are defined using these templates, including B for 'normal beliefs' and K for 'knowledge'. For example, $\neg Bq$ expresses that the operator does not ($\neg$) normally believe (B) that the fan is on (q). This holds in all states of M.

The operator may wish to revise his beliefs and his knowledge. If *the operator gets confirmation that the light is off* (Rev'), information state M' results (see Figure 1). Such 'public updates' can be modelled in dynamic epistemics [GG97, Bal02, vD02]. If *the operator has reason to assume that if the light is on, then the fan is normally also on* (Rev''), he makes an exception of state 10 as well. He has now revised his beliefs but not his knowledge. Information state M'' in Figure 1 results. Unlike proposals where belief revision is treated on the metalevel [KLM90], we propose to express such revisions on the object level, by adding dynamic modal operators for belief revision to the language. For example, $[\mathsf{Rev}']K\neg p$ stands for: 'after having confirmed that (Rev'), the operator knows (K) that the light is off ($\neg p$)' (in terms of [GG97]: $[\neg p]K\neg p$), and

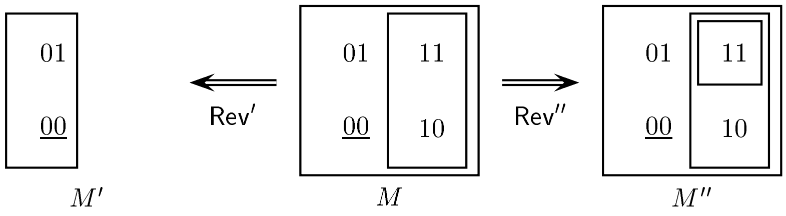


Figure 1: Knowledge revision and belief revision in the light-fan-example. The actual state 00, where light and fan are both off, is underlined. The outermost rectangle encloses all conceivable states, an inner rectangle distinguishes the enclosed states, that are considered more normal, from the rest. Transition Rev' expresses that the agent learns that the light is off, i.e., can definitely exclude states where the light is on. Transition Rev'' expresses that the agent increases her belief in the state where light and fan are both on, but cannot exclude any state from consideration.

$[\mathsf{Rev}'']Bq$ stands for: 'after making situations exceptional where 'light and fan both on' is not true' (Rev''), the operator normally believes (B) that the fan is on (q)'. Both hold in $(M, 00)$. As in dynamic epistemic logic, such modal revision operators can be seen as information state transformers, i.e., as binary relations between information states, as visualized in Figure 1. Unlike in dynamic epistemic logic, the interpretation of revision operators involves relating templated orders corresponding to degrees of belief. Obviously, the dynamic modal operators $[\mathsf{Rev}']$ and $[\mathsf{Rev}'']$ are interpreted relative to some internal structure of 'programmes' Rev' and Rev'', respectively.

Suppose the agent involved here is named a, then Rev' will be described as $L_a?\neg p$, and Rev'' as $*_a(p \wedge q)$. The first stands for: agent a learns that $\neg p$, and the second for 'agent a revises with $p \wedge q$', which corresponds to: agent a makes all states where $p \wedge q$ does not hold an exception.

We start with a Section 2 that supplies motivation for our preferred structures, as visualized in Figure 1, from the viewpoint of standard belief revision semantics. In Section 2 we then properly introduce the notion of a templated equivalence relation. In Section 3 we propose the language for dynamic doxastic logic and define its interpretation. In Section 4 we outline a proof system for the language. We describe some applications in great detail in Section 5.

2 Templated equivalence relations and doxastic epistemic states

2.1 Two traditions in the logic of belief

[This subsection will explain why we can assume '$K45$', or even '$S5$', properties of agents, unlike the more general perspective of templates relative to a state as in Labuschagne's work (to which correspond not necessarily reflexive and/or transitive binary accessibility relations).]

2.2 Doxastic epistemic states

Given are a finite set of *agents* N and a set of *atoms* P.

A *multiagent (Kripke) model* $M = \langle W, R, V \rangle$ consists of a domain W of *worlds* or *factual states*, for each agent $n \in N$ a (at least one) binary *accessibility relation* R_n on W (we will generally write $w \rightarrow_n w'$ instead of $R_n(w, w')$), and a *valuation* $V : P \rightarrow \mathcal{P}(W)$, or in other words: for each atom $p \in P$, a subset V_p of W. Given a model M and a world $w \in M$, (M, w) is called a *multiagent state*. Given a multiagent state (M, w), w is its *point*, and M the model *underlying* it.

If all accessibility relations are equivalence relations, write $w \sim_n w'$ instead of $R_n(w, w')$. We then speak of *epistemic* models and states (or e-model, e-state). If all accessibility relations are transitive and euclidean, we speak of *doxastic* models and states (d-model, d-state).

Set of agents N is the *group* of the model/state, notation: $gr(M) = N$. The group of a set of models/states is the union of the groups of its members. This 'group' or 'set of active agents' is a constantly changing and therefore important parameter when interpreting logical descriptions of knowledge and belief change.

The information states that we propose are a refinement of epistemic states: for each agent, an equivalence relation may contain a transitive and euclidean relation, which may contain another one, etc. An equivalence relation induces a partition on the domain. From the viewpoint of a given world in an equivalence class, that class represents all 'conceivable states/worlds' for that agent. Among those, some may be considered more likely than others. Such preferences are induced by default rules in which they are presented in linguistic (logical) form. This can be captured in the models by a simple *belief function* f mapping each world, given an agent, to a *pair* of indices. The first of pair denotes the equivalence class of the model, and thus relates to the knowledge of that agent. This is essentially just an arbitrary one of an unordered set of labels. The second of the pair corresponds to the degree of belief within a given equivalence class for that agent. The accessibility relations corresponding to agent belief and knowledge are implicit in the belief function but entirely induced by it. In the following, let 0 be the projection on the first of a pair ($0(i, j) = i$), and 1 ditto on the second ($1(i, j) = j$).

Definition 1 (Belief function, templated equivalence, doxastic access)
Given a domain of worlds W, a belief function is a function $f : N \to W \to (\mathbb{N} \times \mathbb{N})$. For $n \in N$, write f_n for $f(n)$. Define

$$\begin{array}{lll} w \sim_n w' & \text{iff} & 0(f_n(w)) = 0(f_n(w')) \\ w \leq_n w' & \text{iff} & 0(f_n(w)) = 0(f_n(w')) \text{ and } 1(f_n(w)) \leq 1(f_n(w')) \\ w \to_n^j w' & \text{iff} & 0(f_n(w)) = 0(f_n(w')) \text{ and } 1(f_n(w')) \leq j \end{array}$$

Par abus de langage, if $1(f_n(w')) \leq j$ we may write $w' \leq_n j$. Relation $\leq_n$ is called a templated equivalence.

All $\sim_n$ are indeed equivalence relations, and will interpret epistemic operators. All $\to_n^i$ are transitive and euclidean, and will interpret belief operators. Let max be the maximum degree of belief, i.e., $\mathsf{max} := \mathsf{MAX}_{w \in W} 1(f_n(w))$, then we also have that $w \sim_n w'$ iff $w \sim_n^{\mathsf{max}} w'$. (This corresponds to knowledge as true belief.) Unlike templates as in [FL02], in the underlying belief functions and corresponding left- and right-linear order $\leq_n$ there is no explicit topmost level of excluded states: these are now implicit, as the domain may now be seen as the disjoint union of a set of total orders.

Unlike in tradional belief revision approaches, the order on states not just induces an order on believed facts but also on arbitrary believed formulas: more specifically and interestingly, on the agent's beliefs about other agents' beliefs.

Definition 2 (Doxastic epistemic model/state (de-state))
A *doxastic epistemic model* (de-model) is a (Kripke) model $M = \langle W, \to, V \rangle$ relative to some belief function f, where for each agent $n \in N$ and for each relevant (i.e. depending on f) degree of belief j, $\to_n^j$ is the binary accessibility relation as induced on W by f. A *doxastic epistemic state* (*de-state*) is a pointed doxastic model.

Example 3
The three different information states in Figure 1 are all de-states. Let 1 be the agent whose beliefs are represented. In $(M, 11)$ we have that $\sim_1$ is the universal relation on the domain $\{00, 01, 10, 11\}$, and that $\{11, 00\} <_1 \{01, 00\}$. In $(M', 11)$ we have that $\sim_1$ is the partition $\{00, 01\}, \{10, 11\}$, where on both classes $\leq_1$ is the identity $(=_1)$. In $(M'', 11)$ we have that $\sim_1$ is the universal relation on the domain $\{00, 01, 10, 11\}$, and that $11 <_1 10 <_1 \{01, 00\}$.

We not just want to express individual knowledge and belief, but also group knowledge and belief. Such notions are standardly interpreted on transitive closures $R_G := (\bigcup_{n \in G} R_n)^+$ of binary relations (or transitive and reflexive closures $R_G := (\bigcup_{n \in G} R_n)^*$). For our structures, in which for different agents we combine different degrees of belief that are not totally ordered, this is somewhat more complex.

Assume that all agents are publicly known to be equally reliable sources of information. Then, when we combine different beliefs, we can simply *add* the degrees of confidence that the different agents have in their beliefs. In terms of the belief function f:

$$1(f_G(w)) := \Sigma_{n\in G} 1(f_n(w))$$

For the first argument of the pair, As far as the possible worlds are concerned from a given one, we can treat this in the standard way: two worlds G-equivalent to each other if there is a finite chain of $\sim_n$ links for agents $n \in G$. Thus we define $\sim_G := (\bigcup_{n\in G} \sim_n)^+$. From our current point of view of belief functions, from which $\sim_n$ is derived, that is an implicit definition. The corresponding explicit definition is as follows: Choose an arbitrary $w \in W$. Define $0(f_G(w) := 0(f_n(w))$ for an arbitrary $n \in G$. (As there are *fewer* equivalence classes for the group, we 'name' it after an arbitrary world w and an arbitrary n-equivalence class contained in it.) For each w' such that there is a chain $0(f_{n_1}(w)) = 0(f_{n_1}(w_1)), 0(f_{n_2}(w_1)) = 0(f_{n_2}(w_2), \cdots, 0(f_{n_t}(w_{t-1})) = 0(f_{n_t}(w')$ with all $n_1, ..., n_t \in G$, define $0(f_G(w') := 0(f_G(w)$ (in other words: we also name it $0(f_n(w)))$. In the remaining worlds of W, choose another world w'' and repeat the procedure, until no worlds are left.

Proposition 4 (Group access)
Accessibility relations $\sim_G$ and $\rightarrow^i_G$ as induced by f_G are, respectively, equivalence relations, and transitive and euclidean relations.

Proof: Obvious. $\dashv$

So far we have not required all 'layers' in the structure to be non-empty. Even if we start with filled layers, certainly in the case of derived belief functions for groups most layers will be empty. Als in the process of interpreting actions in de-states we will compute belief functions with empty layers. We need a simple process of renormalization on the second argument in order to achieve filled layers:

Definition 5 (Renormalization of belief functions)
Let $f_G : W \rightarrow (\mathbb{N} \times \mathbb{N})$ be the belief function for any group of agents G. The *renormalization* $\overline{f}_G$ of f_G is computed as follows: For all $x \in 0(\mathcal{I}mg(f_G))$, list in increasing order all i such that $f_G(w) = (x, i)$: $(x, i_0), \cdots, (x, i_t)$. Then $\overline{f}_G$ is defined by $\overline{f}_G{}^{-1}(x, t) := f_G{}^{-1}(x, i_t)$.

From now on we always assume that introduced belief functions are normal for all individual agents, and that computed belief functions are renormalized for all individual agents.

3 A dynamic doxastic logical language

To a standard multiagent epistemic language with common knowledge for a set N of agents and a set P of atoms [MvdH95, FHMV95], we add dynamic modal operators for programs generically called 'actions' that interpret changes in belief or knowledge. The formulas $\mathcal{L}_N(P)$, the actions $\mathcal{L}^a_N(P)$, and the group *gr* of an action are defined by simultaneous induction:

Definition 6 (Formulas and actions)
The *formulas* $\mathcal{L}_N(P)$ are defined by

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid B^i_G\varphi \mid K_G\varphi \mid [\alpha]\psi$$

where $p \in P$, $n \in N$, $G \subseteq N$, $\alpha \in \mathcal{L}^a_N(P)$, $i \in \mathbb{N}$, and $\psi \in \mathcal{L}_{gr(\alpha)}(P)$. The *actions* $\mathcal{L}^a_N(P)$ are defined by

$$\alpha ::= ?\varphi \mid \ *_G \varphi \mid L_G\beta \mid (\alpha \cup \alpha)$$

where $\varphi \in \mathcal{L}_N(P)$, $G \subseteq N$, $\beta \in \mathcal{L}^a_G(P)$, and where the *group* $gr(\alpha)$ of an action $\alpha \in \mathcal{L}^a_N(P)$ is defined as: $gr(?\varphi) := \emptyset$, $gr(*_G\varphi) := G$, $gr(L_G\alpha) := G$, and $gr(\alpha \cup \alpha') := gr(\alpha) \cap gr(\alpha')$.

Instead of the more familiar notation C_G for the common knowledge (and common belief) operator we now write K_G. In fact, we take common knowledge as primitives and take – permissible within the underlying $S5$ and $K45$ setting – individual knowledge as abbrevations K_n for $K_{\{n\}}$. Formula $B^i_G\varphi$ stands for 'group of agents G commonly believes to degree i that φ', where, again, $B^i_n := B^i_{\{n\}}$ is the individual degree belief for agent n. For the 'most normal' degree of belief 0, we simply write B, i.e.: $B_G := B^0_G$ – common (most normal) belief –, and so $B_n = B^0_n$ – individual belief. The highest degree of belief *is* in fact knowledge, i.e., depending on 'where you are' $K_G := B^{\mathsf{max}}_G$, but for convenience we preferred to introduce knowledge separately from belief.

Other propositional connectives and modal operators are defined by standard abbreviations, in particular $E_G\varphi := \bigwedge_{n\in G} K_n\varphi$. Outermost parentheses of formulas and actions are deleted whenever convenient. As we may generally assume an arbitrary P, write $\mathcal{L}_N$ instead of $\mathcal{L}_N(P)$, and $\mathcal{L}^{\mathsf{act}}_N$ instead of $\mathcal{L}^{\mathsf{act}}_N(P)$. Instead of, e.g., $K_{\{a,b,c\}}$ we always write K_{abc}. For an arbitrary modal ('box'-like) operator B^i_G, write $\overline{B^i_G}$ for its dual ('diamond'). The dual of $[\alpha]$ is $\langle\alpha\rangle$.

The program constructor $*$ in $*_G\varphi$ is called, not surprisingly, the *belief revision* operator. From agent n's point of view, the knowledge base closure for agent n resulting from the operation $*\varphi$ (revision with φ), is approximated as the postconditions of the dynamic operator $[*_n\varphi]$. (There are differences, because we also allow revision about others' beliefs.) The construct $L_G?\varphi$ is pronounced as 'G learn *that* φ' and $*_G\varphi$ is pronounced as 'G revise with φ'. Action $?\varphi$ is a *test*, and $(\alpha \cup \alpha')$ is *nondeterministic choice*.

The group gr was already used for the agents 'occurring' in epistemic states and models. It serves a similar function on actions, whence the overloading. In constructs $[\alpha]\psi$ and $L_G\beta$, the constraint on group gr for the bound subexpressions guarantees that in a de-state for agents N that is the result of action execution, formulas containing modal operators for agents *not* in G are not considered for interpretation.

Example 7
The knowledge and belief descriptions visualized in Figure 1 are indeed exactly as we introduced them. Formula $[L_a?\neg p]K_a\neg p$ describes that after having confirmed that the light is off, operator 1 knows that the light is off. Further note

that though a originally believed the light to be *on*, she now no longer believes that but instead believes that it is off. So a side-effect of the epistemic update is also (pure) belief revision. Formula $[*_a(p \wedge q)]B_a q$ describes that the operator believes that the fan is on, after a belief revision that light and fan are normally on.

3.1 Semantics

A notion of equivalence of de-states is used in Definition 9 of the interpretation of formulas.

Definition 8 (Equivalence of de-states)
Let M, M' be de-states, $w' \in M'$, and $n \in N$. If $n \notin gr(M) \cup gr(M')$, then $(M,w) \sim_n (M',w')$. If $n \in gr(M) \cap gr(M')$, then $(M,w) \sim_n (M,v)$ iff $w \sim_n v$, and $(M,w) \sim_n (M',w')$ iff there is a $v \in M$ such that (M,v) is bisimilar to (M',w') and $(M,w) \sim_n (M,v)$.

Bisimilarity is a notion of sameness between epistemic states that implies equivalence of their logical descriptions (theories), though not vice versa. For a definition, see [BdRV01].

The interpretation of formulas and actions is defined simultaneously.

Definition 9 (Interpretation of formulas and actions)
Let (M,w) be a de-state (for agents N and atoms P), where $M = \langle W, \sim, V\rangle$; let $\varphi \in \mathcal{L}_N(P)$, and let $\alpha \in \mathcal{L}_N^a(P)$. The interpretation $\models$ of φ in (M,w) and the interpretation $[\![\cdot]\!]$ of α in (M,w) are both defined by inductive cases.

$$
\begin{array}{lll}
(M,w) \models p & :\text{iff} & w \in V(p) \\
(M,w) \models \neg\varphi & :\text{iff} & (M,w) \not\models \varphi \\
(M,w) \models \varphi \wedge \psi & :\text{iff} & (M,w) \models \varphi \text{ and } (M,w) \models \psi \\
(M,w) \models B_G^i \varphi & :\text{iff} & \forall w' \in W : w \to_G^i w' \text{ implies } (M,w') \models \varphi \\
(M,w) \models K_G \varphi & :\text{iff} & \forall w' \in W : w \sim_G w' \text{ implies } (M,w') \models \varphi \\
(M,w) \models [\alpha]\varphi & :\text{iff} & \forall (M',w') : (M,w)[\![\alpha]\!](M',w') \Rightarrow (M',w') \models \varphi \\
 & & \\
(M,w)[\![?\varphi]\!](M',w) & :\text{iff} & (M,w) \models \varphi \text{ and } M' = \langle W_\varphi, \emptyset, V|W_\varphi\rangle \quad \text{(see below)} \\
(M,w)[\![*_G\varphi]\!](M^*,w) & & \text{(see below)} \\
(M,w)[\![L_G\alpha]\!](M',w') & :\text{iff} & (M,w)[\![\alpha]\!]w' \text{ and } M' = \langle W', \sim', V'\rangle \qquad \text{(see below)} \\
[\![\alpha \cup \alpha']\!] & := & [\![\alpha]\!] \cup [\![\alpha']\!]
\end{array}
$$

In the clause of action interpretation for 'test': $W_\varphi = \{v \in \mathcal{D}(M) \mid M, v \models \varphi\}$.

In the clause for revision: M^* has the same domain and valuation as M, but associated (renormalized) belief function $\overline{f}'$ (and corresponding induced accessibility) for f' defined as: $0(f'(w)) := 0(f(w))$; if $(M,w) \not\models \varphi$ then $1(f'(w)) := 1(f(w)) + |G|$, else $1(f'(w)) := 1(f(w))$.

In the clause for 'learning': $W' := \{(M'', w'') \mid \exists v \in M : (M, v)[\![\alpha]\!](M'', w'')\}$; for an arbitrary agent n: $(M'', u'') \sim'_n (M'', v'')$ iff $u'' \sim_n v''$ and $(M'', u'') \sim_n (M'', v'')$, where the last $\sim_n$ is equivalence of de-states; and for an arbitrary atom p: $(M'', w'') \in V'_p$ iff $w'' \in V''_p$ (i.e. in M''). These 'new' equivalences are induced by the (first argument of the value of the) following belief function f': introduce an arbitrary assignment of different numbers to different $\sim'_n$ classes as follows: $0(f'_n((M,v))) = 0(f'_n((M,v))) := q$ whenever $(M,v) \sim'_n (M,v')$ in M', for some 'new' $q \in \mathbb{N}$. Further define $1(f'_n((M,v))) := 1(f_n(v))$ if f_n is defined in M (i.e., if $n \in gr(M)$), or else $1(f'_n((M,v))) := 0$. The belief function associated with M' is, of course, not f' itself but $\bar{f}'$.

The notion $\langle\alpha\rangle$ is dual to $[\alpha]$.

Example 10
First we interpret $L_a?\neg p$ in $(M, 00)$. We have that $(M, 00) \models \neg p$. The de-state $(M, 00)[\![?\neg p]\!]$ has underlying model that is the restriction to all worlds where $\neg p$ holds (00 and 01), with empty access. Another world in M where $?\neg p$ can be executed is 01 which leads to de-state $(M, 01)[\![?\neg p]\!]$. These now are the two worlds of the domain of $(M, 00)[\![L_a?\neg p]\!]$. They are a-equivalent because a does not occur in the group (empty!) of either. This leads indeed to the de-state depicted in Figure 1. Part of the interpretation of operator L_a involves recomputing the associated belief function. In this case, normalization is indeed required: in M, we had that (naming the single epistemic class 0) $f_a(00) = f_a(01) = (0,1)$ whereas $f_a(10) = f_a(11) = (0,0)$. Therefore, $1(f_a((M,00)[\![?\neg p]\!])) = 1(f_a(00)) = 1$ and also $1(f_a((M,01)[\![?\neg p]\!])) = 1(f_a(00)) = 1$. However, the most normal level has been eliminated during the interpretation (the actual world was not an expected one). After renormalization, $1(\overline{f}_a((M,00)[\![?\neg p]\!])) = 1(\overline{f}_a((M,01)[\![?\neg p]\!])) = 0$. In other words: worlds 00 and 01 have now become 'most normal'. This completes the computation of the de-state $(M', 00)$.

Observe that formula $[L_a?\neg p]K_a\neg p$ indeed holds in state $(M, 00)$ (because $K_a\neg p$ holds in $(M', 00)$). Further, though in state $(M, 00)$ agent a believed the light to be *on* – $(M, 00) \models B_a p$ –, in state $(M', 00)$ she no longer believes that but instead believes that it is off – $(M', 00) \models B_a\neg p$. So a side-effect of the epistemic update is also (pure) belief revision. The revision of this particular belief is the result of the renormalization of layers in the belief function computed for M'.

Example 11
We now compute the result of $*_a(p \wedge q)$ on $(M, 00)$. All non-models of $p \wedge q$ get a 'higher rating' in their belief function, i.e., from $f_a(00) = f_a(01) = (0,1)$ and $f_a(10) = f_a(11) = (0,0)$ we derive f'_a with $f'_a(00) = f'_a(01) = (0,2)$, $f'_a(10) = (0,1)$ and $f'_a(01) = (0,0)$. Normalization is not required in this case. State $(M'', 00)$ indeed results.

Formula $[*_a(p \wedge q)]B_a q$ holds in state $(M, 00)$. Before this revision, a did *not* believe that the fan was on: $(M, 00) \models \neg B_a q$. This is because 10 is $<_a$-minimal in M, so from $(M, 10) \models \neg q$ follows $(M, 10) \models \overline{B^0_a}\neg q$, i.e., $(M, 10) \models \neg B^0_a q$. So a has indeed revised her beliefs.

After the revision, she believes ($B_a = B_a^0$) that light and fan are both on – $(M'', 00) \models B_a^0(p \wedge q)$ –, she doesn't strongly believe that, but she *does* strongly believe that the light is on – $(M'', 00) \models B_a^1 p$. Last but not least, she *still* hasn't ruled out that the light may be off – $(M'', 00) \models \overline{B_a^2} \neg p$.

Standardly, validity of formulas is defined as follows: $M \models \varphi$ iff for all $w \in M$, $(M, w) \models \varphi$; 'φ is valid', notation $\models \varphi$, if for all M, $M \models \varphi$.

We now mention some relevant validities. These are far from comprehensive, but merely jumping grounds towards a as yet incomplete axiomatization of the logic.

Common and individual belief Most normal common belief implies most normal individual belief. For arbitrary $n \in G$:

$$B_G\varphi \to B_n\varphi$$

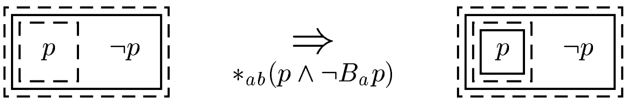
Meaning: B_G^0, and B_n^0. For other levels of belief, common belief typically does not imply individual belief.

Unsuccessful belief updates An unsuccessful *epistemic* update is an epistemic formula that after utterance is no longer true (relative to a given epistemic state). A typical example is the utterance 'you don't know that p' for some true fact p, after which you *do* know that p, so the utterance cannot be made another time, so to speak. Something similar holds for belief revision! In an information state where φ is consistent with the beliefs of the agents, it may be false that φ after a belief revision $*\varphi$. (We expect this to be a surprising statement for the traditional belief revision community.) An example:

Consider a de-state for two agents a and b and about some fact p, where neither a nor b are knowledgable about p, but where b believes p and a does not believe p. Suppose a revises with b telling her that p and that a does not believe that. The revision formula is $p \wedge \neg B_a p$. In the current de-state, this formula indeed holds. Execution of $*_{ab}(p \wedge \neg B_a p)$ results in a de-state where both a and b believe that p. (An – equivalent – alternative description would be $L_{ab} *_a (p \wedge \neg B_a p)$. The two different descriptions suggest that we may at some future stage choose our primitives differently.) As $B_a p$ entails $\neg(p \wedge \neg B_a p)$, the negation of the revision formula therefore holds in that state. In other words, in the original state it holds that

$$[*_{ab}\,(p \wedge \neg B_a p)]\neg(p \wedge \neg B_a p)$$

The revision executed is visualized below. Access for a is solid, access for b is dashed.

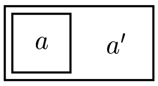


The more interesting validities involve the postconditions of belief revision. As the unsuccessful belief revision example shows, this is not trivial.

Action models Apart from the belief revision operator, the chosen dynamic epistemic operators form the 'descriptive core' of the dynamic epistemic language in [vD02]. The revision operator appears to be quite adaptable to other dynamic epistemic languages, e.g. to the very appealing action language of [BMS03], where actions are interpreted as action models. Action models are relational structures that are quite similar to information states, but where the worlds of the structure are characterized not by valuations of atoms but by preconditions for action execution. Worlds in the product model that represents the result of executing an action in an information state have the form (w, a), where w is a world in the model wherein action a, a point in the action model, can be executed.

We can extend this semantics to include degrees of belief by associating belief functions with action models in the exact same way as they define de-models. One may then compute the 'degree of belief after action execution as follows: define $f((w, a)) := f(w) + f(a)$, i.e., the degree of belief in the next state is a function (the sum) of the degree of belief in the chosen state w and ditto in the chosen action a.

The action model corresponding to a revision by G with φ consists of two points, say a and a', with $pre(a) = \varphi$ and $pre(a') = \neg\varphi$, and associated belief function that induces for any agent $n \in G$ that $a \sim_n a'$ and $a <_n a'$ (i.e. $f_n(a) = (0, 0)$ and $f_n(a') = (0, 1)$). Such an action model can be visualized as below:



4 Proof system for dynamic doxastic logic

We do not have a proof system yet. This will be based on the complete proof systems for similar languages as provided in [BMS03, vDvdHK03].

5 Applications

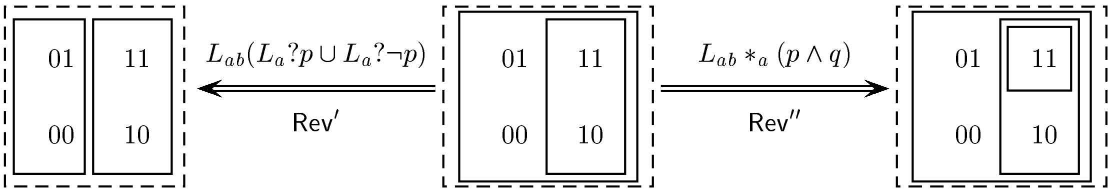
5.1 The light-and-fan system

The scenario is as outlined in the introduction, only now for more than one agent:

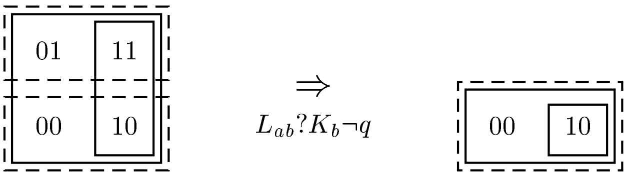
Two operators Anne and Bert monitor a control system that consists of a light and a fan. Both can be on and off. The operators use factual knowledge and defeasible beliefs to reason about the system. Defeasible beliefs are based on default rules, such as 'the light is normally on'.

In all figures, access for Anne (a) is solid and for Bert (b) is dashed. We make no assumptions about the actual world, so do not underline such worlds in the pictures.

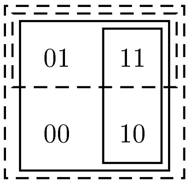
Example 12
First we repeat the example from the introduction for two agents, assume that Bert is known to be totally ignorant, and that action Rev' more precisely (for two agents) describes that Anne checks *whether* (not *that*) the light is on. The following transitions then take place:



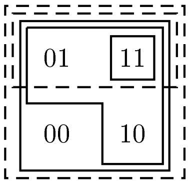
Example 13
We start in an information state where it is commonly known that operator b knows whether the fan is on, and operator a believes that the light is on. After operator b says that the fan is off, operator a (still) believes that the light is on. This is described by $[L_{ab}?\neg q]B_a p$.



Example 14
We start in an information state where it is commonly known that operator a believes that the light is on, whereas operator b believes that the fan is on. Operator a says: "I find your beliefs quite reasonable, b; I will share your confidence in your beliefs." This is the action $L_{ab} *_a q$ $(*_a bq)$. After this action, Anne believes that light and fan are both on: $B^0_a(p \wedge q)$. The corresponding de-state transition is:

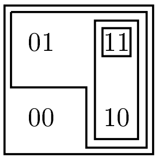


$\Rightarrow \quad L_{ab} *_a q$

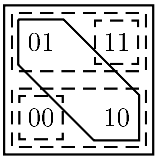


In the resulting state, worlds 01 and 10 are now considered equally likely by Anne, whereas before, she thought 10 to be more likely. A somewhat different scenario is the one wherein Anne and Bert *pool* their beliefs: $*_{ab}q$ and after that $*_{ab}p$ (we have not added a concurrency operator, so have to model this as sequential execution). The resulting model is then as above, but with access for Bert similar to the access for Anne in that model (incidentally, even before executing these actions, this access was already induced by their 'group' belief function f_{12}).

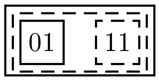
Yet another information state can be reached if Anne decides to take over Bert's belief in q, but merely as a refinement on her current beliefs. More precisely, whenever p, q is preferred, and whenever $\neg p$, q is also preferred. This corresponds to two (sequential) belief revisions $*_a(p \to q)$ and $*_a(\neg p \to q)$. In that case, the following model results:



Example 15
We start in a state where it is commonly known that Anne believes that light and fan are not both on nor both off, and where Bert knows whether the fan is on, and if so believes that the light is on too, but if not believes that the light is off too. The following conversation now takes place: Bert: "I believe that the light is on." Anne: "Actually, I think you are wrong, because I now believe that the light is off." Bert: "Sure, I already knew that."



$\Rightarrow \quad L_{ab}?B^0_b p$



Bert's announcement that he believes that the light is on can only be successfully executed in states 01 and 11 (because in the bottom two states, he believes that it's off). In the resulting state, on the right, whatever the actual state, it indeed holds that $B_a\neg p$ and K_aB_bp, and B_bp and $K_bB_a\neg p$.

5.2 How to get to the railway station

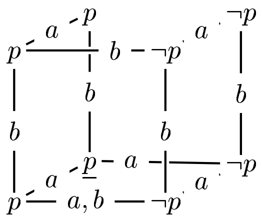
We now present a variation on a famous example that explains how various pragmatic constraints can be semantically modelled and that originates with Johan van Benthem [vBDvE$^+$02]: two strangers meet on the road, and one asks the other the way to the railway station. The point of this communication is that by asking the question, that agent declares his ignorance of the fact asked for, which results in an (implicit) update, even before the answer is given by the responding agent.

Our analysis of this example is more involved, and we hope to convince the reader of its feasibility. First we present the scenario in more detail:

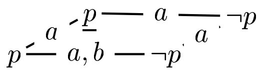
> *Anne approaches Bert on the street. She addresses him with "Good afternoon, sir." – Bert nods in assent –* (i) *and continues with "Would you by any chance know the way to the railway station?"* (ii) *Bert responds with "Certainly, madam, I can help you out."* (iii) *and continues with "First to the left that way, and then second to the right. You can't miss it."* (iv) *Anne concludes with "Thank you very much, sir" and continues her way in the pointed direction.* (v)

Obviously, these people aren't from New Zealand, as in that country the – by now – generic term 'mate' appears to have replaced both 'sir' and 'madam' in everyday conversation. What other meaning in terms of information exchange can we give to this scenario?

Although the 'approach on the street' may be given some meaning, e.g. in an ominous sense, we take as a basic assumption that the communicating agents are, so to speak, behaving in a maximally cooperative and communicative way, according to the Gricean maxims. Therefore, Anne's combined approach and introduction (i) suggests a reason for that contact. More precisely, (i) *raises the issue*, as yet unknown. Let's choose the absolute simplest solution to that: we call the issue the truth value of some fact p. However, even more information than that, is already contained in (i): because Anne doesn't call Bert by his name, and because of Bert's implicit confirmation of that, Anne and Bert confirm that they are strangers to each other. In this way they have not merely raised an issue p, but also publicly declared ignorance about each other, i.e. both are now commonly known be ignorant about the other's knowledge and beliefs. This entails that all ignorance is about the issue (and that this is publicly known). The following model represents such maximal factual ignorance; instead of boxes representing equivalence classes, in this visualization worlds in the same n-class are linked by a chain of n-connections (e.g., the 'bottom' four worlds are in the same a-class):



Anne continues by asking Bert a question (*ii*). Now this is not immediately obvious at the outset. E.g., she might have approached Bert with the intention of supplying him with information on the issue, e.g. handing him a brochure on Greenpeace or on an upcoming garage sale in town. Therefore, on Gricean assumption, by (*ii*) she declares her ignorance on the issue. This results in the elimination of all worlds where she is knowledgable and the following model results:



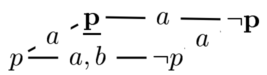
Bert further reduces the options with his response (*iii*), which reveals him as knowledgeable about p (another and perfectly acceptable answer would have been "No, I'm actually a stranger myself in this town" which would have resulted in a different information state):

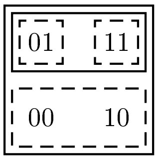
$$\underline{p} \overset{a}{\text{———}} \neg p$$

And he concludes the scenario by declaring that p:

$$\underline{p}$$

So far, we have only modelled the agents' knowledge and not their beliefs. Some defaults have actually been used in the course of this conversation. The most obvious one is that by asking Bert a question, Anne considers it not merely possible but also likely that he knows the answer. This results in, the following (in tentative visualization above, to show the correspondence – more normal worlds for Anne are in bold), and using the more familiar visualization below):





One might suggest that by asking the question, Anne does not actually declare ignorance but merely makes it public that she is more likely to be knowledgable about p than to be ignorant about it. However, this would contradict the informative goals of that communication and thus the Gricean assumptions. We therefore have correctly modelled this as a confirmation of ignorance.

At a previous state in the scenario however, such beliefs about her knowledge of p still appear to be relevant, and also something corresponding to the default knowledge assumed in Bert. The most common reason to approach a stranger on the street is to ask a question, and not to inform him or her, and to assume that this stranger is more likely to be able to answer the question than not to answer it. These two default assumptions can be modelled in the initial 'cube of factual ignorance'. A 'proper' visualization would involve blocks containing smaller blocks, etc. Suffices to point out that for Anne, from the bottom four worlds the back two are preferred (namely the ones where Bert knows p) – this is as above –, whereas for Bert, from the front four worlds the bottom two are preferred (where Anne is ignorant), from the left-back two worlds the bottom one is preferred (where Anne is ignorant), and from the right-back two worlds also the bottom one is preferred.

From this information state, which is a de-state, the de-state pictured above results from (*ii*). The results of (*iii*) and (*iv*) in terms of de-states are similar to those already pictured before: no defaults play a role in the remainder of the conversation. This concludes the analysis of the example.

Essentially, what we claim to have done by this careful analysis, is to show that pragmatic assumptions do not exist as separate from the semantic level but merely require a more advanced semantics.

6 Conclusion

[to be added later]

References

[AGM85] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[Bal02] A. Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54(1):1–45, 2002.

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001. Cambridge Tracts in Theoretical Computer Science 53.

[BMS98] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements and common knowledge. Proceedings of TARK 98, 1998.

[BMS03] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. Manuscript, originally presented at TARK 98, 2003.

[FHLvdW03] D. Ferguson, J. Heidema, W.A. Labuschagne, and L. van der Westhuizen. Information templates. Manuscript, 2003.

[FHMV95] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge MA, 1995.

[FL02] D. Ferguson and W.A. Labuschagne. Information-theoretic semantics for epistemic logic. In *Proceedings of LOFT 5*, Turin, Italy, 2002. ICER. no page numbers.

[Ger99] J.D. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, University of Amsterdam, 1999. ILLC Dissertation Series DS-1999-01.

[GG97] J.D. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.

[KLM90] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Mey01] T.A. Meyer. Basic infobase change. *Studia Logica*, 67:215–242, 2001.

[MLH00a] T.A. Meyer, W.A. Labuschagne, and J. Heidema. Infobase change: A first approximation. *Journal of Logic, Language, and Information*, 9:353–377, 2000.

[MLH00b] T.A. Meyer, W.A. Labuschagne, and J. Heidema. Refined epistemic entrenchment. *Journal of Logic, Language, and Information*, 9:237–259, 2000.

[MvdH95] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge, 1995.

[Pla89] J.A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.

[Seg99] K. Segerberg. Two traditions in the logic of belief: bringing them together. In H.J. Ohlbach and U. Reyle, editors, *Logic, Language, and Reasoning*, pages 135–147, Dordrecht, 1999. Kluwer Academic Publishers.

[tC02] B. ten Cate. Internalizing epistemic actions. In Maricarmen Martinez, editor, *Proceedings of the NASSLLI-2002 student session*, Stanford University, 2002.

[vB96] J.F.A.K. van Benthem. *Exploring logical dynamics*. CSLI Publications, 1996.

[vBDvE+02] J.F.A.K. van Benthem, P. Dekker, J. van Eijck, M. de Rijke, and Y. Venema. *Logic in Action*. ILLC, Amsterdam, 2002.

[vD00] H.P. van Ditmarsch. *Knowledge games*. PhD thesis, University of Groningen, 2000. ILLC Dissertation Series DS-2000-06.

[vD02] H.P. van Ditmarsch. Descriptions of game actions. *Journal of Logic, Language and Information*, 11:349–365, 2002.

[vDvdHK03] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Concurrent dynamic epistemic logic. In V.F. Hendricks, K.F. Jorgensen, and S.A. Pedersen, editors, *Knowledge – Foundations and Applications*, Dordrecht, 2003. Kluwer Academic Publishers. To appear.