Belief Revision and Dynamic Epistemic Logic

Hans van Ditmarsch, University of Otago, New Zealand, hans@cs.otago.ac.nz

In 'standard' belief revision [AGM 1985] a deductively closed theory T is revised with a formula φ resulting in a revised theory $T * \varphi$. Typically, the negation of φ is in T and in order to accommodate φ one therefore has to give up some former beliefs. A fairly recent way to model belief revision is within a setting of dynamic epistemics [Segerberg 2002, van Ditmarsch & Labuschagne 2003, Aucher 2003].

One option is to see the replacement of ignorance by knowledge that results from epistemic actions as belief revision. An agent that does not know p and learns that p, 'revises' the theory containing his ignorance $\neg Kp$ by 'retracting' that ignorance and then 'expanding' the contracted theory with his knowledge Kp. When replacing ignorance by knowledge in such ways one has to overlook some complications related to negative introspection. Can we really say that the knowledge of the agent has increased by this revision? Because some knowledge has now disappeared as well, such as the knowledge of this ignorance $K\neg Kp$ that was still there before the revision. Also, revision of objective knowledge is not possible in this setting: for that, one appears to need some notion of 'forgetting' and that can currently not be modelled in the available languages for dynamic epistemics. So another drawback of this approach appears to be that it describes *irrevocable* belief revision only.

We propose a different framework for belief revision in dynamic epistemics, namely one where one can also revise objective knowledge, and can do that repeatedly. A typical example is where an agent believes that some atomic fact p is true but when provided with evidence to the contrary revises his beliefs and having done that then believes $\neg p$. Later, he might be forced to revise his beliefs again, such that they include p once more, etc. So we have $p \in T$, and $\neg p \in T * \neg p$, and $p \notin T * \neg p$. In the dynamic epistemic setting that we propose we 'simulate' this revision as follows. In some given pointed Kripke model (M, s) for the theory T the agent *believes* p, i.e. $M, s \models Bp$; a dynamic modal operator $[*\neg p]$ corresponds to, 'all in one', contraction of the agent's belief in p and (consistent) expansion of the agent's beliefs with $\neg p$, and this operator $[*\neg p]$ is interpreted as a state transformer (binary relation) $[\![*\neg p]\!]$ such that $B\neg p$ is true in the resulting epistemic state $(M, s)[\![*\neg p]\!]$. So we end up with $M, s \models [*\neg p]B\neg p$. Features of 'standard' belief revision expressed as "if $\psi \in T$ then $\chi \in T * \varphi$" therefore correspond to 'doxastic correctness statements' of the form "$B\psi \rightarrow [*\varphi]B\chi$ is valid".

The suitability of such a dynamic revision operator (for which different semantics are outlined below) can be assessed by matching the AGM requirements to corresponding properties of the dynamic logical semantics. One can then observe the following. 'Success' can no longer be required, because the revision formula is not necessarily an objective formula. A typical example is to attempt 'revision' with $p \wedge \neg Bp$: after that you believe that p so it is no longer the case that you don't believe it. Instead, one may demand success on objective formulas only. AGM requirements concerning inconsistency, that are mainly relating to expansion only, all go up in smoke. An inconsistent theory $T + \varphi$ is a theory without models, whereas any state transformer $[\![+\varphi]\!]$ either induces a relation between models, or not. Consistency requirements have meaning in the (logical) 'linguistic' setting of standard belief revision, but not in a 'structural' setting of dynamic modal operators interpreted on relational structures.

The contraction of a theory with $\neg\varphi$ that may be necessary to accommodate consistent expansion with φ is generally carried out relative to an ordering of either semantic or syntactic objects (states, formulas, theories): a 'system of spheres' or some 'preference relation' [Lewis 1973, Grove 1988, Spohn 1988]. The innermost sphere or most preferred states correspond to what is normally believed. An original idea by [Labuschagne 2002] allows for an 'outermost sphere' of 'inaccessible states' that are infinitely less preferable than any other. Within this setting we can now define both different degrees i of belief B^i such that $B^i\varphi \rightarrow B^j\varphi$ iff $i \geq j$, and also knowledge K as true belief of an arbitrarily high degree. The static picture thus becomes a variation of the [Kraus & Lehmann 1988] interpretation of knowledge as true belief. We now outline its semantics. For simplicity, we only consider the single agent version in this abstract. All our results are for any finite number of agents.

Given a domain S, and a state s in the domain, an agent's preference relation is a well-founded discrete total order, and more particularly: a partial function $<^s$ from S onto an initial fragment of $\mathbb{N}$. In other words: there are no empty levels. We need a countably infinite number of levels, for reasons related to a multiagent setting. We overload $<^s$ as both binary and unary operation. Accessibility relations $\rightarrow^i$ are the smallest that are induced by: for all i, if $<^s(s') = i$ then $s \rightarrow^i s'$, and $\rightarrow^i \subseteq \rightarrow^{i+1}$. Relations $\rightarrow^i$ define corresponding modal operators B^i in the obvious way: in a state s, a proposition $B^i\varphi$ holds if and only if φ holds in all states s' such that $s \rightarrow^i s'$. For B^0, we write B, as this corresponds to 'normal belief'. 'Knowledge' K is an infinitary modal operator with accessibility $\rightarrow^\omega := \bigcup_{i \in \mathbb{N}} \rightarrow^i$. If we additionally require that all accessibility relations $\rightarrow^i$ are serial, and positively and negatively introspective, and that, for all states s in the domain, $<^s(s) \in \mathbb{N}$ (i.e., $\rightarrow^\omega$ is reflexive), then all belief operators B^i satisfy the 'standard' properties of belief and K the 'standard' properties of knowledge. Instead of $s \rightarrow^\omega s'$ we now write $s \sim s'$. Subject to additional multimodal interaction – namely $B^i\varphi \rightarrow B^jB^i\varphi$, for $j > i$ – the preference relation $<^s$ becomes independent from s, so we may write $<$. We can think of $<$ as a set of disjoint total orders that represent epistemic equivalence classes. Our structures therefore are *doxastic epistemic models* $\langle S, <, V\rangle$ with induced access $\rightarrow^0$, $\rightarrow^1$, ..., $\sim$. Figure 1 gives an example.

The dynamic revision $*\varphi$ is seen as *tentative* public announcement of φ:

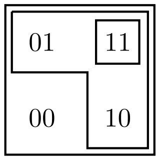

Figure 1: A doxastic epistemic model $M = \langle \{00, 01, 10, 11\}, <, V \rangle$ for a single (anonymous) agent and two facts p and q, such that $11 < 01 = 10 < 00$, and $V_p = \{10, 11\}$ and $V_q = \{01, 11\}$. For example, $M \models B(p \wedge q)$, $M \models B^1(p \vee q)$, and $M \models K\varphi \leftrightarrow B^2\varphi$.

both φ and $\neg\varphi$ may be true, but it is thought more likely that φ is true. It is therefore a nondeterministic action. For $M = \langle S, <, V \rangle$ we define

$$\langle S, <, V \rangle, s \models [*\varphi]\psi \text{ iff } \langle S, <^*, V \rangle, s \models \psi$$

where preference $<$ is 'changed relative to φ' resulting in $<^*$. We give five examples $<^{*1}, <^{*2}, ...$ of that. The definitions involve a 'normalization' – the normalization of some $<$ is written as $\overline{<}$ – that removes possible gaps that have appeared between levels. Some explanations follow the definitions, and we suggest to apply the definitions to Figure 1 and thus get the results of Figure 2.

- $<^{*1} = \overline{<'}$, where $<'(s) = <(s)$ if $M, s \models \varphi$ and else $<'(s) = <(s) + 1$
- $<^{*2} = \overline{<'}$, where $<'(s) = <(s)$ if $M, s \models \varphi$ and else $<'(s) = \omega$
- $<^{*3} = \overline{<'}$, where $<'(s) = max(<(s), 0)$ if $M, s \models \varphi$ and else $<'(s) = max(<(s), 1)$
- $<^{*4} = \overline{<'}$, where $<'(s) = <(s)$ if $M, s \models \varphi$ and else $<'(s) = max\{<(s') \mid M, s' \models \varphi\} + 1$
- $<^{*5} = \overline{<'}$, where $<'(s) = <(s) - min\{<(s') \mid M, s' \models \varphi\}$ if $M, s \models \varphi$ and else $<'(s) = <(s) + 1 - min\{<(s') \mid M, s' \models \neg\varphi\}$

Revision $*^1$ is due to [van Ditmarsch 2004] and considers the source of the revision formula equally reliable as the revising agent, unlike the standard assumption in belief revision that an agent only wants to revise with information that it considers *more* acceptable than its current beliefs. Revision $*^1$ is therefore not (necessarily) successful on objective formulas. Revision $*^2$ is standard public update, where we write $<(s) = \omega$ to denote that those states have become inaccessible (definitely excluded). Revision $*^3$ is a possibly surprising generalization of the standard [Baltag, Moss, Solecki 2003] framework, namely for more than one modal operator per agent. See also below. Revision $*^4$ is due to [van Benthem 2003] where it was suggested within a slightly different context of conditional implication; revision $*^4$ is not entirely faithful to that proposal. Revision $*^5$ is a specific action that fits the general framework by [Aucher 2003] that is an admirable dynamic epistemic setting according to [Baltag, Moss & Solecki] of a proposal originally found in [Spohn 1988].

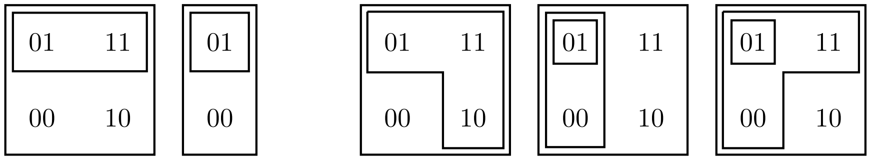

Figure 2: The induced state transformation $[\![*^i\neg p]\!]$ by belief revision with $\neg p$ on the model in Figure 1 for all five distinguished forms of belief revision $*^1, ..., *^5$ from, respectively, left to right in the figure. Note that only $*^2, *^4, *^5$ are successful. Typical examples of formulas valid in the original model M are $[*^1\neg p]Bq$, $[*^2\neg p]K\neg p$, $[*^3\neg p]B(p \vee q)$, $[*^4\neg p]\neg B^1 p$, $[*^5\neg p]B(\neg p \vee q)$.

For all these dynamic belief revisions, one can then consider interaction principles describing the relation between belief, knowledge, and dynamic revision. With the exception of $*^3$ (for which an alternative definition without normalization exists) these interactions are unclear at this stage. This is even so in the case of $*^4$ and $*^5$. We cannot straightforwardly use the suggestions and results of these authors because of the complication of normalization. Apart from the five defined here, many other ways to model dynamic belief revision are conceivable, e.g. according to two other proposals also found in [Spohn 1988].

This dynamic framework for belief revision can be generalized to arbitrary 'doxastic epistemic actions' following either the approach of [Baltag, Moss & Solecki 2003] or that of [van Ditmarsch, van der Hoek & Kooi 2003]. In the first case we can model dynamic belief revision as a restricted modal product of a doxastic epistemic (static) model with a doxastic epistemic action model. This was observed by [van Ditmarsch] and in a somewhat different setting independently by [Aucher 2003]. For an example, note that $*\varphi$ can be seen as a nondeterministic action model consisting of two simple actions α and β with preconditions φ and $\neg\varphi$, respectively, where we allow preferences among actions similar to those among states, such that α is preferred to β. In the second case, we can see belief revision as a basic action constructor $*\varphi$ defined as before as a state transformer resulting in changed preferences *but now resulting in a state with empty access for all agents*. The preferences are only 'activated' by learning operators such as L_N. Surprisingly, we appear then to be able to plug in all the previously distinguished notions $*^1$, ..., $*^5$ again. The constructor $*\varphi$ is therefore a generalization of the 'test' operator $?\varphi$ in that language, and test $?\varphi$ has now become the special case $*^2\varphi$: 'public belief revision $*^2\varphi$' as above is within this more general action language described as $L_N *^2 \varphi$, i.e. as $L_N ?\varphi$.

The $\mathbb{N}$ degrees of preference proposed in this abstract, can be generalized to the mere assumption of a total order, that allows for fully qualitative reasoning. The observed frame correspondences and different revision operators match that generalization well. Further details and references can be found in 'Prolegomena to dynamic belief revision', available on `www.cs.otago.ac.nz/staffpriv/hans/`.