To appear in Erkenntnis.

DDL UNLIMITED: DYNAMIC DOXASTIC LOGIC FOR INTROSPECTIVE AGENTS*

Sten Lindström and Wlodek Rabinowicz

The theories of belief change developed within the AGM-tradition are not logics in the proper sense, but rather informal axiomatic theories of belief change. Instead of characterizing the models of belief and belief change in a formalized object language, the AGM-approach uses a natural language — ordinary mathematical English — to characterize the mathematical structures that are under study. Recently, however, various authors such as Johan van Benthem and Maarten de Rijke have suggested representing doxastic change within a formal logical language: a dynamic modal logic.[1] Inspired by these suggestions Krister Segerberg has developed a very general logical framework for reasoning about doxastic change: *dynamic doxastic logic* (DDL).[2] This framework may be seen as an extension of standard Hintikka-style doxastic logic (Hintikka 1962) with dynamic operators representing various kinds of transformations of the agent's doxastic state.

Basic DDL describes an agent that has opinions about the external world and an ability to change these opinions in the light of new information. Such an agent is *non-introspective* in the sense that he lacks opinions about his own belief states. Here we are going to discuss various possibilities for developing a dynamic doxastic logic for *introspective agents: full DDL* or *DDL unlimited*. The project of constructing such a logic is faced with difficulties due to the fact that the agent's own doxastic state now becomes a part of the reality that he is trying to explore: when an introspective agent learns more about the world, then the reality he holds beliefs about

---

* Many of the ideas of this paper have been presented at various seminars, workshops and conferences in Uppsala, Umeå, Lund, Stockholm, Pittsburgh, Aix-en-Provence, Warsaw, Paris and Konstanz. We are grateful to the participants for their comments and friendly criticism. This paper would not have come into existence, however, were it not for the Workshop on Coherence and Dynamics of Belief – Konstanz, February 5-7, 1998. For this we owe special thanks to the organizers Wolfgang Spohn, Volker Halbach and Erik Olsson. Intellectually and personally, our greatest debt is to Krister Segerberg. We are grateful to him as well as to John Cantwell, Sven Ove Hansson and Tor Sandqvist for inspiration and advice. We also wish to thank an anonymous referee for helpful comments. Some of the material presented here has appeared earlier in Lindström and Rabinowicz (1997).

[1] Cf. de Rijke (1994).

[2] Cf. Segerberg (1995a), (1995b), (1996a), (1996b) and (1997).

undergoes a change. But then his introspective (higher-order) beliefs have to be adjusted accordingly. In the paper we shall consider various ways of solving this problem.

1. Background: AGM and LR

In a doxastic logic of Hintikka-type, with a modal operator **B** standing for "the agent believes that", it is possible to represent and reason about the *static* aspects of an agent's beliefs about the world. Such a logic studies various constraints that a rational agent or a set of rational agents should satisfy. A Hintikka-type logic cannot, however, be used to reason about doxastic change, i.e., various kinds of *doxastic actions* that an agent may perform. The agent may, for instance, *revise* his beliefs by adding a new piece of information, while at the same time making adjustments to his stock of beliefs in order to preserve consistency. Or he may *contract* his beliefs by giving up a proposition that he formerly believed. Such operations of doxastic change are studied in the theories of *rational belief change* that started with the work of Alchourrón, Gärdenfors and Makinson in the 80's: the so-called *AGM-approach*.[3] According to AGM, there are three basic types of doxastic actions:

> *Expansion:* The agent adds a new belief α to his stock of old beliefs without giving up any old beliefs. If G is the set of old beliefs, then $G+\alpha$ denotes the set of beliefs that results from *expanding* G with α. To expand is dangerous, since $G+\alpha$ might very well be logically inconsistent; and inconsistency is something that we should try to avoid in our beliefs.
>
> *Contraction:* The agent gives up a proposition α that was formerly believed. This requires that he also gives up other propositions that *logically imply* the proposition α. We use $G–\alpha$ to denote the result of contracting α from the old set G of beliefs.
>
> *Revision:* The revision $G*\alpha$ of the set G with the new information α is the result of adding α to G in such a way that consistency is preserved whenever possible. The idea is that $G*\alpha$ should be a set of beliefs that preserves as much as possible of the information that is contained in G and still contains α. $G*\alpha$ should be a minimal change of G that incorporates α.

The following is an important guiding principle when revising and contracting belief sets:

> *The Principle of Conservatism*: Try not to give up or add information to your original belief set unnecessarily.

---

[3] Cf. Alchourrón, Gärdenfors, Makinson (1985) and Gärdenfors (1988).

Within the AGM approach, the agent's belief state is represented by his *belief set*, i.e., the set G of all sentences α such that the agent believes that α. An underlying classical consequence operation Cn is assumed and the operation of expansion + is defined by

$$G+\alpha = Cn(G \cup \{\alpha\}).$$

By contrast, the operations of contraction and revision are characterized only axiomatically. Thus, the operation of revision is assumed to satisfy the axioms:

(R1)	$Cn(G) = G$	*(Logical Closure)*
(R2)	$\alpha \in G{*}\alpha$	*(Success)*
(R3)	$G{*}\alpha \subseteq G+\alpha$	*(Inclusion)*
(R4)	if $\neg\alpha \notin G$, then $G \subseteq G{*}\alpha$	*(Preservation)*
(R5)	if $\bot \notin Cn(\{\alpha\})$, then $\bot \notin G{*}\alpha$	*(Consistency)*
(R6)	if $Cn(\{\alpha\}) = Cn(\{\beta\})$, then $G{*}\alpha = G{*}\beta$	*(Congruence)*
(R7)	$G{*}(\alpha \wedge \beta) \subseteq (G * \alpha)+\beta$	
(R8)	if $\neg\, \beta \notin G{*}\alpha$, then $(G * \alpha) \subseteq G{*}(\alpha \wedge \beta)$.	

The first four axioms imply:

if $\neg\alpha \notin G$, then $G{*}\alpha = G+\alpha$, *(Expansion)*

i.e., if the new information α is consistent with G, then $G{*}\alpha$ is simply the expansion of G with α. Consistency says that if α is consistent, then $G{*}\alpha$ is also consistent. According to Congruence, if α and β are logically equivalent, then revising G with α yields the same result as revising G with β. In view of (R1) and (R2), the last two axioms yield:

$\neg\, \beta \notin G{*}\alpha$, then $G{*}(\alpha \wedge \beta) = (G{*}\alpha)+\beta$ *(Revision by Conjunction)*

i.e., if β is consistent with $G{*}\alpha$, then revising G with $(\alpha \wedge \beta)$ yields the same result as first revising G with α and then expanding the result with β.

AGM also contains axioms for *contraction* (omitted here) as well as the following bridging principles:

$G{*}\alpha = (G{-}\neg\alpha)+\alpha$ *(The Levi identity)*

$G{-}\alpha = (G{*}\alpha) \cap (G{*}\neg\alpha)$ *(The Harper identity)*

The Levi identity says that the result of revising the belief set G by the sentence α equals the result of first making room for α by (if necessary) contracting G with $\neg\alpha$ and then expanding the result with α. The Harper identity says that the result of contracting α from G is the common part of G revised with α and G revised with $\neg\alpha$.

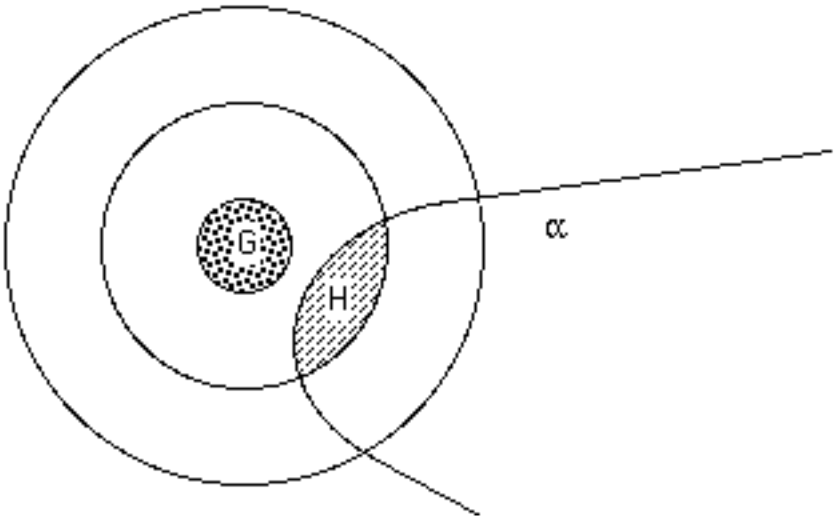
In his (1988) paper, Grove presents two closely related possible worlds modellings of AGM-type belief revision, one in terms of a family of "spheres" around the agent's belief set (or theory) G and the other in terms of an epistemic entrenchment ordering of propositions.[4] Intuitively, a proposition α is at least as entrenched in the agent's belief set as another proposition β if and only if the following holds: provided the agent would have to revise his beliefs so as to falsify the conjunction $\alpha \wedge \beta$, he should do it in such a way as to allow for the falsity of β.

Grove's spheres may be thought of as possible "fallback" theories relative to the agent's original theory: theories that he may reach by deleting propositions that are not "sufficiently" entrenched (according to standards of sufficient entrenchment of varying stringency). To put it differently, fallbacks are theories that are closed upwards under entrenchment: if T is a fallback, α belongs to T, and β is at least as entrenched as α, then β also belongs to T. The entrenchment ordering can be recovered from the family of fallbacks by the definition: α is at least as entrenched as β if and only if α belongs to every fallback to which β belongs.

Representing propositions as sets of possible worlds, and also representing theories as such sets (rather than as sets of propositions), the following picture illustrates Grove's family of spheres around a given theory G and his definition of revision. Notice that the spheres around a theory are "nested", i.e., linearly ordered. For any two spheres, one is included in the other. Grove's family of spheres closely resembles Lewis' sphere semantics for counterfactuals, the main difference being that Lewis' spheres are "centered" around a single world instead of a theory (a set of worlds).

---

[4] Actually, Grove works with an ordering of epistemic *plausibility*. But as Gärdenfors (1988, sect. 4.8) points out, the notions of plausibility and entrenchment are interdefinable. Thus, a proposition α is at least as plausible as a proposition β given the agent's beliefs if and only if non-β is at least as entrenched as non-α in the agent's belief set. The notion of epistemic entrenchment is primarily defined for the propositions that belong to the agent's belief set: one adopts the convention that propositions that are not believed by the agent are minimally entrenched. On the other hand, the notion of plausibility primarily applies to the propositions that are incompatible with the agent's beliefs (the propositions that are compatible with what he believes are all taken to be equally and maximally plausible). Thus, this is a notion of *conditional* plausibility. α is at least as plausible as β in this sense iff the following holds: on the condition that I would have to revise my beliefs with $\alpha \vee \beta$, I should change them in such a way as to allow for α.



The shaded area H represents the revision of G with a proposition α. The revision of G with α is defined as the strongest α-permitting fallback theory of G expanded with α. In the possible worlds representation, this is the intersection of α with the smallest sphere around G that is compatible with α. (Any revision has to contain the proposition we revise with. Therefore, if α is logically inconsistent, the revision with α is taken to be the inconsistent theory.)

In a series of papers, we have proposed a generalization *LR* of the AGM approach according to which belief revision was treated as a *relation* $GR_{\alpha}H$ between theories (belief sets) rather than as a function on theories.[5] The idea was to allow for there being several equally reasonable revisions of a theory with a given proposition. Thus, $GR_{\alpha}H$ means that H is one of those reasonable revisions of the theory G with the new information α. AGM, of course, assumes that belief revision is functional (or deterministic), that is,

$$\text{if } GR_{\alpha}H \text{ and } GR_{\alpha}H', \text{ then } H = H'.$$

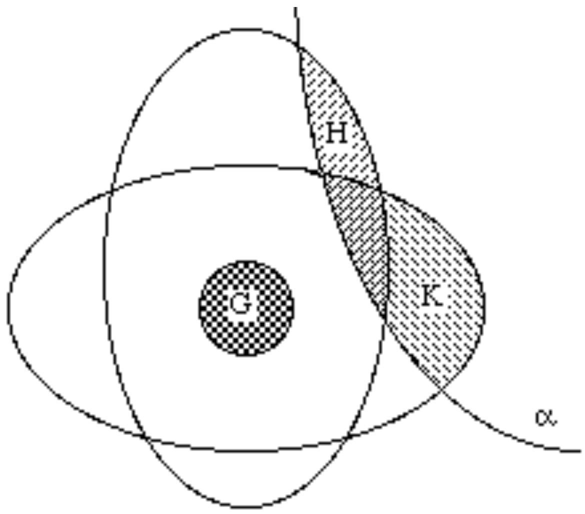
Given this assumption, one can define:

$$G{*}\alpha = \text{the theory H such that } GR_{\alpha}H.$$

The relational notion of belief revision results from weakening epistemic entrenchment by not assuming it to be *connected.* In other words, we allow that some propositions may be incomparable with respect to epistemic entrenchment. As a result, in LR the family of fallbacks around a given theory will no longer be nested. It will no longer be a family of spheres but rather a family of "ellipses". This change opens up for the possibility of several different ways of revising a theory with a given proposition.

---

[5] Cf. Lindström and Rabinowicz (1989), (1990), (1992) and Rabinowicz and Lindström (1994).



In this figure, the two ellipses represent two different fallback theories for G, each of which is a strongest α-permitting fallback. Consequently, there are two possible revisions of G with α: each one of H and K is the intersection of α with a strongest α-permitting fallback.

2. Hypertheories, topology and the problem of iterated belief change

Above we have assumed that for any consistent proposition α, there always exists at least one strongest α-permitting fallback. Without making further assumptions, we have no guarantee, however, that this will always be the case. If α is incompatible with G, then for every α-permitting fallback, there may exist a stronger α-permitting fallback in the fallback family for G. Normally, this possibility is excluded by imposing some form of "limit assumption" on fallback families. We shall, however, deal with this problem in another way.

We are using a modelling in which propositions and theories are represented by sets of points ("possible worlds") in some underlying space U. The set **Prop** of propositions forms a Boolean set algebra. Theories are represented by so-called *closed sets*, i.e., arbitrary intersections of propositions. It is easily shown that any intersection of a family of closed sets is itself closed; and that the same applies to any finite union of closed sets. This means that the family **C** of all closed sets determines a *topology* **T** over U that consists of all the *open* sets, i.e., the subsets of U that are complements of the sets in **C**. In this topology, the propositions are the *clopen* sets, i.e., the sets that are both closed and open. Since propositions are thought to correspond to sentences in the object language and since the underlying logic is taken to be compact, we impose the corresponding compactness condition on the topology.

> Any family of closed sets with an empty intersection includes a finite subfamily that also has an empty intersection. *(Compactness)*

In the following, we refer to families of fallbacks as "hypertheories" (a term introduced by Segerberg). Formally, we define a *hypertheory* as any family **H** of closed sets such that (i) $\cap\mathbf{H}$

$\in$ **H**; and (ii) U $\in$ **H**. The set $\cap$**H** represents the beliefs of the agent, i.e., he believes a proposition P if and only if $\cap$**H** $\subseteq$ P.

Suppose that the agent wants to revise his beliefs with a non-empty proposition P. Since, U $\in$ **H**, **H** contains at least one P-permitting fallback X. If X includes a strongest P-permitting fallback Y, then Y $\cap$ P is a possible result of the revision. But consider the case when there is an infinitely descending chain starting with X of stronger and stronger P-permitting fallbacks. Such a chain can always be extended to a *maximal* chain **K** of that kind (by the so-called Hausdorff's maximal principle, which is equivalent to the axiom of choice). It may well be the case that the "limit" for **K**, i.e., $\cap$**K**, lies outside of **H**. Still, it can then be shown, given Compactness, that this limit is itself P-permitting. We now propose to use such limits of maximal P-permitting chains for the purpose of revision even in those cases when they lie outside the hypertheory itself.

Let **H** be a hypertheory with intersection X (representing an agent's "belief set" G). Consider any proposition P. We say that Y is a *possible revision* of X with P if and only if either (i) P = $\varnothing$ and X = $\varnothing$; or (ii) P $\neq$ $\varnothing$ and there exists a maximal chain **K** of P-permitting fallbacks in **H** and Y = ($\cap$**K**) $\cap$ P

It is easy to verify that any such possible revision Y is a closed set. To show that Y is non-empty if P is non-empty, we need to show that $\cap$**K** in the second clause of the definition is P-permitting. Suppose that it is not, i.e., that ($\cap$**K**) $\cap$ P = $\varnothing$. Then, by Compactness, there is a finite subset **K**' of **K** such that ($\cap$**K**') $\cap$ P = $\varnothing$. But then, let Z be the minimal element of **K**'. Since **K**' is a chain, Z = $\cap$**K**', so that Z $\cap$ P = $\varnothing$, contrary to the hypothesis.

This way of constructing revision has an advantage: we don't need to impose strong limit assumptions on hypertheories. Having such a very general notion of a hypertheory, that satisfies just a few conditions, is helpful when it comes to constructing new hypertheories out of old ones.

A belief state of the agent, as given by a hypertheory, specifies both his beliefs and his dispositions for belief change. The construction of revision that we have presented yields for a given hypertheory and a proposition the new beliefs of the agent, but it does not tell us anything about his new dispositions for belief change. We know what his new theory may look like, but what about his new hypertheory? Until we have answered this question, we cannot say anything interesting about iterated belief change.

Segerberg (1997) has made some tentative proposals about how new hypertheories can be constructed out of old ones. We get the following recipes for contraction and expansion if we adjust his proposals to our present construction in which we are allowed to move out of **H** in the search for limits for P-permitting chains:

We define a *P-permitting limit* with respect to H to be any intersection $\cap$**K** of a maximal chain **K** of P-permitting fallbacks in **H**. As we have seen, such a P-permitting limit is indeed P-permitting and is, of course, the greatest lower bound for all the P-permitting fallbacks in **K**. A limit of this kind may sometimes be a member of **H** itself, in which case it is a genuine fallback, or it may lie outside.

We can now define contraction and expansion of a hypertheory in such a way that the result is a new hypertheory:

> *Contraction* of a hypertheory **H** with P: If P $\neq$ U, take any minimal –P-permitting limit S with respect to **H** and let the new hypertheory consist of S together with all those fallbacks in **H** that include S. If P = U, let the new hypertheory be **H** itself (necessary propositions cannot be given up).
>
> *Expansion* of a hypertheory **H** with P: Take as the new hypertheory all the old fallbacks in **H** together with all their intersections with P.

According to the definitions, expansion is functional while contraction is relational. *Revision* is defined via the Levi-identity: a revision of **H** with P is any hypertheory obtained by first contracting **H** with –P and then expanding with P.

3. *Dynamic doxastic logic*

Up to now we have been concerned with the semantic modelling of belief change. Let us now turn our attention to the object language and its logic. Segerberg's basic DDL consists of a static part – the logic for the belief operator **B** – and a dynamic part: the logic for the dynamic doxastic operators. The latter represent various kinds of transformations of the agent's doxastic state. Segerberg writes $+\alpha$, $*\alpha$, and $-\alpha$, respectively, for the *doxastic actions* of *expanding*, *revising* and *contracting* the agent's beliefs with the sentence α. To different doxastic actions correspond different doxastic operators:

$[+\alpha]\beta$	"If the agent were to *expand* his beliefs with α, then it would be the case that β".
$[*\alpha]\beta$	"If the agent were to *revise* his beliefs with α, then it would be the case that β".
$[-\alpha]\beta$	"If the agent were to *contract* his beliefs with α, then it would be the case that β".

In basic DDL the points in the space U represent different states of the (external) world, where the agent's beliefs and doxastic dispositions are not considered to belong the world that he has beliefs about. Consequently, doxastic actions of the agent do not affect the world. Thus, if β expresses a *worldly proposition*, i.e., a proposition that only concerns the (external) world, then a doxastic action does not influence its truth-value. In basic DDL it is assumed that atomic sentences express worldly propositions Consequently, the same applies to all Boolean formulas (i.e., formulas that are built up from atomic sentences by means of the standard Boolean connectives). Therefore, we have

$$[\tau]\beta \leftrightarrow \beta,$$

for every Boolean formula β and every doxastic action τ.

So the interesting case is the one in which β contains doxastic operators. In particular, we are interested in statements of the forms: $[\tau]\mathbf{B}\beta$.

For example,

$$[*\alpha]\mathbf{B}\beta$$

means: if the agent were to revise his beliefs with α, he would believe β.

DDL allows for the possibility of belief change being *nondeterministic*: in accordance with the LR-approach, there may be many different ways for the agent of revising his beliefs with α . Hence, we must distinguish between:

$[*\alpha]\mathbf{B}\beta$ "If the agent were to revise his beliefs with α, he *would* believe that β".

$<*\alpha>\mathbf{B}\beta$ "If the agent were to revise his beliefs with α, he *might* believe that β".

$<*\alpha>$ is definable in terms of $[*\alpha]$ in the standard way:

$$<*\alpha>\beta = \neg[*\alpha]\neg\beta.$$

In the same way, one can define $<\tau>$ for any doxastic operator $[\tau]$. For theories like the original AGM-theory in which belief change is deterministic, one would have $<*\alpha>\beta \leftrightarrow [*\alpha]\beta$, and similarly for contraction. Expansion, is of course always deterministic, i.e., $<+\alpha>\beta \leftrightarrow [+\alpha]\beta$.

The object language for basic DDL can be described as follows. We define the sets **Term**, **BForm** and **Form** of *terms*, *Boolean formulas* and *formulas* to be the smallest sets satisfying the following conditions:

(i) for any $n < \omega$, the propositional letter P_n belongs to **BForm**

(ii) $\perp \in$ **BForm**

(iii) if $\alpha, \beta \in$ **BForm**, then $(\alpha \rightarrow \beta) \in$ **BForm**

(iv) if $\alpha, \beta \in$ **Form**, then $(\alpha \rightarrow \beta) \in$ **Form**

(v) if $\alpha \in$ **BForm**, then $\alpha \in$ **Form**

(vi) if $\alpha \in$ **BForm**, then $\mathbf{B}\alpha \in$ **Form**.

(vii) if $\alpha \in$ **BForm**, then $+\alpha, -\alpha, *\alpha \in$ **Term**.

(viii) if $\tau \in$ **Term** and $\alpha \in$ **Form**, then $[\tau]\alpha \in$ **Form**.

The Boolean connectives $\neg\alpha$, $(\alpha \wedge \beta)$, etc. are defined from $\perp$ and $\rightarrow$ in the usual way.

As is easily seen, basic DDL is severely limited in its expressive power. To begin with, the belief operator **B** only operates on Boolean formulas. Thus introspection is disallowed, i.e., formulas such as $\mathbf{B}\neg\mathbf{B}\alpha$ or $\mathbf{B}[*\alpha]\beta$ are not well-formed. Secondly, the formula α that we contract, revise, or expand with, must always be Boolean. Thus, formulas such as $[*\neg\mathbf{B}\alpha]\beta$ are not well-formed either. The reason for these limitations is obvious. Since the agent only holds beliefs about the world that his doxastic state is not a part of, he has no "higher order" beliefs. And since he only receives information that concerns the external world, he cannot revise his beliefs with propositions about his own doxastic state.

What happens if we remove these limitations? What if we let **B** and the dynamic doxastic operators operate on arbitrary formulas, without restriction?

Now our object language can be defined in a much simpler way:
We define the sets **Term** and **Form** of *terms* and *formulas* to be the smallest sets satisfying the conditions:

(i) For any $n < \omega$, the propositional letter P_n belongs to **Form**.
(ii) $\perp \in$ **Form**.
(iii) If $\alpha, \beta \in$ **Form**, then $(\alpha \to \beta) \in$ **Form**.
(iv) If $\alpha \in$ **Form**, then $\mathbf{B}\alpha \in$ **Form**.
(v) If $\alpha \in$ **Form**, then $+\alpha, -\alpha, *\alpha \in$ **Term**.
(vi) If $\tau \in$ **Term** and $\alpha \in$ **Form**, then $[\tau]\alpha \in$ **Form**.

What would the semantics for such an *unlimited DDL* look like?

4. General semantics for unlimited DDL

When giving a semantic interpretation for unlimited DDL, we distinguish between:

(i) The *total state* (a "possible world") comprehending both the state of the agent and the state of the (external) world.
(ii) The state of the (external) world. We also use the term *world-state* for this component.
(iii) The *doxastic state* of the agent.

Let U be the set of all *total states*. We refer to the elements of U as x, y, z,... . We let **w** and **d** be two functions that to each state x in U assign a world state **w**(x) and a doxastic state **d**(x) of the agent. We let

$$W = \{\mathbf{w}(x) : x \in U\}$$

be the set of all *world-states* and

$$D = \{\mathbf{d}(x) : x \in U\}$$

be the set of all doxastic states. We refer to the elements of W as w, w',... and to the elements of D as d, d',... .

We make the following assumptions:

(1) For all x, y $\in$ U, x = y iff $\mathbf{w}(x) = \mathbf{w}(y)$ and $\mathbf{d}(x) = \mathbf{d}(y)$.

(2) For any combination of a world state w and a doxastic state d, there exists a (unique) x $\in$ U such that $\mathbf{w}(x) = w$ and $\mathbf{d}(x) = d$. We use the notation (w, d) for this total state x.

Even though we use the pair notation (w, d) for total states, the reader should not identify total states with *ordered pairs* built up from world states and doxastic states. Such an identification would be inappropriate because in more specialized versions of the semantics for DDL, doxastic states will be identified with hypertheories interpreted as set-theoretic constructs built up from total states in U. Given such a construction, reduction of states to ordered pairs of world-states and doxastic states becomes impossible, unless we allow sets to be non-well-founded. The notation (w, d) is simply a shorthand for "the total state having w and d as its world state and its doxastic state, respectively".

We say that two points x = (w, d) and x' = (w', d') are *world-equivalent* if their worldly components w and w' are identical. They are *doxastically equivalent* if and only if d = d'.

Certain subsets of U are called *propositions*. We assume that the set of all propositions forms a Boolean set algebra with domain U. We say that a proposition P is *worldly* if and only if it is closed under world-equivalence:

whenever x $\in$ P and $\mathbf{w}(x) = \mathbf{w}(y)$, then y $\in$ P.

Analogously, P is *doxastic* if and only if it is closed under doxastic equivalence:

whenever x $\in$ P and $\mathbf{d}(x) = \mathbf{d}(y)$, then y $\in$ P.

The worldly propositions are the ones whose truth-values are independent of the agent's doxastic state. The doxastic propositions are independent of the state of the world.

In addition to the elements that have already been mentioned, a model should contain special components that correspond to the different doxastic operators: either accessibility relations between (total) states, or – what amounts to the same thing – functions from states to sets of states. These components should be made dependent on the **d**-function. Thus, if we let **b** be the function that to each state x assigns the set of states that are compatible with what is believed in x (i.e., if **b** is to be the component of the model that corresponds to the operator **B**), then we should impose the following restriction on **b**:

(i) If $\mathbf{d}(x) = \mathbf{d}(y)$, then $\mathbf{b}(x) = \mathbf{b}(y)$.

For doxastic dynamic operators, the dependence relationships are somewhat more complex. Let R^{τ} be the accessibility relation on states that corresponds to the operator $[\tau]$. Since we take τ to be a doxastic action, that only modifies the doxastic state but does not "touch" the (external) world, we must assume that:

(ii) If $R^{\tau}(x, y)$, then $\mathbf{w}(x) = \mathbf{w}(y)$.

Furthermore,

(iii) If $\mathbf{d}(x) = \mathbf{d}(x')$, $\mathbf{d}(y) = \mathbf{d}(y')$, $\mathbf{w}(x) = \mathbf{w}(y)$ and $\mathbf{w}(x') = \mathbf{w}(y')$, then $R^{\tau}(x, y)$ iff $R^{\tau}(x', y')$,

i.e., doxastic actions are dependent only on the doxastic components of the states involved.

The above assumptions seem to be sufficient as far as a general semantics for DDL is concerned. Thus, we define a *model* $\mathfrak{M}$ to be a structure $<U, \mathbf{Prop}, \mathbf{w}, \mathbf{d}, \mathbf{b}, R, \models>$, where U, **w**, **d**, **b** are as described above.

Prop is a Boolean set-algebra with domain U, the elements of which are called *propositions*. In terms of **Prop**, we define a topology **T** in which the *closed sets* are the intersections of the subsets of **Prop**. We assume that **T** is compact.

R is a function that for every term τ yields an accessibility relation $R^{\tau} \subseteq U \times U$. $\models$ is a relation between elements of U and formulas that satisfies the conditions:

(i) If α is a formula, then the set $[\![\alpha]\!] = \{x \in U: x \models \alpha\}$ belongs to **Prop**. Moreover, we assume that if α is atomic, then $[\![\alpha]\!]$ is a worldly proposition. It follows that all Boolean formulas express worldly propositions.

(ii) It is not the case that $x \models \bot$.

(iii) $x \models (\alpha \rightarrow \beta)$ iff it is either the case that not: $x \models \alpha$ or it is the case that $x \models \beta$.

(iv) $x \models \mathbf{B}\alpha$ iff $\mathbf{b}(x) \subseteq [\![\alpha]\!]$).

(v) If τ is a term, then

$x \models [\tau]\alpha$ iff for all y such that $R^{\tau}(x, y)$, $y \models \alpha$.

Here we have suppressed the reference to the model $\mathfrak{M}$. When we need to be fully explicit, we write $\mathfrak{M}, x \models \alpha$ instead of $x \models \alpha$.

Let X be a class of models. We then define the notions of X-consequence and X-validity in the expected way. α is an *X-consequence* of a set of formulas Γ (in symbols, $\Gamma \models_X \alpha$) if and only if, for any model $\mathfrak{M}$ in X and any state x in $\mathfrak{M}$, if $\mathfrak{M}, x \models \beta$ for every β in Γ, then $\mathfrak{M}, x \models \alpha$. α is *X-valid* (in symbols, $\models_X \alpha$) if and only if, for every model $\mathfrak{M}$ in X and every state x in $\mathfrak{M}$, $\mathfrak{M}, x \models \alpha$.

5. Segerberg-style semantics for unlimited DDL

In a Segerberg-style semantics, we let doxastic states be hypertheories. While Segerberg in his semantics for basic DDL took hypertheories to be families of sets of world-states, we take them to be families of sets of total states that have both a worldly and a doxastic component. This change is necessary when we provide semantics for *unlimited* DDL. To be more precise, we let a hypertheory be any family **H** of *closed* sets of U that contains U and $\cap\mathbf{H}$.

For each x, we let $\mathbf{d}$(x) be a hypertheory. We then identify $\mathbf{b}$(x) — the set of states that are compatible with what is believed in x — with $\cap\mathbf{d}(x)$, the intersection of all the subsets of U that belong to $\mathbf{d}$(x). That is, we have:

$$x \models \mathbf{B}\alpha \text{ iff } \cap\mathbf{d}(x) \subseteq [\![\alpha]\!].$$

In accordance with the recipes provided in Section 2 above, we define the accessibility relations corresponding to the operators of *expansion* and *contraction* ($[+\alpha]$ and $[-\alpha]$, with α being an arbitrary formula) as follows.

Let us first define two operations on hypertheories. For any hypertheory **H** and any proposition P, let us define the *expansion* of **H** with P as:

$$\mathbf{H} + P = \mathbf{H} \cup \{X \cap P\colon X \in \mathbf{H}\}.$$

The *restriction* of **H** to a set Z is defined as:

$$\mathbf{H}\,|\,Z = \{X \in \mathbf{H}\colon Z \subseteq X\}.$$

A hypertheory **H'** is a *contraction* of **H** with P if and only if either (i) $P = U$ and $\mathbf{H}' = \mathbf{H}$; or (ii) $P \neq U$ and there is some (U–P)-permitting limit Z with respect to **H** such that $\mathbf{H}' = \{Z\} \cup \mathbf{H}\,|\,Z$.

We can now define the two accessibility relations between total states corresponding to expansion and contraction:

$R^{+\alpha}(x, y)$ iff
(i) $\mathbf{w}(x) = \mathbf{w}(y)$, and
(ii) $\mathbf{d}(y) = \mathbf{d}(x) + [\![\alpha]\!]$.

$R^{-\alpha}(x, y)$ iff
(i) $\mathbf{w}(x) = \mathbf{w}(y)$, and
(ii) $\mathbf{d}(y)$ is a contraction of $\mathbf{d}(x)$ with $[\![\alpha]\!]$.

Given these definitions, it might seem straightforward to specify the accessibility relation corresponding to the revision operator $[*\alpha]$:

$$R^{*\alpha}(x, y) \text{ iff for some } z,\ R^{-(\neg\alpha)}(x, z) \text{ and } R^{+\alpha}(z, y).$$

Since $R^{*\alpha}$ is defined as the relative product of $R^{-(\neg\alpha)}$ and $R^{+\alpha}$, the operator $[*\alpha]$ is explicitly definable as $[-(\neg\alpha)][+\alpha]$. However, as we shall argue, this Levi-style definition of $*$ might have to be given up in view of the problem that will be presented next.

6. *Paradoxes*

Whether we choose to accept this particular Segerberg-style modelling for unlimited DDL or prefer to work with the general model, we encounter the following difficulty. Since an introspective agent's own doxastic state is itself a part of the reality that he has views about, when such an agent learns more about the world, then the reality that he confronts undergoes a change. This feature of introspective reasoning leads to difficulties for the theory of belief revision.

Let us say that revision $*$ is *strongly paradoxical*, if for every state x and every formula α, the following formula is true in x:

(Strong Paradox) $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \rightarrow [*\alpha]\mathbf{B}\bot$.

The opposite of strong paradoxicality just requires that there should be a state x and a formula α such that $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \wedge \neg[*\alpha]\mathbf{B}\bot$ holds in x. This seems to be a very reasonable requirement on any belief revision operation.

Suppose for example that (α) it is actually raining in Lund, but I don't believe it ($\alpha \wedge \neg\mathbf{B}\alpha$), nor do I believe the opposite ($\neg\mathbf{B}\neg\alpha$). As a matter of fact, I also correctly believe that I don't believe that it is raining in Lund ($\mathbf{B}\neg\mathbf{B}\alpha$). Now, someone informs me that it is in fact raining in Lund. Surely, we would expect that $\neg[*\alpha]\mathbf{B}\bot$. It does not seem reasonable to assume that after receiving the information α, I will acquire inconsistent beliefs. However, it can be shown that the following holds:

Lemma 1. Suppose that $*$ satisfies *Preservation and Success,* while **B** satisfies *Positive Introspection*:

(P)	$\neg\mathbf{B}\neg\alpha \rightarrow (\mathbf{B}\beta \rightarrow [*\alpha]\mathbf{B}\beta)$	*(Preservation)*
(S)	$[*\alpha]\mathbf{B}\alpha$	*(Success)*
(PI)	$\mathbf{B}\alpha \rightarrow \mathbf{B}\mathbf{B}\alpha$.	*(Positive Introspection)*

Then, if the operator $[*\alpha]$ satisfies closure under logical implication:

$$\text{if} \vDash \beta \rightarrow \gamma, \text{ then } \vDash [*\alpha]\beta \rightarrow [*\alpha]\gamma,$$

and both $[*\alpha]$ and **B** satisfy closure under conjunction, $*$ is strongly paradoxical.[6]

Proof: Suppose that $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha$ holds in x. Then, by (P),

(1) $[*\alpha]\mathbf{B}\neg\mathbf{B}\alpha$

is true in x. But by Success it is also true in x that:

(2) $[*\alpha]\,\mathbf{B}\alpha$.

---

[6] This lemma is closely related to Fuhrmann's (1989) "paradox of serious possibility". In present terms, Fuhrmann proves $\neg\mathbf{B}\neg\alpha \wedge \neg\mathbf{B}\alpha \rightarrow [*\alpha]\mathbf{B}\bot$, but he relies on Negative Introspection in addition to the positive one. Cf. also Levi (1988).

If $[*\alpha]$ is closed under logical implication, (2) and (PI) imply that:

(3) $[*\alpha]\,\mathbf{BB}\alpha$.

If, in addition, $[*\alpha]$ and **B** are closed under conjunction, (3) and (1) imply:

(4) $[*\alpha]\mathbf{B}(\mathbf{B}\alpha \wedge \neg\mathbf{B}\alpha)$,

which in turn yields the result: $[*\alpha]\mathbf{B}\bot$. Q. E. D.

Note that even in the absence of Positive Introspection, Preservation plus Success will yield unacceptable results. Say that $*$ is *paradoxical* if and only if, for every x and α, the following formula is true in x:

(Paradox) $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \rightarrow [*\alpha](\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha)$.

This means, in particular, that if the agent holds no opinion as regards α and correctly believes that he does not believe α, then, upon revision with α, he will believe that α and, at the same time, believe that he does not believe α. But then he has at least one false belief, namely that he does not believe α. The requirement that $*$ should not be paradoxical in this sense seems eminently plausible.

Lemma 2. Suppose that the $*$ satisfies *Preservation and Success*. Then if $[*\alpha]$ is closed under conjunction, $*$ is paradoxical.

Proof: Suppose that $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha$ holds in x. Then, by (P) and (S), respectively,

(1) $[*\alpha]\mathbf{B}\neg\mathbf{B}\alpha$

and

(2) $[*\alpha]\,\mathbf{B}\alpha$,

are true in x. But then, if $[*\alpha]$ is closed under conjunction,

(3) $[*\alpha](\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha)$

is true in x. Q. E. D.

A natural conclusion is that we should give up *Preservation* for $*$: If I originally neither believe nor disbelieve α and am aware of this fact and if I then learn that α is true, some of my original beliefs must be given up. In particular, I have to give up my original belief that I do not believe α.

As we have seen, it is Preservation that leads to the paradoxical results.[7] This does not mean, however, that Introspection, neither Positive nor Negative, is an unproblematic requirement. It seems to us that Positive and Negative Introspection ($\mathbf{B}\alpha \rightarrow \mathbf{BB}\alpha$ and $\neg\mathbf{B}\alpha \rightarrow \mathbf{B}\neg\mathbf{B}\alpha$, respectively) should also be given up, but for a different reason, having to do with *contraction* rather than with revision. Let us first consider why Negative Introspection is inappropriate as a general requirement. When we originally do not believe α and then contract with $\neg\mathbf{B}\alpha$ (i.e., stop believing that we do not believe α), then $\neg\mathbf{B}\alpha$ should still be true in the contracted state (contracting with $\neg\mathbf{B}\alpha$ should not make us believe α) but it won't be believed any longer: $\mathbf{B}\neg\mathbf{B}\alpha$ will be false. Thus, in this contracted state, Negative Introspection will be violated.

That *Positive* introspection will also sometimes be violated is less obvious, but think of an agent who originally believes α and believes that he does believe α. Suppose he is invited to contract his beliefs with $\mathbf{B}\alpha$ (i.e., stop believing that he believes α). In the contracted state, it is no longer true that $\mathbf{BB}\alpha$, but we would like to allow that it is still true that $\mathbf{B}\alpha$. This is, however, impossible unless positive introspection is violated after the contraction. If we insisted on positive introspection being valid, we would have to stop believing α just because we stop believing that we believe α. This seems wrong.

While Positive and Negative Introspection should probably be given up, it seems that we instead might still insist on their converses: we might insist that an (ideal) agent's beliefs concerning his own beliefs are never mistaken:

(VPI) $\mathbf{BB}\alpha \rightarrow \mathbf{B}\alpha$ *(Veridicality of Positive Introspection)*

(VNI) $\neg\mathbf{B}\bot \rightarrow (\mathbf{B}\neg\mathbf{B}\alpha \rightarrow \neg\mathbf{B}\alpha)$. *(Veridicality of Negative Introspection)*

(The latter requirement is a slightly qualified converse of Negative Introspection: the qualification in the antecedent is added in order to allow states in which the agent holds inconsistent beliefs.) In our general semantic framework, these requirements are validated by the following conditions:

---

[7] Lemmas 1 and 2 point to an analogy between higher order beliefs and acceptance of so-called *Ramsey conditionals*, i.e., conditionals > that satisfy the *Ramsey test*: $\neg\mathbf{B}\bot \rightarrow (\mathbf{B}(\alpha > \beta) \leftrightarrow [*\alpha]\mathbf{B}\beta)$. Gärdenfors' impossibility result (1988), proved for DDL in Lindström and Rabinowicz (1997), shows that one cannot accommodate Ramsey conditionals within an AGM-type theory of belief revision without giving up Preservation. Thus, Ramsey conditionals and higher-order beliefs are alike in that they should sometimes be given up when we add new information to our stock of beliefs. It is not surprising that Ramsey conditionals behave like beliefs about beliefs in this respect. After all, what the Ramsey test says is that the agent should accept the conditional "If α, then β" just in case he is disposed to believe β, if he were to learn α. That is, the agent's belief in conditionals should *reflect* his conditional dispositions to believe. In the light of new information compatible with what the agent believes, it might very well be rational to relinquish some of these conditional dispositions. But then, according to the Ramsey test, the agent should also cease to believe the corresponding conditionals.

In Lindström and Rabinowicz (1997), Gärdenfors' impossibility theorem is discussed in the context of DDL. For a more comprehensive discussion of the Ramsey test, see Lindström and Rabinowicz (1995).

If $y \in \mathbf{b}(x)$, there is some $z \in \mathbf{b}(x)$ such that $y \in \mathbf{b}(z)$.
If $\mathbf{b}(x) \neq \varnothing$, there is some $y \in \mathbf{b}(x)$ such that $\mathbf{b}(y) \subseteq \mathbf{b}(x)$.

Note that both conditions would follow from the following restriction on the model:

If $\mathbf{b}(x) \neq \varnothing$, there is some $y \in \mathbf{b}(x)$ such that $\mathbf{b}(y) = \mathbf{b}(x)$.

According to this condition, the agent is never mistaken about his beliefs.

For future reference, we may also mention an even stronger condition according to which an agent is never mistaken about his *doxastic state.* Thus, he does not make mistakes – neither about his beliefs nor about his policies for belief change. He might not be fully informed about his doxastic state (in particular, he might violate positive and negative introspection) but the beliefs he holds about it are never false:

Full Veridicality of Introspection
If $\mathbf{b}(x) \neq \varnothing$, there is some $y \in \mathbf{b}(x)$ such that $\mathbf{d}(y) = \mathbf{d}(x)$.

Let us now return to our problem with revision. If revision is not to be paradoxical, it should not be fully preservative: in particular, certain beliefs about one's own beliefs need to be given up when one receives new information. In particular, when receiving the information α, the agent should give up his original belief that he does not believe α. How can we achieve this result? Here is a suggestion.

Levi-style revision with α consists in two steps: we first contract with $\neg\alpha$ and then expand with α. In some cases, the first step is vacuous, $\neg\alpha$ is not believed to begin with. Then revision reduces to expansion. These are precisely the cases for which Preservation is meant to hold: revision has been supposed to be preservative simply because expansion is cumulative: all the old beliefs are kept when we expand with a new belief. Our suggestion is to replace the expansion step in the process of revision with what might be called *cautious expansion:* before we expand with α, we should first make sure that we give up the belief that we do not believe α. Clearly, this belief should not survive our coming to believe that α. Thus, unlike standard expansion, cautious expansion is not fully cumulative: certain beliefs have to be given up when new beliefs are added. This suggests the following definition of the *cautious expansion* operator $[\oplus\alpha]$:

$$[\oplus\alpha]\beta =\text{df } [-(\neg\mathbf{B}\alpha)][+\alpha]\beta.$$

Thus, cautious expansion with α is itself a two-step process: we first contract with $\neg\mathbf{B}\alpha$ and only then expand with α.

We can then define revision with α in a new way – as contraction with $\neg\alpha$ followed by cautious expansion with α:

$$[*\alpha]\beta =\text{df } [-(\neg\alpha)][\oplus\alpha]\beta.$$

How does this relate to the Segerberg-style semantics for unlimited DDL? The definitions of the accessibility relations that correspond to contraction and (standard) expansion may be kept unchanged. But the accessibility relation that corresponds to revision will have to be modified. $R^{*\alpha}$ will now be interpreted as the relative product of $R^{-(\neg\alpha)}$ and $R^{\oplus\alpha}$, where $R^{\oplus\alpha}$ will itself be the relative product of $R^{-(\neg \mathbf{B}\alpha)}$ and $R^{+\alpha}$.

One might wonder, however, if our cautious expansion is sufficiently cautious. It is easy to see that the suggested definition of cautious expansion would not be cautious enough if introspection weren't assumed to be veridical. When we prepare the ground for the expansion with α, we give up the second-order belief that we do not believe α. But couldn't there be some higher-order beliefs that should also be given up? Suppose that in the original state in which the agent does not believe that α, he is fully reflective, so that $\neg\mathbf{B}\alpha$, $\mathbf{B}\neg\mathbf{B}\alpha$, $\mathbf{BB}\neg\mathbf{B}\alpha$, etc., are all true in that state. If he then contracts with $\neg\mathbf{B}\alpha$, as the first step in cautious expansion with α, will then these higher-order beliefs automatically disappear? This would be desirable, but to make them disappear we need introspection to be veridical. When the agent contracts with $\neg\mathbf{B}\alpha$, then – given the veridicality of introspection – he will lose not just his belief in $\neg\mathbf{B}\alpha$ but also all his higher-order beliefs: not just $\mathbf{B}\neg\mathbf{B}\alpha$, but also $\mathbf{BB}\neg\mathbf{B}\alpha$, etc., will all be false. Otherwise, if he kept one of these higher-order beliefs, some of his introspective beliefs would not be veridical. Proof: Suppose that n ($n > 1$) is the lowest number such that $\mathbf{B}^n\neg\mathbf{B}\alpha$ is still true after contraction. Then the agent has a false introspective belief that $\mathbf{B}^{n-1}\neg\mathbf{B}\alpha$. And he would hold on to that false belief after the second step of the cautious expansion. This would clearly be an unwanted result.

In fact, it seems desirable to accept Full Veridicality of Introspection. Otherwise, when contracting with $\neg\mathbf{B}\alpha$, we might not get rid of some of the original beliefs concerning outcomes of potential belief change — beliefs that are dependent on our belief in $\neg\mathbf{B}\alpha$ and that would become false when belief in $\neg\mathbf{B}\alpha$ is removed. As long as we demand Full Veridicality of Introspection, this possibility need not worry us.

Still, what *is* worrying is that the proposal we have just sketched is so dependent on the assumption of veridicality. This gives it an air of *ad hocness* and suggests that the cautious expansion approach does not really go to the heart of the problem. In fact, it can be shown that imposing veridicality is a rather dangerous medicine: while veridicality solves some of the problems of the cautious approach, it creates at the same time new problems, at least as devastating.

Consider the following story: as in our previous example, the agent has no opinion about α, but now we assume that α happens to be true. In particular, then, it is true that

(1) $\alpha \wedge \neg\mathbf{B}\alpha$.

The agent is now informed that (1) holds; he has received true information. Since (1) is true, it is clearly a consistent proposition. We would therefore expect that revision with (1) will not lead the agent to an inconsistent belief state. That the revision of a consistent doxastic state with a consistent proposition always leads to a consistent doxastic is, in fact, one of the fundamental principles of any reasonable theory of belief revision. In particular, then, it should be the case that

(2) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\neg\mathbf{B}\bot$.

By Veridicality of Negative Introspection, (2) implies that, after revision with (1), all the agent's beliefs to the effect that he does not believe some β must be veridical. In particular:

(3) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}\neg\mathbf{B}\alpha \rightarrow [*(\alpha \wedge \neg\mathbf{B}\alpha)]\neg\mathbf{B}\alpha$.

But we also know that revision, *whether cautious or not*, is supposed to satisfy Success. Thus, upon the revision with (1), the agent must believe that (1) holds:

(4) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}(\alpha \wedge \neg\mathbf{B}\alpha)$.

Given that * is closed under logical consequence and **B** distributes over conjunction, (4) implies that

(5) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha$.

(3) and (5), taken together, imply

(6) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}\alpha \wedge \neg\mathbf{B}\alpha$,

which implies

(7) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\bot$.

Contrary to what we should expect, revision with a true proposition such as (1) turns out to be impossible!

The puzzles that we were discussing earlier (cf. Lemmas 1 and 2) – that the cautious expansion approach was designed to solve – concerned the fact that in unlimited DDL we are studying agents that have beliefs about their own beliefs. The problem was how these higher-order beliefs should be adjusted when an agent receives new information. The cautious expansion approach would then, perhaps, be satisfactory as long as we were only considering what happens when an agent receives new information about the world. But in unlimited DDL, an agent can revise his belief state not only with worldly propositions, but also with propositions that concern his own doxastic state. As we saw, the possibility of revising ones beliefs with doxastic propositions, for example of the form $\alpha \wedge \neg\mathbf{B}\alpha$, lead to difficulties that the cautious expansion approach is unable to handle. In order to focus on these problems more sharply, we state the following lemma:

Lemma 3. Suppose that **B** and $[*\alpha]$ are normal modal operators in unlimited DDL that satisfy the following principles:

(S) $[*\alpha]\mathbf{B}\alpha$ *(Success)*

(PR) $\neg[*\alpha]\bot$ *(Possibility of Revision)*

(C) $\alpha \to [*\alpha]\neg\mathbf{B}\bot$. *(Consistency Principle)*

Then:

(a) If the operator **B** satisfies:

(VPI) $\mathbf{BB}\alpha \to \mathbf{B}\alpha$ *(Veridicality of Positive Introspection)*

then all the agent's beliefs must be true, i.e.,

$$\mathbf{B}\alpha \to \alpha.$$

(b) If the operator **B** satisfies:

(VNI) $\neg\mathbf{B}\bot \to (\mathbf{B}\neg\mathbf{B}\alpha \to \neg\mathbf{B}\alpha)$, *(Veridicality of Negative Introspection)*

then the agent believes every true proposition, i.e.,

$$\alpha \to \mathbf{B}\alpha,$$

which in turn implies that the agent is either inconsistent or completely accurate in his beliefs, i.e.,

$$\neg\mathbf{B}\bot \to (\mathbf{B}\alpha \leftrightarrow \alpha).$$

Proof: We first notice that if **B** is a normal modal operator satisfying VPI then it satisfies:

(a) $\mathbf{B}(\mathbf{B}\alpha \wedge \neg\alpha) \to \mathbf{B}\bot$,

i.e., only an inconsistent agent satisfying VPI can believe $\mathbf{B}\alpha \wedge \neg\alpha$. Similarly, if **B** is a normal modal operator satisfying (VNI), then it satisfies:

(b) $\mathbf{B}(\alpha \wedge \neg\mathbf{B}\alpha) \to \mathbf{B}\bot$.

Suppose now that **B** satisfies (VPI). By Success, we have:

(1) $[*(\mathbf{B}\alpha \wedge \neg\alpha)]\mathbf{B}(\mathbf{B}\alpha \wedge \neg\alpha)$.

In view of (a), he will then also satisfy:

(2) $[*(\mathbf{B}\alpha \wedge \neg\alpha)]\mathbf{B}\bot$

for every α. Suppose now that for some particular α, it is true that:

(3) $\mathbf{B}\alpha \wedge \neg\alpha$ (i.e., $\neg(\mathbf{B}\alpha \to \alpha)$)

Then, by the Consistency Principle (C), it is also true that:

(4) $[*(\mathbf{B}\alpha \wedge \neg\alpha)]\neg\mathbf{B}\bot$.

From (2) and (4) we get:

(5) $[*(\mathbf{B}\alpha \wedge \neg\alpha)]\bot$,

which contradicts (PR). Hence, by reductio, we get:

(6) $\mathbf{B}\alpha \rightarrow \alpha$.

The proof of (b) is parallel:
We first use (VNI) and Success to prove:

(7) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}\bot$.

But now, if $\alpha \wedge \neg\mathbf{B}\alpha$, then by the (C),

(8) $[*(\alpha \wedge \neg\mathbf{B}\alpha)]\neg\mathbf{B}\bot$.

But (7) and (8) together with (PR) yield a contradiction.
Hence, for all α,

(9) $\alpha \rightarrow \mathbf{B}\alpha$.

Q.E.D.

(S), (PR) and (C) appear to be valid principles on any reasonable view of belief revision. The cautious expansion approach does not touch these principles. Moreover, this approach is based on VNI. Thus, as long as we do not put any restrictions on the formulas that we revise with, the cautious expansion approach also leads to paradoxical results.

Clearly, then, the cautious approach does not solve all our problems. We need a more comprehensive solution. The next section delineates a rather radical proposal that might give us a way out of our difficulties.

7. The two-dimensional approach

When an introspective agent gets new information, his doxastic state undergoes a change. Thereby the total state changes as well. What are then his beliefs about? The original state or the new one? One would like to say that he has beliefs about the old state as well as about the new one. In general, therefore, we have to distinguish between the state in which beliefs are held (the *point of evaluation*) and the state about which certain things are believed (the *point of reference*).[8] This means that our semantics has to be made much more powerful.

As before we assume that each total state x has both a worldly component $\mathbf{w}(x)$ and a doxastic component $\mathbf{d}(x)$. But now $\mathbf{d}(x)$ is a function that to each possible point of reference y

---

[8] The first general treatment of two-dimensional modal logic occurred in Segerberg (1973). Both authors have previously adopted two-dimensional approaches to epistemic logic in connection with the so-called *Paradox of Knowability* (Cf., Rabinowicz and Segerberg (1994) and Lindström (1997)). The reader who would like to compare these different approaches should be warned that, apart from some differences of substance, the terminology differs between our earlier papers and the present one.

assigns a doxastic state $\mathbf{d}(x)(y)$ that specifies the agent's views *in* the evaluation point x *about* the reference point y. Instead of $\mathbf{d}(x)(y)$, we shall write $\mathbf{d}_y(x)$. We may speak of $\mathbf{d}_y(x)$ the agent's *doxastic state in* x *about* y. In a Segerberg-style semantics, we can identify each such $\mathbf{d}_y(x)$ with a hypertheory. In the same way, we relativize $\mathbf{b}(x)$ to various reference points y and write $\mathbf{b}_y(x)$ for each such relativization. Intuitively, $\mathbf{b}_y(x)$ is the set of points that are compatible with the agent's beliefs in x about y. In a Segerberg-style semantics, we have $\mathbf{b}_y(x) = \cap\mathbf{d}_y(x)$.

In addition, the accessibility relations that correspond to different dynamic doxastic operators should be made sensitive to different points of reference. The idea is that a doxastic action in x is always supposed to consist in a transformation of some definite belief state $\mathbf{d}_y(x)$ about some point of reference y. Hence, we associate an accessibility relation $R(\tau, y)$ with every doxastic action τ and every point of reference y. Intuitively, $R(\tau, y)(x, z)$ holds if and only if z is a possible result of performing τ on the doxastic state $\mathbf{d}_y(x)$ (which in its turn may necessitate some adjustments in other parts of $\mathbf{d}(x)$). One might say that y is the *point of reference* of the accessibility relation $R(\tau, y)$.

Formally, we define a model $\mathfrak{M}$ to be a structure:

$$\mathfrak{M} = \langle U, W, D, \mathbf{Prop}, \mathbf{w}, \mathbf{d}, \mathbf{b}, R, \models\rangle$$

such that:

(1) U, W and D are non-empty sets of total states, world-states and doxastic states, respectively.

(2) **Prop** is, as before, a Boolean set algebra with the set U as its domain. In terms of **Prop**, we define a topology **T** in which the *closed sets* are the intersections of the subsets of **Prop**. We assume that **T** is compact.

(3) **w** is a function from U onto W.

(4) **d** is a function from $U \times U$ onto D. We usually write $\mathbf{d}_y(x)$ for $\mathbf{d}(x, y)$ and let $\mathbf{d}(x)$ be the function $(\lambda y \in U)\mathbf{d}_y(x)$.

(5) **b** is a function from $U \times U$ to $\wp(U)$. Usually we write $\mathbf{b}_y(x)$ instead of $\mathbf{b}(x, y)$. Thus, for all x, y, $\mathbf{b}_y(x) \subseteq U$.

(6) For each doxastic action term τ and every $x \in U$, $R(\tau, x) \subseteq U \times U$.

(7) **w**, **d**, **b** and R satisfy the conditions:

(i) $x = y$ iff $\mathbf{w}(x) = \mathbf{w}(y)$ and $\mathbf{d}(x) = \mathbf{d}(y)$.

(ii) for any $w \in W$ and any $x \in U$, there is a $z \in U$ such that $\mathbf{w}(z) = w$ and $\mathbf{d}(z) = \mathbf{d}(x)$. Hence, we can write U on the form $\{(w, d): w \in W \text{ and } d = \mathbf{d}(x), \text{ for some } x \in U\}$.

(iii) If $\mathbf{d}_z(x) = \mathbf{d}_z(y)$, then $\mathbf{b}_z(x) = \mathbf{b}_z(y)$.

(iv) If $R(\tau, z)(x, y)$, then $\mathbf{w}(x) = \mathbf{w}(y)$.

(v) If $\mathbf{d}(x) = \mathbf{d}(x')$, $\mathbf{d}(y) = \mathbf{d}(y')$, $\mathbf{w}(x) = \mathbf{w}(y)$ and $\mathbf{w}(x') = \mathbf{w}(y')$, then $R(\tau, z)(x, y)$ iff $R(\tau, z)(x', y')$.

These conditions should be compared to the corresponding conditions for the one-dimensional models that we defined earlier.

(8) The formulas of the language are no longer assigned truth-values at single points but rather at ordered pairs $<x, y>$ of points, where x is the point of evaluation and y the point of reference. The truth-relation $\vDash$ satisfies the following requirements (we write $x, y \vDash \alpha$ for $<x, y> \vDash \alpha$):

(i) If α is a formula, then the set $[\![\alpha]\!]_y = \{x \in U: x, y \vDash \alpha\}$ belongs to **Prop**. Moreover, we assume that if α is atomic, then α expresses one and the same worldly proposition $[\![\alpha]\!]_y$ relative to every point of reference y.

(ii) It is not the case that $x, y \vDash \bot$.

(iii) $x, y \vDash (\alpha \rightarrow \beta)$ iff it is either the case that not: $x, y \vDash \alpha$ or it is the case that $x, y \vDash \beta$.

(iv) $x, y \vDash \mathbf{B}\alpha$ iff $\mathbf{b}_y(x) \subseteq [\![\alpha]\!]_y$

(v) If τ is a term, then
$x, y \vDash [\tau]\alpha$ iff for all z such that $R(\tau, y)(x, z)$, $z, y \vDash \alpha$.

We read $x, y \vDash \alpha$ as: α is *true at* the point x *with reference to* the point y.

(9) We also extend the language with a new operator † that takes the current point of evaluation and makes it the point of reference:[9]

(vi) $x, y \vDash \dagger\alpha$ iff $x, x \vDash \alpha$.

We introduce † in order to be able to distinguish between an agent's *posterior beliefs about his original state* (the one he is in before performing a doxastic action):

(1) $x, x \vDash [\tau]\mathbf{B}\alpha$

and his *posterior beliefs about the posterior state* (the one he is in after the action):

(2) $x, x \vDash [\tau]\dagger\mathbf{B}\alpha$.

(1) is equivalent to:

(1') for all z such that $R(\tau, x)(x, z)$, $z, x \vDash \mathbf{B}\alpha$,

while (2) can be written as:

(2') for all z such that $R(\tau, x)(x, z)$, $z, z \vDash \mathbf{B}\alpha$.

---

[9] The †-operator is discussed in Lewis (1973), Section 2.8. See also Segerberg (1973).

We say that a formula α is *ordinary* if its truth or falsity does not depend on the point of reference, i.e., if for all x, y, z:

$$x, y \models \alpha \text{ iff } x, z \models \alpha.$$

A formula is *special*, if it is not ordinary.

An ordinary formula α expresses one and the same proposition (written, $[\![\alpha]\!]$) with reference to every point of reference. It is easily seen that:

(a) Boolean formulas are ordinary.

(b) For any formula α, $\dagger\alpha$ is ordinary.

If α is ordinary, then

$$x, y \models \dagger\alpha \text{ iff } x, y \models \alpha.$$

Thus, for ordinary α,

$$[\![\dagger\alpha]\!] = [\![\alpha]\!].$$

The proposition

$$[\![\dagger\alpha]\!] = \{x \in U: x, x \models \alpha\},$$

we call the *diagonal proposition* corresponding to α.

We say that a formula α, when interpreted in $\mathfrak{M}$, is *true at the point* x if and only if x, x $\models \alpha$. In other words, α is true at x if and only if the proposition $[\![\alpha]\!]_X$ expressed by α with reference to x is true at the point x itself.

We say that a formula α is *valid* (or *weakly valid*) in the model $\mathfrak{M}$ (in symbols, $\mathfrak{M} \models \alpha$) if and only if α is true at every point in $\mathfrak{M}$.

Let us say that a pair <x, y> of points in U is *normal* if x = y. We have defined truth at a point x as truth relative to the normal pair <x, x>, and we have defined validity in a model $\mathfrak{M}$ as truth relative to all *normal pairs* in $\mathfrak{M}$.

There is another notion of validity in a model: We say that α is *strongly valid* in the model $\mathfrak{M}$ if and only if, for every pair of points <x, y> in $\mathfrak{M}$, <x, y> $\models \alpha$. Of course, if α is strongly valid in $\mathfrak{M}$, then α is valid in $\mathfrak{M}$. The converse does not hold in general. Consider, for example,

$$\alpha \leftrightarrow \dagger\alpha.$$

Every instance of this schema is weakly valid in every model. However, if α is a special formula, then $\alpha \leftrightarrow \dagger\alpha$ is not strongly valid.

Notice, however, that for every model $\mathfrak{M}$, α is weakly valid in $\mathfrak{M}$ if and only if $\dagger\alpha$ is strongly valid in $\mathfrak{M}$.

Let **K** be a class of models. We say that α is ***K**-valid (strongly **K**-valid)* if α is valid (strongly valid) in every model in **K**.

Let us now define a two-dimensional Segerberg-style semantics in which the doxastic states are hypertheories. As before, we let a hypertheory be a family **H** of closed sets of U that contains U and $\cap\mathbf{H}$.

For each x, y $\in$ U, we let $\mathbf{d}_y(x)$ be a hypertheory. We then let $\mathbf{b}_y(x) = \cap\mathbf{d}_y(x)$. That is, we have:

$$x, y \models \mathbf{B}\alpha \text{ iff } \cap\mathbf{d}_y(x) \subseteq [\![\alpha]\!]_y.$$

This means that $\mathbf{B}\alpha$ is true relative to a pair <x, y>, consisting of a point of evaluation x and a point of reference y, if and only if the proposition expressed by α with respect to y is true in every point z that is compatible with everything that the agent believes in x about the point y.

The accessibility relations corresponding to the operations of *expansion* and *contraction* are now characterized in the following way:

If $R(+\alpha, z)(x, y)$, then

(i) $\mathbf{w}(x) = \mathbf{w}(y)$, and

(ii) $\mathbf{d}_z(y) = \mathbf{d}_z(x) + [\![\alpha]\!]_z$.

If $R(-\alpha, z)(x, y)$, then

(i) $\mathbf{w}(x) = \mathbf{w}(y)$, and

(ii) $\mathbf{d}_z(y)$ is a contraction of $\mathbf{d}_z(x)$ with $[\![\alpha]\!]_z$,

where the notion of a contraction of a hypertheory with a proposition is defined in exactly the same way as in Section 5.[10]

We then define the accessibility relation that corresponds to *revision* by means of the following Levi-style condition:

$$R(*\alpha, z)(x, y) \text{ iff for some } u, R(-(\neg\alpha), z)(x, u) \text{ and } R(+\alpha, z)(u, y).$$

Let us now see how the two-dimensional semantics handles the paradoxes of the previous section. The two-dimensional Segerberg-style semantics that we have just outlined, strongly validates the following conditions:

---

[10] Notice that we no longer provide full definitions of the accessibility relations that correspond to expansion and contraction. Instead, these relations are *constrained* by the above Segerberg-style conditions. By not providing full definitions, we leave room for the possibility of the agent making adjustments in other parts of $\mathbf{d}(x)$ when he performs some doxastic action on

(P) $\neg\mathbf{B}\neg\alpha \to (\mathbf{B}\beta \to [*\alpha]\mathbf{B}\beta)$ *(Preservation)*

(S) $[*\alpha]\mathbf{B}\alpha$ *(Success)*

From these two conditions, we proved, in Section 6 (Lemma 2), the condition:

(1) $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \to [*\alpha](\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha)$,

which we called Paradox. This formula also follows in the present framework. That is, it is strongly valid. So for every x, y $\in$ U in every model, we have:

$$x, y \vDash \neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \to [*\alpha](\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha).$$

However, the meaning of the formula (1) has changed from the old semantics to the new one: it is no longer paradoxical. To see this, one should compare (1) with:

(2) $\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \to [*\alpha]\dagger(\mathbf{B}\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha)$,

which is indeed paradoxical. However, (2) is, of course, not even weakly valid. While (1) is about what the agent, after having learned α, would believe about the state prior to the revision, (2) is about what he then would believe about the state obtaining after the revision. There is no reason to suppose that (2) would hold.

Consider now Lemma 1 of Section 6. Every two-dimensional model strongly validating Preservation, Success and

(PI) $\mathbf{B}\alpha \to \mathbf{B}\mathbf{B}\alpha$ *(Positive Introspection)*

will indeed also satisfy:

$\neg\mathbf{B}\neg\alpha \wedge \mathbf{B}\neg\mathbf{B}\alpha \to [*\alpha]\mathbf{B}\bot$ *(Strong Paradox)*

But here the way out of the Paradox has to do with the assumption of PI. This principle is only plausible for normal pairs <x, x>, where the point of evaluation and point of reference are the same.[11] So we can only assume this principle to be weakly valid in a given model. But weak validity is not closed under the principle:

$$\text{if } \vDash \beta \to \gamma, \text{ then } \vDash [*\alpha]\beta \to [*\alpha]\gamma.$$

Hence, we cannot infer from the weak validity of $\mathbf{B}\alpha \to \mathbf{B}\mathbf{B}\alpha$ in a model, to the weak validity of $[*\alpha]\mathbf{B}\alpha \to [*\alpha]\mathbf{B}\mathbf{B}\alpha$ in the same model. But this step is needed for the proof of Lemma 1 to go through in the new setting.

Finally, consider the case in which:

$\mathbf{d}_Z(x)$. We return to this possibility of making adjustments when we speak about *transfer principles* below.

[11] Suppose the two points are different and that x, y $\vDash$ $\mathbf{B}\alpha$. Then, if we intuitively assume that the agent in x is aware of his beliefs, then, in x, he believes *about x* that he there believes α about y. This is not the same as x, y $\vDash$ $\mathbf{B}\mathbf{B}\alpha$, which would mean that he in x believes *about y* that he there believes α about y.

(1) $x, x \vDash \alpha \wedge \neg\mathbf{B}\alpha$,

The agent then learns (1) and revises his beliefs about the point x with this information. By Success:

(2) $x, x \vDash [*(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}(\alpha \wedge \neg\mathbf{B}\alpha)$,

but there is nothing paradoxical about (2), since the beliefs that are referred to in the formula following the revision operator are all about the prior state x and not about the one posterior to the revision. In contrast to (2), the following situation would be paradoxical:

(3) $x, x \vDash [*(\alpha \wedge \neg\mathbf{B}\alpha)]\dagger\mathbf{B}(\alpha \wedge \neg\mathbf{B}\alpha)$.

But the formula occurring in (3) is of course not (even weakly) valid.

But what about the following formula?

$$[*\dagger(\alpha \wedge \neg\mathbf{B}\alpha)]\mathbf{B}\dagger(\alpha \wedge \neg\mathbf{B}\alpha).$$

Isn't this formula valid, by Success? Yes, indeed it is. What it says, is that if one revises one's original beliefs with the diagonal proposition $\dagger(\alpha \wedge \neg\mathbf{B}\alpha)$, then, in the posterior state, one will have the belief about the prior state that $\dagger(\alpha \wedge \neg\mathbf{B}\alpha)$ was true then. In our example, however, this posterior belief about the prior state is in fact true. Hence, there is nothing paradoxical about it.

The last formula may be contrasted with:

$$[*\dagger(\alpha \wedge \neg\mathbf{B}\alpha)]\dagger\mathbf{B}(\alpha \wedge \neg\mathbf{B}\alpha),$$

which says that the agent after revision with $\dagger(\alpha \wedge \neg\mathbf{B}\alpha)$ would believe $\alpha \wedge \neg\mathbf{B}\alpha$ about his *posterior* state. This would indeed be paradoxical. But this formula is not even weakly valid, so no paradox is forthcoming.

In the two-dimensional semantics, we can impose various introspection principles, like (PI), (NI), (VPI), (VNI). These principles do not lead to trouble as long as we only assume them to be weakly, rather than strongly, valid.

Our conclusion is that the two-dimensional semantics avoids the original paradoxes, without – as far as we can see – creating new ones. This semantics has one serious drawback, however: it only determines the agent's posterior beliefs about the *prior* state:

(1) $[*\alpha]\mathbf{B}\beta$

What we would like to infer, however, are posterior beliefs about the posterior state:

(2) $[*\alpha]\dagger\mathbf{B}\beta$.

Thus, we would like to have some *transfer principles*: at least for all the Boolean formulas β, we would like it to be valid that

(3) $[*\alpha](\mathbf{B}\beta \to \dagger\mathbf{B}\beta)$,

(4) $[*\alpha](\neg \mathbf{B}\beta \rightarrow \dagger\neg \mathbf{B}\beta)$.

Such principles would allow us to infer from one of $[*\alpha]\mathbf{B}\beta$, $<*\alpha>\mathbf{B}\beta$, $[*\alpha]\neg\mathbf{B}\beta$, $<*\alpha>\neg\mathbf{B}\beta$, to the corresponding statements about the posterior state: $[*\alpha]\dagger\mathbf{B}\beta$, $<*\alpha>\dagger\mathbf{B}\beta$, $[*\alpha]\dagger\neg\mathbf{B}\beta$, $<*\alpha>\dagger\neg\mathbf{B}\beta$, respectively.

In fact, we want such transfer principles for posterior beliefs in worldly propositions to be weakly valid not just for all revisions $*\alpha$ but for all doxastic actions τ: if β is a Boolean formula, then

(3') $[\tau](\mathbf{B}\beta \rightarrow \dagger\mathbf{B}\beta)$,

(4') $[\tau](\neg\mathbf{B}\beta \rightarrow \dagger\neg\mathbf{B}\beta)$.

In order to validate these principles, however, we need to impose appropriate conditions on the accessibility relations. Thus, for all doxastic actions τ and all worldly propositions P, we should require that

$$\text{if } R(\tau, x)(x, y), \text{ then } \mathbf{b}_X(y) \subseteq P \text{ iff } \mathbf{b}_Y(y) \subseteq P.$$

In fact, even with respect to beliefs in doxastic propositions, there should be a large measure of agreement between posterior beliefs of this kind regarding the prior and the posterior state. If I initially believe that I believe the earth to be round, then after the revision with some information about, say, the weather in Sweden, I will keep my beliefs about what I believe to be the shape of the earth both with regard to my prior state and with regard to the posterior state. But transfer principles for posterior beliefs in doxastic propositions are much more difficult to formulate: many posterior beliefs about doxastic propositions are *not* transferable, as we have seen.

This shows that there is work that remains to be done. Still, we have at least made some first steps towards the development of a two-dimensional semantics for unlimited DDL. It is to be hoped that this project can be further developed.

References

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

Fuhrmann, A.: 1989, 'Reflective modalities and theory change', Synthese **81**, 115-134.

Grove, A.: 1988, 'Two modellings for theory change'', *Journal of Philosophical Logic* **17**, 157-170.

Gärdenfors, P.: 1988, *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Bradford Books, Cambridge, MA: MIT Press.

Hintikka, J.: 1962, *Knowledge and Belief*, Ithaca, N. Y.: Cornell University Press.

Levi, I.: 1988, 'Iteration of conditionals and the Ramsey test', Synthese **76**, 49-81.

Lewis, D.: 1973, *Counterfactuals.* Oxford: Blackwells.

Lindström, S.: 1997, 'Situations, truth and Knowability — a situation-theoretic analysis of a paradox by Fitch", in Ejerhed and Lindström (eds.) *Logic, Action and Cognition: Essays in Philosophical Logic*, Kluwer 1997, 183-209.

Lindström, S., and Rabinowicz, W.: 1989, 'On probabilistic representation of non-probabilistic belief revision', *Journal of Philosophical Logic* **18**, 69-101.

Lindström, S., and Rabinowicz, W.: 1990, 'Epistemic entrenchment with incomparabilities and relational belief revision', in Fuhrmann and Morreau (eds.), *The Logic of Theory Change*, Lecture Notes in Artificial Intelligence no. **465**, Springer Verlag, 93-126.

Lindström, S., and Rabinowicz, W.: 1992, 'Belief revision, epistemic conditionals and the Ramsey test', *Synthese* **91**, 195-237.

Lindström, S., and Rabinowicz, W.: 1997, 'Extending dynamic doxastic logic: accommodating iterated beliefs and Ramsey conditionals within DDL', in Lindahl, Needham and Sliwinski (eds.) *For Good Measure*, Uppsala Philosophical Studies 46, 126-153.

Rabinowicz, W. and Lindström, S.: 1994, 'How to model relational belief revision', in D. Prawitz and D. Westerståhl (eds.), *Logic and Philosophy of Science in Uppsala,* Kluwer, 1994, 69-84.

Rabinowicz, W. and Segerberg, K.: 1994, 'Actual truth, possible knowledge', *Topoi* **13**, 101-115.

de Rijke, M.: 1994, 'Meeting some neighbours: a dynamic modal logic meets theories of change and knowledge representation'. In *Logic and Information Flow*, edited by Jan van Eijck and Albert Visser, 170-196. Cambridge, MA: MIT Press.

Segerberg, K.: 1973, 'Two-dimensional modal logic', *Journal of Philosophical Logic* **2**, pp. 77-96.

Segerberg, K.: 1995a, 'Belief revision from the point of view of doxastic logic'. *Bulletin of the IGPL*, vol. 3, 535-553.

Segerberg, K.: 1995b, 'Some questions about hypertheories', In *Logic for a Change*, edited by Sven Ove Hansson and Wlodek Rabinowicz, Uppsala Prints and Preprints in Philosophy, 1995:9, 136-153.

Segerberg, K.: 1996a, 'Two traditions in the logic of belief: bringing them together'. *Uppsala Prints and Preprints in Philosophy*, 1996, number 9.

Segerberg, K.: 1996b, 'Further questions about hypertheories.' In *Odds and Ends*, edited by Sten Lindström, Rysiek Sliwinski and Jan Österberg, Uppsala Philosophical Studies, vol. 45, 171-184.

Segerberg, K.: 1997, 'Proposal for a theory of belief revision along the lines of Lindström and Rabinowicz', *Fundamenta Informaticae* **32**, 183-191.