Towards a “sophisticated” model of belief dynamics*

Brian Hill

26th October 2006

Abstract

It is well-known that classical models of belief are not realistic representations of human doxastic capacity; equally, models of actions involving beliefs, such as decisions based on beliefs, or changes of beliefs, suffer from similar inaccuracies. In this paper, a general framework is presented which permits a more realistic modelling both of instantaneous states of belief, and of the operations involving them. This framework is motivated by the inadequacies of existing models, which it overcomes, whilst retaining technical rigour in so far as it relies on known, natural logical and mathematical notions. As an illustration of this framework, it will be applied to the particular case of belief revision. A model of belief revision shall be obtained which, firstly, recovers the Gärdenfors postulates in a well-specified, natural yet simple class of particular circumstances; secondly, can accommodate iterated revisions, recovering several proposed revision operators for iterated revision as special cases; and finally, offers an analysis of Rott’s recent counterexample to several Gärdenfors postulates [23], elucidating in what sense it fails to be one of the special cases to which these postulates apply.

Keywords  Representations of belief, bounded rationality, logical omniscience, awareness, logical locality, belief dynamics, iterated revision, Gärdenfors postulates, rational choice theory, framing effect.

For several years now, the “realism” of the classical representations of belief proposed by logicians, philosophers, and economists has been the source of anxiety and debate. The realism of the models of doxastic actions which rely on such representations, such as those models proposed by decision theory, choice theory, and, more recently, belief revision, has given rise to similar worries. The purpose of this paper is to propose and motivate a framework which supports a more realistic model of doxastic states, of the changes they undergo, and of the role they play in action and decision. This framework shall be developed and applied to the case of belief revision, a paradigm example of an operation involving beliefs, and a field which has recently seen some concern about the realism of traditional approaches.

Indeed, belief revision illustrates nicely the two aspects of the problem of realistically modelling beliefs and the operations in which they are involved. One concerns the representation of states of belief: classical models – be it as sets of possible worlds, or sets of sentences closed under logical consequence – which are accepted in traditional

*Submitted to Studia Logica, October, 2006
models of belief revision [10], suffer from well-known problems such as “logical omniscience”, in so far as, loosely speaking, they imply that an agent believes all the consequences of his beliefs [14, 6]. A second problem concerns the models of operations involving beliefs: in the case of belief revision, it is the traditional Gärdenfors postulates for belief revision, and corresponding models, which come under attack as unrealistic. To take but one important example, Rott has recently given a purported counterexample to two of the basic Gärdenfors postulates for belief change [23]. His counterexample is intended to cast doubt on the “realism” of these postulates, and thus motivate the search for “more sophisticated models of belief formation”.

This paper takes up Rott’s challenge, setting it in a wider stage. It assumes the view that, given that the point of modelling belief is to use such models to capture what happens to, and with, belief, and given furthermore that every model of what happens to, and with, belief relies on a model of the states of belief, the question of representing belief and the question of representing operations involving beliefs are closely related. A model of a belief state will not be “sophisticated” if it cannot support an equally “sophisticated” model of belief dynamics or the role of belief in decision and action; conversely, a “sophisticated” model of belief change or of the role of belief in action shall have to rely on a “sophisticated” model of the belief state at a particular moment. The challenge is to propose a representation of belief which not only models accurately the doxastic state of an agent at a particular moment, but also permits a realistic model of the changes and effects of these beliefs. The proposed framework shall have the machinery to deal with the alleged insufficiencies of traditional theories, both with respect to beliefs themselves, and with respect to the use and development of these beliefs. In the case of belief revision, to which the proposed framework shall be applied, this means that it provides not just “sophisticated models of belief formation”, but equally “sophisticated models of belief states”.

The main concern of this article is to present the technical framework and to illustrate it on the particular case of belief revision; accordingly, a general discussion of the perspective taken on beliefs, of the notion of “realistic” model, of the conception of idealisation, and of related topics shall be reserved for another place. Suffice it to say that the proposed model shall be an improvement on the traditional framework, as far as the question of realism is concerned, in two senses. Firstly, the motivation and application of the model avoids the aforementioned problems with the traditional frameworks. Secondly, the traditional models can be recovered as special cases: in this concrete sense, they appear as idealisations of the more general, and realistic, framework proposed here.

In the first part of the paper, the general framework shall be developed in two stages – firstly a representation of the instantaneous state shall be proposed and motivated, then an operation capturing the dynamics of this state shall be defined. In the second part of the paper, the framework shall be applied to the case of belief revision. A model of belief revision shall be obtained which, firstly, recovers the Gärdenfors postulates as applying in particular circumstances; secondly, can accommodate iterated revisions, recovering several proposed revision operators for iterated revisions as special cases; and finally, offers an analysis of Rott’s aforementioned counterexample to several Gärdenfors postulates [23], elucidating in what sense it fails to be among the circumstances to which these postulates apply.
1 General framework

All systems purporting to represent beliefs or operations involving them assume an underlying language, with its own logic (for the most part, the classical consequence relation). The fundamental observation motivating the proposed model of beliefs is that, between any two moments, the languages which are effective or “in play” at these moments – the languages in which the beliefs active at these moments are couched – may differ. A similar point seems to hold for the logics of these languages, in so far as they are comparable. Let us call the combination of language and logic effective at a particular moment, the local logical structure at that moment. The model developed shall be more “realistic” or “sophisticated” in that it pays explicit attention to and indeed represents formally the local logical structures effective at particular moments, as well as the changes in these structures as new information comes into the fray.

In the first part of this section, a model of the local logical structure shall be motivated and proposed; in Section 2, this model shall form the base of the representation of the beliefs involved at that moment. Given that the local logical structure varies over time, modelling the local logical structure at a particular moment will not suffice: any interesting approach must also capture the change in local logical structure when new information comes into play. In the second part of this section, an operation shall be defined which shall be used to represent such changes; it shall be used, in Section 2 to model belief dynamics.

1.1 Interpreted Algebras

The technical notion used to model local logical structures – interpreted algebras – can be motivated in two stages, loosely corresponding to two sorts of failure of logical omniscience. Firstly, there is the important notion of a sentence or an issue being in play at a particular moment. If, at time $t$, an agent believes (actively or explicitly) that he has a meeting at 10.00, without apparently believing that he has a meeting at 10.00 and there are infinitely many primes, it is not just that there is no belief that there are infinitely many primes, but rather that the whole question of the number of primes – the sentence “there are infinitely many primes”, if you prefer – is out of play at time $t$. To give a more mundane example, if the agent forgets to go to his meeting, it is not that at 10.00 he believes there is no meeting, nor that at 10.00 he neither believes that there is a meeting nor that there is no meeting, but rather that the subject of the meeting doesn’t cross his mind”. It doesn’t enter “into play”.

Traditional models of belief cannot account for this notion of ‘in play’, because they can only allocate one of three (doxastic) statuses to each sentence of the language – believed to be true, believed to be false, or neither – and the notion of ‘in play’ is orthogonal to this triple distinction. Attempts to account for such phenomena have generally consisted, somehow or other, in bringing in some sort of syntactic apparatus to represent the sentences which are in play. Fagin and Halpern’s awareness models [6] are classic examples of this: they extend the ordinary Hintikka-styled models of belief with sets of sentences of which the agent is “aware”, and allow (explicit) belief only relative to these sentences. This is the moral to be taken from these sorts of examples: to capture the local logical structure effective at a given moment, it is necessary to render explicit the sentences in play at that moment. The set of sentences in play at a particular moment shall henceforth be called the local language (at that moment).

A second important aspect of logical omniscience concerns the failure to recognise the logical equivalence of sentences. An agent may accept that he needs to go to the
eye-doctor without accepting that he need go to the ophthalmologist, despite the fact that the two sentences are (intensionally) equivalent. This too cannot be represented in the traditional framework, since these two sentences are instances of the same proposition, or, to put it another way, they have the same truth values in all possible worlds. Nor will the addition of sets containing sentences of which the agent is aware help in this case, since the subject is aware of both sentences; he just does not recognise the logical relationship between them. Suggestions for dealing with this sort of example often consist of alterations to the semantic structure of the agent’s beliefs. The typical example, dating back to [26], and proposed in [6] under the name of the “society of minds model”, models the belief state of the agent as a set of (consistent) sets of beliefs (that is, a set of sets of possible worlds). The agent’s belief state is fragmented into consistent “clusters” – or “minds” – the simple union of which will usually be inconsistent, thus his lack of logical omniscience. For example, in one of the agent’s “minds” (sets of beliefs), he believes that he needs to see the eye-doctor, in another, he does not believe he needs to see the ophthalmologist; the fragmentation into these consistent “minds” prevents “logical” conflict between contradictory beliefs.

The moral of this sort of example shall be taken to be the need, when modelling local logical structures, to represent accurately the logical relationships between sentences, in so far as they figure in that local logical structure at that moment. The example shows that the local logical structure does not necessarily respect the global logical structure pertaining to some global language.

These two aspects of the local logical structure – the set of sentences in play and the logical relationships between them in so far as they figure in the local logical structure – shall be captured by a model with a two-levelled structure. In the (classical) propositional case which shall be considered here, the appropriate structure, called an interpreted algebra, is defined as follows.

**Definition 1 (Interpreted Algebra).** An interpreted algebra \( \mathcal{B} \) is a triple \((B_1, B, q)\), where:

- \( B_1 \) is the free Boolean algebra generated by a set \( I \) (the interpreting algebra);\(^1\)
- \( B \) is a Boolean algebra (the base algebra);
- \( q : B_1 \to B \) is a surjective Boolean homomorphism.

An element of \( \mathcal{B} \) is a pair \((\phi, q(\phi))\), \( \phi \in B_1 \). Elements of an interpreted algebra shall be referred to (without risk of confusion) by the appropriate elements of the interpreting algebra, and, for this reason, shall often be called “sentences”. The set of elements of \( \mathcal{B} \) shall be denoted \( |\mathcal{B}| \). \((|\mathcal{B}|, \wedge, \neg, \top)\) is the Boolean algebra on \(|\mathcal{B}|\), induced by the Boolean structure on the interpreting algebra \( B_1 \) (the ordinary Boolean connective symbols \( \wedge, \vee, \neg \) shall be used when speaking of this structure, with \( \wedge \) and \( \neg \) taken as primitive for the purposes of definitions and proofs).

Finally, the consequence relation \( \Rightarrow \) on \(|\mathcal{B}|\) is defined as follows:

\[
\forall \phi, \psi \in |(B_1, B, q)|, \phi \Rightarrow \psi \iff q(\phi) \leq q(\psi)
\]

\( \Rightarrow \) will designate the derived equivalence relation.

\(^1\)A Boolean algebra is a distributive complemented lattice; the order will be written as \( \leq \), meet, join, complementation and residuation as \( \wedge, \vee, \neg \), the top and bottom elements as \( \top \) and \( \bot \). The free Boolean algebra generated by a set \( X \) shall be noted as \( \mathcal{B}_X \) for the rest of the paper; details on this and the other notions used in this paper may be found in [17].
**Explication**  This structure can be understood in the following way.

The interpreting algebra models the local language effective at the moment in question, with \( I \) being the set of *locally* atomic sentences in play at that moment. It is, so to speak, the “syntax” of the local logical structure.

The base algebra is the *local logic* on this language. It is, so to speak, the “semantics” of the local logical structure. Just as the elements of the interpreting algebra may be thought of as the *sentences* of the local logical structure, the elements of the base algebra may be thought of as the (local) *propositions*. Accordingly, \( q \) is the map taking sentences to propositions, and may be thought of as the *valuation* of the sentences of the language.\(^2\) Elements of the interpreted algebra consist of a sentence and the proposition which it expresses; the Boolean connectives are operators on the sentences (and, via \( q \), on the propositions), whereas the consequence relation on elements of the interpreted algebra arises from relations between the propositions they express. Non-trivial logical equivalences between sentences (cf. the eye-doctor / ophthalmologist example above) are modelled by the fact that the two sentences express the same (local) proposition; that is, they are mapped to the same element of \( B \) by \( q \).

Given that the local logical structure is, in practice, finite, the interpreting algebra, and thus the base algebra, are finite (see below), and therefore atomic.\(^3\) The atoms of the base algebra can be thought of as “states” or “small worlds” – worlds in the sense that every sentence of the local language receives a valuation in each world (thanks to \( q \); *small* in the sense that only the sentences of the local language receive any valuation in these worlds. Although the algebraic perspective on the semantics (ie. treating \( B \) as the algebra of local propositions) shall prove more fruitful for modelling the dynamics of local logical structures, the extensional view (ie. treating \( B \) as the sets of small worlds) will sometimes permit a more concise – and for some, more intuitive – discussion.

It is worth mentioning several assumptions that are – and aren’t – implicit in this model and in particular, in the use of a free Boolean algebra \( (B_1) \) to model the local language. Firstly, the local language is assumed to be closed under ordinary linguistic connectives such as “and”, “or”, “not”. Philosophically, this seems a harmless assumption: if the sentence “you have a meeting at 10.00” is in play, and the sentence “there are infinitely primes” is in play, then the sentence “you have a meeting at 10.00 and there are infinitely primes” is in play.

Secondly, and more importantly, in this model, these connectives on elements of the interpreted algebra retain their ordinary Boolean properties; so for example, if the sentence “you have a meeting at 10.00” is in play, then the sentence “you do have a meeting at 10.00 and you don’t have a meeting at 10.00” is contradictory. This assumption is both technically and philosophically less audacious than it may seem at first.

Technically, a more evident model of the local language, which does not involve these assumptions, is a so-called “term algebra”.\(^4\) However, since such an algebra would have to be mapped into the Boolean algebra \( B \) (the base algebra) in such a way that the connectives are taken to the Boolean operators, this mapping ends up factoring

---

\(^{2}\) The fact that it is a Boolean homomorphism guarantees that the ordinary conditions on valuations are satisfied.

\(^{3}\) Standard terminology is employed here: an atom of a Boolean algebra is an element \( a \in B \), such that, for all \( x \in B \), \( \bot \leq x \leq a \), either \( x = \bot \) or \( x = a \). Note furthermore that the assumption of finiteness is not required for any of the definitions or results in this paper; the weaker assumption that \( B \) is atomic is sufficient.

\(^{4}\) For example, [4, Ch 5].
through the interpreting algebra $B_I$. Most of the framework presented here can be expressed, in a messier and less transparent fashion, in terms of term algebras.

As far as the philosophical—or “realistic”—credentials of this model are concerned, one might question the extent to which the connectives involved at a particular moment do satisfy the ordinary Boolean conditions. To ease this worry, it should be noted that no suppositions have been put on the set $I$ of locally atomic, or primitive, sentences. These sentences are those taken to be primitive at the particular moment and in the particular situation in question, and certainly not in any deeper or larger sense. Therefore, the model can account for a connective featuring in a sentence which does not obey the Boolean properties by taking the whole sentence, rather than the appropriate clauses, as an element of $I$. If “it is cold and it is wet” is in play, it does not necessarily mean that “it is cold” and “it is wet” are individually implied; indeed, they may not even be in play as separate linguistic entities. The assumption that the connectives satisfy the ordinary Boolean conditions is rather weaker than it seems at first blush, since it only applies to those connectives connecting sentences which themselves figure in the local language.

In defence of the assumption, it does have the advantage of reducing the size of the local language: although recursive application of Boolean connectives on the elements of (a non-empty set) $I$ yields an infinite set, the set of equivalence classes under Boolean equivalence — that is, the set of elements of $B_I$ — will be much smaller, and may be finite. This seems a faithful rendering of the intuition that the local language at a particular moment is, in practice, finite: not only are there a finite number of (locally) primitive sentences in play ($I$ is finite), but there are effectively only a finite number of linguistic entities which can be formed from them, since one naturally discounts such differences as those between ‘$A$ and $A$ and $A$ and $A’ and ‘$A’ As noted above, in this paper, the interpreted algebras shall be taken to be finite.

Furthermore, the model of the local language as a free algebra ensures that no relationships on the elements of the algebra other than the ordinary Boolean ones are assumed. If a non free algebra were used, it would represent the sentences as entering into such non trivial (i.e. non Boolean) relationships; by using a free algebra, it follows that the only non trivial relationships between sentences are those arising from the relationships between the local propositions they express, that is, those expressed in terms of the consequence $\Rightarrow$.

Here are three examples of basic, but important, sorts of interpreted algebra.

**Example 1.** A trivial interpreted algebra $B$ is an algebra of the form $(B_I, 0, q)$, where 0 is the one-element Boolean algebra $(\top = \bot)$ and $q : B_I \mapsto \top$.

The point interpreted algebra for the sentence $\phi$, $B_{\phi} = (B_I, 1, q)$, where 1 is the two element Boolean algebra $(\{\top, \bot\})$, and $q : \phi \mapsto \top$.

The simple interpreted algebra for the sentence $\phi$, $B_{\phi} = (B_I, 2, q)$, where 2 is the four element Boolean algebra $(\{\top, \bot, x, x'\})$, and $q : \phi \mapsto x$.

Trivial algebras are the inconsistent local logical structures. All (and only) the sentences of the local language are equivalent to the (local) true (and, equally, to the local false).

Point algebras and simple algebras are the two basic possibilities for representing a (consistent) local logical structure which has essentially one sentence $(\phi)$ in play (that is, there is the one sentence and those which can be formed from it with Boolean connectives). In the point algebra, this sentence is accepted as a (local) logical truth in

---

5 Recall (footnote 1) that $B_I(\phi)$ is the free Boolean algebra generated by $\{\phi\}$. 

©Brian Hill Towards a “sophisticated” model of belief dynamics
the language (in terms of small worlds, there is one world, where $\phi$ holds, $\neg\phi$ holding at no world in this interpreted algebra). The simple algebra admits the “possibility” that the sentence may be true as well as false (there are two worlds, one where $\phi$ holds, the other where $\neg\phi$ holds).

Finally, it is easy to see that this sort of model accommodates naturally the two sorts of phenomena related to the lack of logical omniscience described above. On the one hand, the fact for sentences to be in or out of play (cf. the example of the meeting and the number of primes) is captured by their presence or absence with respect to the interpreting algebra. On the other hand, logical inconsistencies relative to sentences which are both in play is captured by the structure of the base algebra (and the homomorphism into this algebra): in such an algebra, the global logical relationship between sentences (those referring to eye-doctor and those referring to ophthalmologist express the same propositions) is not respected. Although it will divert us from the topic of this paper to elaborate the point, it should be noted that this model has the advantage over, say, the families of models proposed in [6], of capturing both of these phenomena in a single framework. A more important advantage lies however in the ease of modelling changes in the local logical structure—a question which is seldom posed in the literature on logical omniscience and realistic models of belief. This is the question which shall now be addressed.

1.2 Fusion

Investment in a model which captures the logical imperfectness of an agent’s instantaneous belief state seems worthless if it is not accompanied by an account of how this state can change. If a model captures the “awareness” of an agent at a moment, it should surely account for the frequent changes of awareness from one moment to the next. If it models the agent’s “several minds”, it should equally yield an understanding of how these “minds” interact with each other over time. In terms of the framework proposed here, a proposed model of the local logical structure at particular moments is of little use unless it can also model the changes in the local logical structure which occur from one moment to the next. In this section, a fusion operation shall be defined which will model the change in the local logical structure as new information comes into play.

The changes to local logical structures which shall be dealt with here are those brought about by the incoming information. Typically, in models of belief (or knowledge) and their changes, new information comes in the form of a sentence (or set of sentences) of the language. However, no global or overarching language is assumed in the current framework; indeed, given that the only language present is the local language of the current local logical structure, the whole problem is how to deal with sentences which do not necessarily belong to this language. It is therefore necessary to endow the incoming information with its own fragment of language, with the sort of basic logical structure which always accompanies such fragments of language. To put it another way, the new information comes in the form of (at least) a local language with a local logic. It shall thus be modelled using interpreted algebras.

---

6 This is the case not only in belief revision [10], but equally in epistemic dynamic logic (for example, the case of public announcements [8, 28]), or in typical Bayesian update theory [24].

7 It should be recalled that the current discussion concerns the general framework: in Section 2.2.2, a richer model of new information, obtained by adding extra structure to the interpreted algebra, shall be proposed.
The flexibility of the notion of interpreted algebra permits it to capture the diverse, more or less complicated, forms which incoming information might take. At one end of the spectrum, rich local languages (large \(B_1\)) with interesting logical structures \((B, q)\) can accurately model an input which does not consist of a simple sentence, but comprises a complex of diverse information, about how such a sentence comes into play, how it was learnt, what justifies it, and so on. At the other end of the spectrum, the simple traditional cases of a single sentence entering into play can be captured using simple or point interpreted algebras (Example 1).

Once new information is modelled by an interpreted algebra, the change in the face of new information takes the form of an operation (or operations) sending a pair of interpreted algebras – one modelling the current local logical structure, the other the new information – to another interpreted algebra – the local logical structure resulting from the input of the new information. There is, however, one final aspect that needs to be taken into account before such an operation can be proposed.

Given that no overarching language is assumed, but only the local languages contained in the individual interpreted algebras, there is a priori no way of identifying sentences belonging to different interpreted algebras. However, when new information comes into play, it could certainly be the case that some of the sentences involved belong to the local logical structure already in play. The representation of the prior local logical structure and the new information by interpreted algebras cannot account for the fact that the two may have certain sentences in common: supplementary technical apparatus is required to capture this. The identification of sentences between different interpreted algebras shall be represented using an appropriate relation – called identification – defined as follows.

**Definition 2 (Identification between algebras).** For \(B_1 = (B_{11}, B_1, q_1)\), \(B_2 = (B_{12}, B_2, q_2)\) two interpreted algebras, \(\simeq \subseteq |B_1| \times |B_2|\) is an identification of sentences between the two algebras if:

(i) \(\simeq\) is the restriction of an equivalence relation on \(|B_1| \uplus |B_2|\), to \(|B_1| \times |B_2|\), where \(\uplus\) is the disjoint union;

(ii) \(\simeq\) respects the Boolean connectives:

- if \(\phi \simeq \phi', \psi \simeq \psi'\), then \(\phi \land \psi \simeq \phi' \land \psi'\);
- if \(\phi \simeq \phi'\), then \(\neg \phi \simeq \neg \phi'\);

(iii) \(\simeq\) is generated by a relation \(\simeq \subseteq I_1 \times I_2: \phi \simeq \psi\) iff, for \(\phi = \bigvee_{m \in M} (\bigwedge_{i \in I_{1m}} \phi_i)\), \(\psi = \bigvee_{m \in N} (\bigwedge_{k \in I_{2m}} \psi_k)\), \(\phi_i \in I_1, \psi_k \in I_2\), the normal disjunctive conjunctive forms for \(\phi\) and \(\psi\), there is a bijection \(\mu\) between \(M\) and \(N\), and bijections \(\sigma_m\) between \(I_{1m}\) and \(I_{2m}\) for every \(m \in M\), such that, \(\forall m \in M, \forall j \in J_m, \phi_j \sim \psi_{\sigma_m(j)}\).

**Remark 1.** Clause (i) is expressed in this form because it clarifies the extensions of the notion to cases of more than two interpreted algebra and to the case of a single interpreted algebra (see below). An equivalent formulation would be, for \(\phi, \phi' \in |B_1|\), \(\psi, \psi' \in |B_2|\), if \(\{\chi | \phi \simeq \chi\} \cap \{\chi | \phi' \simeq \chi\} \neq \emptyset\), then \(\{\chi | \phi \simeq \chi\} = \{\chi | \phi' \simeq \chi\}\); and if \(\{\pi | \pi \simeq \psi\} \cap \{\pi | \pi \simeq \psi'\} \neq \emptyset\), then \(\{\pi | \pi \simeq \psi\} = \{\pi | \pi \simeq \psi'\}\).

\(8\) I.e., \(\phi\) and \(\psi\) have the same disjunctive conjunctive form, with equivalent elements of \(I_1\) and \(I_2\). For the existence of disjunctive conjunctive form in free Boolean algebras, see any basic logic textbook. This condition guarantees that the top elements of the two algebras are identified: \(\top \simeq \top\).
In keeping with the spirit of this definition, a notion of identification relation on sentences of a single interpreted algebra may be defined as follows.

**Definition 3.** An identification relation on $B$ is a congruence relation on $(|B|, \wedge, \neg, \top)$ generated by an equivalence relation on $I$, in the sense of (iii) above.¹⁰

An extended discussion of the philosophical consequences of identifications between local logical structures will have to be left for another paper. Here it is only important to understand the consequence for the question of changes in the local logical structure in the face of new information: the task now becomes that of proposing an operation taking two interpreted algebras, with an identification of sentences between them, and yielding an interpreted algebra which respects the identification of the sentences. The operation of fusion of interpreted algebras does just this. It can be defined from two simple operations on interpreted algebras (which are little more than operations on Boolean algebras lifted to interpreted algebras).

The first is the operation of free product on interpreted algebras.

**Definition 4 (Free Product of interpreted algebras).** The free product of interpreted algebras $A_1 = (B_1, B_1, q_1)$ and $A_2 = (B_2, B_2, q_2)$ is $B_1 \otimes B_2 = (B_{I_1 \cup I_2}, \otimes, B_1 \otimes B_2, q_1 \otimes q_2)$, where $\otimes$ is the free product on Boolean algebras and Boolean homomorphisms respectively.¹¹

At the level of languages, the new local language obtained is the closure under Boolean operations of the disjoint union of the two initial local languages. On the semantic side, the set of small worlds or states in the resulting interpreted algebra is the cartesian product of the sets of small worlds or states of the initial algebras¹²; the valuation on these worlds (the homomorphism $q$) is the naturally derived valuation. The free product adds the initial local languages, without identifying any of the sentences, and combines the small worlds of the initial algebras, so to speak, to obtain “enriched” small worlds, without imposing any additional logical structure on these worlds.

The operation used to identify or render identical elements of algebras is the well-known operation of quotient. In the current case, it is defined as follows.

**Definition 5 (Quotient of an interpreted algebra).** The quotient of an interpreted algebra $B = (B_1, B, q)$ by a identification relation $\sim$ on $B$, is the algebra $B/\sim = (B_1/\sim, B/\sim^q, q_\sim)$, where

- $B_1/\sim$ is the quotient of the ordinary Boolean algebra $B_1$ by the congruence relation $\sim$;
- $\sim$ is the congruence relation on $B_1$ induced by the relation $\sim$ on elements of $B$ $(\phi \sim \psi$ iff $(\phi, q(\phi)) \sim (\psi, q(\psi)))$.

¹⁰That is: there is an equivalence relation $\sim$ on $I$, with $\phi \sim \psi$ iff, for $\phi = \bigvee_{m \in M} (\bigvee_{j \in J_m} \phi_j)$, $\psi = \bigvee_{n \in N} (\bigvee_{k \in K_n} \psi_k)$, $\phi_j, \psi_k \in I$, the normal disjunctive conjunctive forms for $\phi$ and $\psi$, there is a bijection $\mu$ between $M$ and $N$, and bijections $\sigma_m$ between $J_m$ and $K_{\mu(m)}$ for every $m \in M$, such that, $\forall m \in M, \forall j \in J_m, \phi_j \sim \psi_{\mu(m)}(j)$.

¹¹$B_{I_1 \cup I_2} = B_{I_1} \otimes B_{I_2}$, so the free product of interpreted algebras is just the free product of the interpreting and base algebras, with the free product of the homomorphisms. This is thus a well-defined operation on interpreted algebras. For details on the technical notions, see [17].

¹²For atomic algebras $B_1$ and $B_2$, with sets of atoms $S_1$ and $S_2$ respectively, $B_1 \otimes B_2$ is atomic with its set of atoms isomorphic to $S_1 \times S_2$. 
- $\simeq^q$ is the smallest congruence relation on $B$ containing $\simeq'$, where $X \simeq' Y$ iff
  \[ \exists \phi, \psi \in B_1, q(\phi) = X, q(\psi) = Y, \text{ and } \phi \simeq \psi; \]
- $q_{\simeq'}([\phi]) = [q(\phi)]_{\simeq'}$.

**Observation 1.** The quotient of an interpreted algebra by an identification relation is well-defined.\(^\text{13}\)

Two different elements of a Boolean algebra which are related by a congruence relation are taken, under the quotient with respect to this relation, to the same element in the resulting algebra. In terms of local languages, the quotient operation identifies the sentences which were $\simeq$-equivalent in the initial local language; that is to say, such sentences in $B_l$ have a common image in $B_l/ \simeq$. In terms of the local logic, the propositions corresponding to sentences which are $\simeq$-equivalent in the initial local logical structure are identified in the resultant structure; that is, they have the same image. Equivalently, quotienting on the semantic level removes the small worlds which are witness to differences between any pair of $\simeq$-equivalent sentences $\phi$ and $\psi$; that is, worlds where the valuations of $\phi$ and $\psi$ differ.\(^\text{14}\)

The free product of two interpreted algebras puts them together, without identifying any of the sentences between them. The quotient identifies sentences which are $\simeq$-equivalent in an interpreted algebra. The operation of fusing two interpreted algebras, whilst respecting the identification of sentences between them, would seem to require exactly these two steps. Such an operation will model the change in the local logical structure in the face of incoming information. The following preliminary definition is required.

**Definition 6.** If $\simeq$ is an identification relation between $B_1$ and $B_2$, let $\simeq' \subseteq B_1 \otimes B_2$ be $\{(e_1(\phi), e_2(\psi)) : (\phi, \psi) \in \simeq\}$, where $e_1$ (resp. $e_2$) is the canonical homomorphism from $B_1$ (resp. $B_2$) into $B_1 \otimes B_2$. Let $\simeq$ be the smallest congruence relation on $B_1 \otimes B_2$ containing $\simeq'$,\(^\text{15}\)

It is easily seen that $\simeq$ is an identification relation on $B_1 \otimes B_2$. Furthermore, it is the natural identification relation on $B_1 \otimes B_2$ generated by the relation $\simeq$ between $B_1$ and $B_2$.\(^\text{16}\) It allows the following definition of the fusion of two interpreted algebras.

**Definition 7 (Fusion $\ast$).** Given two interpreted algebras $B_1$ and $B_2$, with an identification relation $\simeq$ between them, the fusion of the two algebras respecting this relation is defined as:

\[ B_1 \ast_{\simeq} B_2 = (B_1 \otimes B_2)/\simeq \]

In subsequent discussion, identification relations shall generally be assumed between appropriate interpreted algebras, and the fusions written simply as $B_1 \ast B_2$.

\(^{12}\) In other words, $\simeq^q$ is the closure under the appropriate conditions of $\simeq'$, which fails to be a congruence relation because it is not transitive.

\(^{13}\) Proofs of this and other statements are to be found in the appendix.

\(^{14}\) Formally: the $x$ with $x \leq q(\phi) \triangle q(\psi)$, where $\triangle$ is the symmetric difference, are removed in the quotient operation.

\(^{15}\) In other words, $\simeq$ is the closure of $\simeq'$ under Boolean operations, reflexivity, transitivity and symmetry.

\(^{16}\) This can be seen from the fact that $\simeq$ is unique in the following sense: if $\simeq''$ is an identification relation on $B_1 \otimes B_2$ satisfying $e_1 \times e_2(\simeq') = \simeq'' |_{e_1 \times e_2(B_1 \otimes B_2)}$, then $\simeq'' = \simeq$. 

10
This operation models the change in the local logical structure under new incoming information: both the original local logical structure and the new information are modelled by interpreted algebra; the resulting local logical structure is the resulting interpreted algebra. This model is intuitive: in fusing the new information (with its fragment of language) with the existing logical structure, the “sum” of the two languages is taken (free product), and then appropriate sentences figuring in the different languages are identified (the quotient). Given that the operation to be modelled is that of “merging” or “combining” two fragments of language, one would expect it to be commutative: no priority should be given to one over the other. The operation \(*\) has this property.

Examples and relation to the literature Two examples shall serve to illustrate this sort of operation.

Example 2. For \(\phi\) in \(B\), the fusions with the relevant simple and point algebras (Example 1) are as follows:

Simple algebra \(B \ast B_\phi\) is isomorphic to \(B\);

Point algebra \(B \ast B_{\phi q}\) is isomorphic to \((B_1, B/(\phi), q')\), where \(B/(\phi)\) is the quotient of \(B\) by the smallest congruence relation such that \(\phi \simeq T\), and \(q'\) the composition of \(q\) with the quotient homomorphism.

The first example illustrates that the fact of bringing into play a sentence which is already in play, in such a way that no extra logical structure is allocated to it, does not alter the algebra. As one would expect, the fusion operation does not change anything when the fusion is with something already present.

The second example concerns fusion with a sentence already in play, but such that the sentence, in so far as it figures as new information, is endowed with extra logical structure: namely, it is taken to be equivalent to the true (of the local language). This leads to a change in the local logical structure to accommodate this information: the fusion results in a logical structure with the same local language, but such that the sentence is now equivalent to the true (or alternatively true in all small worlds).

This second example is interesting because, put in terms of small worlds, it essentially says that fusion with \(B_{\phi q}\) does not change the language but removes all the small possible worlds, or states, where \(\phi\) is false. This sort of operation, which plays an important role in the literature on public announcement and dynamic epistemic logic [8, 28], is thus reproduced as a special case in the framework proposed here. More generally still, it is not difficult to see that the model of epistemic programs proposed by Baltag [2, 3] is based on the sort of fusion operation proposed here. Further discussion of the relationship with this work and its consequences shall be left for another paper.

The task of this section was to propose a model for the local logical structure effective at a particular moment, and of the dynamics of this structure. The notion of

---

17 Recall that only the logico-linguistic structures in which beliefs and new information are coched are being considered at this point. Fusions of such structures are expected to treat the two structures equally. This does not mean that the revision of beliefs by new information should not prioritise one over the other; indeed, the model of beliefs and new information proposed in the Section 2.2 shall give priority to the new information.

18 Leaving aside the locality of languages, which is not present in [2, 3], and modalities, which are not (yet) present in the basic framework proposed here, Baltag's Update product and the operation of Fusion defined above turn out to be essentially the same.
interpreted algebra captures the local logical structure in play, in such a way that it can account for some of the most pressing logical imperfections in our beliefs and behaviour, such as the phenomena of awareness and lack of recognition of intensional equivalence. The operation of fusion of interpreted algebras models the change of the local logical structure under incoming information.

This framework is abstract, and intentionally so. It can be applied to several different questions in several different fields; in each, the basic notions may assume a different philosophical interpretation. Examples of applications may include context and conversation, knowledge and communication, belief and decision, counterfactuals and nonmonotonic logic. And belief and its revision. The next section, intended to be an extended example of the application of this sort of framework to phenomena already considered by logicians, will propose a model of belief and belief revision which is based on the general framework introduced above.

2 Belief revision

2.1 Introduction and state of play

A model of belief revision consists of (1) a model of the belief state, (2) a representation of new information with which the state is to be revised, and (3) a representation of the revision of the former by the latter, enjoying appropriate properties. Firstly, the operation is generally taken to satisfy certain number of belief revision postulates, of which the most popular are the so-called Gärdenfors postulates [1, 10]; furthermore, one often expects the framework to be general enough to accommodate any revision operation which satisfies these postulates (it supports a representation theorem). Typically, in the AGM paradigm, the state of belief is taken to be a set of sentences (of a given language \( L \)) closed under a (given) logical consequence relation, and the new information consists of a sentence of this language \([1, 10]\). A popular model of belief revision in this paradigm, proposed by Grove [9], will serve as a useful example. It uses a (reflexive) order \( \preceq \) on the set \( S \) of maximal consistent sets of a language \( L \) – or if you prefer, possible worlds with respect to \( L \) – which has the following properties:

(S \( \preceq 1 \)) \( \preceq \) is connected \((\forall x, y \in S, x \preceq y \text{ or } y \preceq x)\);

(S \( \preceq 2 \)) \( \preceq \) is transitive;

(S \( \preceq 3 \)) \( \preceq \) is finitarily stoppered: for all \( \phi \in L, |\phi| \neq \emptyset \) implies that \( \{x \in |\phi| | x \preceq y \forall y \in |\phi| \} \neq \emptyset \), where \( |\phi| = \{x \in S | x \models \phi\} \) (that is, the set of worlds where \( \phi \) is true).

Such an order shall henceforth be called a Grove order.

In this model, the set of beliefs are the set of sentences true in the \( \preceq \)-minimal worlds. A sentence \( \psi \) is believed after revision by \( \phi \) if it is true in all the \( \preceq \)-minimal worlds satisfying \( \phi \). This model satisfies the Gärdenfors postulates (and supports a representation theorem with respect to them).  

\[^{19}\text{In AGM theory, the operation of contraction – removal of a belief – is taken as primitive. Here only the question of belief revision shall be dealt with; contraction can be recovered, if appropriate care is taken, with the help of the Harper identity. See [10] for details.}\]

\[^{20}\text{For the uninitiated, it may be useful to compare this model with Lewis' semantics for counterfactuals [18]; for a detailed comparison, see [9].}\]
Since the original models of belief revision were proposed, two other desirable properties of models of belief revision have come to fore. On the one hand, there is the question of iterated belief: it is desirable to have a model such that, whatever results from the revision of belief, it can itself be revised in the face of subsequent information. The traditional AGM models, and indeed the Grove model described above, do not satisfy this condition. The Grove model, for example, yields a set of sentences after revision, but no order \( \preceq \) on \( S \) which would be appropriate for use in further revision. Since then, several models supporting iterated revision, and indeed various postulates on iterated revisions which these models satisfy, have been proposed [5, 25, 19, 16, 22]. It is a generally accepted desideratum for models of belief revision that they permit iterated revisions, and therefore satisfy at least some of the proposed postulates.

A second sort of development has already been mentioned: the question of the realism of the proposed theories of belief revision. Doubts over this issue have taken several forms: some of the debates have been localised to the validity of particular postulates [20, 21], whereas other authors have seen a general need for more “sophisticated” or “realistic” models which give a general account of the apparent failings of certain postulates in particular situations and the apparent idealisations underpinning traditional models. Hansson [12] counts as the first problem of belief revision that of finding models which are more faithful to the finiteness of agents; Rott [23] proposes his counterexample to two of the Gärdenfors postulates in order to motivate a search for more “sophisticated” models of belief and belief change. It is incumbent upon any model seeking to capture more accurately such phenomena, such as the model proposed in this paper, to show how it deals with such problematic examples. Rott’s counterexample shall be taken as a test case: a model should make it clear to what extent the Gärdenfors postulates hold, and why they do not seem to hold in this counterexample.

In the following section, a model of belief revision shall be proposed. In Section 2.3, it shall be shown how this theory satisfies the desiderata for models of belief revision mentioned above: it satisfies the Gärdenfors postulates in appropriate cases, it models iterated revision and recovers iterated revision operations proposed in the literature as special cases, and it supports an enlightening, “sophisticated” analysis of Rott’s counterexample.

2.2 A model of belief revision

2.2.1 Belief states

In Section 1.1, a model of the local logical structure effective at a particular moment was proposed, in the form of what was called interpreted algebra. The locality of this language (not all sentences of some accepted overarching language are contained in an interpreted algebra), and of its logic (the logical consequence at play in the interpreted algebra does not necessary coincide with some global or absolute notion of logical consequence) respond well to certain limits in real agents’ belief states. The basic proposal for modelling the belief state of an individual is to employ traditional models of beliefs, but, instead of using some fixed language and logical structure, considering the beliefs of the agent at a particular moment as couched in a local logical structure which is effective at that moment. This is, so to speak, a model of the beliefs of which the agent is “aware”, in the agent’s own language, at a particular moment. Philosophical discussion of the consequences of this perspective on modelling beliefs shall be left for another paper; here, concern shall centre mainly on details of the model.
The simplest model of beliefs would be, following the tradition, as a set of sentences closed under logical consequence – that is, the logical consequence of the local logical structure which is operative at the appropriate moment.\(^{21}\) However, it has been suggested that correct representations of the belief states (sometimes called “epistemic states”) of an individual should include information not only about his current beliefs, but also about how he would revise them, or, alternatively, about how “entrenched” they are [5]. Such a model of belief states shall be employed. It consists in adding a Groove order – representing not only the agent’s beliefs but potential revisions of these beliefs – to an interpreted algebra – representing the local logical structure in play at the moment in question. The resulting structure is called an ordered algebra.

**Definition 8 (Ordered algebra).** An ordered algebra is a pair \((B, \preceq)\) where \(B = (B_f, B, q)\) is an interpreted algebra and \(\preceq\) is a reflexive order on the atoms of \(B\) which is connected, transitive and finitarily stoppered (conditions (S \(\preceq\) 1-3)).\(^{22}\)

Here are three examples of ordered algebra (see Example 1 for terminology and notation).

**Example 3.** For any trivial interpreted algebra \(B\), \((B, \preceq_0)\) is a trivial ordered algebra, where \(\preceq_0\) is the empty order.

The point ordered algebra for sentence \(\phi\), \((B_{\phi_p}, \preceq_{\phi_p})\), has, for interpreted algebra, the point algebra for \(\phi\), and, for order, the only possible one.

The simple ordered algebra for sentence \(\phi\), \((B_{\phi}, \preceq_{\phi})\), has, for interpreted algebra, the simple algebra for \(\phi\), and, for order, the order favouring \(\phi\): \(q(\phi) \preceq q(\neg \phi)\).

**Explication** Ordered algebras provide subtle models of the agent’s doxastic states, in that they permit, for any given sentence, a range of “doxastic statuses”. To clarify the discussion, a little preliminary terminology will prove useful.

**Definition 9.** For an element \(X\) of the base algebra of an interpreted algebra \(B = (B_f, B, q)\), let \(|X| = \{|\phi| \in B | q(\phi) \geq X\}\).\(^{23}\)

The centre of an ordered algebra \((B, \preceq)\), denoted \(|B|\times\{x\times\preceq\text{-minimal}\}\); that is, the set of elements of \(B\) true in all the small worlds minimal with respect to \(\preceq\). The centre of a trivial ordered algebra \((B_0, \preceq_0)\) shall be taken to be the set of its elements.

An element \(A \in |B|\) is a generator if \(q(A) = \{x| x \preceq\text{-minimal}\}\).\(^{24}\) In a trivial ordered algebra, every element is a generator.

An element \(A \in |B|\) is a local tautology if \(q(A) = \top\).\(^{25}\) All the elements of a trivial ordered algebra are local tautologies.

For a sentence \(\phi\) in an ordered algebra \((B, \preceq)\), there are two general senses in which it may be “believed”. It might in the centre of the algebra; furthermore, it may be a

---

\(^{21}\) This model implies that the agent is locally logically omniscient. However, this seems a generally correct assumption if an agent believes ‘\(A\)’, if \(A\), then ‘\(B\)’.迗eads these beliefs as such, and ‘\(B\)’ is in play, then it would seem that he believes ‘\(B\)’.

\(^{22}\) Given that, in general, the interpreted algebra involved here are finite, the condition (S \(\preceq\) 3) is redundant. It is retained for coherence.

\(^{23}\) Recall that \(\geq\) is the order of the Boolean algebra. In traditional notation (read for example in terms of small worlds), this definition would be expressed as \(|X| = \{|\phi| \times x \geq \phi\}\), which is, \(|X| = \{|\phi| \times x \geq \phi\}\), that is, \(|X| = \{|\phi| \times x \geq \phi\}\).\(^{24}\) Since q is surjective, there is always a generator. It is natural in the finite case appropriate here.

\(^{25}\) This notion can be defined on interpreted algebras, but shall only be used in the context of ordered algebras.
local tautology. The first case is what are called “beliefs” in Grove’s model [10]; this set of “beliefs” may be revised if new information forces one to move to worlds where not all of them hold. The second case corresponds what have been called “doxastic commitments” or “irrevocable beliefs” [25]; no revision of such beliefs is admissible, since there is no world (falling under the relation $\preceq$) where they do not hold.

However, because the ordering only applies to the local logical structure, these possibilities take on different meanings than in traditional frameworks. Invariably, a fixed language and notion of logical consequence is presupposed in the literature, so that “irrevocable” beliefs are just the tautologies of this language. However, in the framework proposed here, where the language and the logic are local, it is not necessary that the local tautologies are tautologies of some fixed language (see Section 1.1); in this sense, the believer is not modelled as omniscient. Moreover, as opposed to most traditional models, not only the centre of the ordered algebra modelling the agent’s belief state, but also the local tautologies, may change in time.

The local tautologies, rather than being some fixed set of logical truths in a strong sense, are more adequately understood as those opinions which the agent cannot envisage giving up at that particular moment. Perhaps the term ‘commitments’ is appropriate for them, perhaps ‘presuppositions’ is more fitting; both capture the idea of acceptance without question (though the second is more amenable to change). The sentences in the centre are the most preferred sentences amongst those which are in play; assuming the traditional terminology, they shall be called (explicit, instantaneous) ‘beliefs'. For an ordered algebra representing the agent’s belief state, the centre will be the set of beliefs, often denoted by $K$ (or $K_{B, \preceq}$ where necessary). A generator is a belief (such as the conjunction of elements of $K$) which sums up exactly the belief set.

Among those sentences which are neither beliefs nor presuppositions (nor presupposed false), there are two sorts. Firstly, there are those which are in play for the agent, that is, which are contained in the interpreted algebra $B$. These are true in some small worlds, but not all of the $\preceq$-minimal ones. The ordinary method for revising beliefs in Grove models apply to such sentences: the set of beliefs after revision by $\phi$ are those sentences true in all $\preceq$-minimal small worlds where $\phi$. The ordered algebra thus represents the agent’s opinion on how he would revise his beliefs by sentences which are in play for him at that moment. As shall be discussed shortly, this cannot be the actual revision operation, since such an operation would have to take account of revision by sentences not belonging to the local language. So the ordered algebra only provides envisaged revision (or, equivalently, envisagedrenchment of beliefs), but not a full measure of actual revisions. Note that, since revision using a Grove order satisfies the Gärdenfors postulates [9], these envisaged revisions satisfy Gärdenfors postulates relative to the local logical structure effective at that moment (i.e. $B$).

**Observation 2.** An ordered algebra $(B, \phi)$ satisfies the Gärdenfors postulates relative to the interpreted algebra $B$ (that is, for the sentences of this algebra, and its notion of logical consequence).

Finally, a sentence may simply not figure in the underlying interpreted algebra; it may simply be out of play. The agent has no opinions at this moment about such a sentence. However, for the purposes of revision, it is as necessary to model the agent’s reaction to such a sentences as it is to capture his response to a sentence of which he is already aware. The next question is thus that of the representation of the new information.
2.2.2 New information

Often new information with respect to which beliefs are to be revised is treated as a simple sentence of some fixed language. However, in the framework proposed here, where no use is made of such a fixed language, incoming information will generally require a local logical structure of its own, because it may involve sentences which do not figure in the local logical structure relevant for the current state of belief (Section 1.2). It was suggested that this structure be modelled as an interpreted algebra, and this suggestion shall be retained without repeating the arguments which weigh in its favour (flexibility, ability to capture the context in which a sentence is learnt, and so on). However, the interpreted algebra only models the logical structure in which the incoming information is couched; it does not specify which sentences in this structure are learnt, or the extent to which the sentences of the local language are to be accepted.

For example, if the local logical structure pertinent for a case where \( \phi \) is learnt is modelled as a simple algebra (Example 1), some supplementary structure on this algebra would be needed to represent the fact that it is \( \phi \) and not \( \neg \phi \) which is to be accepted. It would seem natural to represent this fact with an order on the states or small worlds of this algebra which favours (the small world where) \( \phi \) to (that where) \( \neg \phi \). So doing, one obtains an ordered algebra – in fact, one obtains a simple ordered algebra (Example 3). On the other hand, if one uses a point algebra as model for the local logical structure, \( \phi \) is automatically specified as input information, because it is a local tautology of this algebra. But in this case too, the structure modelling the incoming information is (trivially) an ordered algebra (the point ordered algebra, Example 3). In the most basic cases, new information can be represented as ordered algebra; the suggestion is that this sort of representation is appropriate in general.

Indeed, representing new information with ordered algebras inherits the advantages of ordered algebras which have been emphasised above. As with the case of belief states (Section 2.2.1), different statuses of the different elements of incoming information may be captured by ordered algebras. A sentence learnt irrevocably – accepted without any envisaged possibility of challenging the new information – corresponds to a local tautology of the ordered algebra representing the incoming information (cf. the point algebra example above).\(^2\) On the other hand, often the information acquired comes in a context which admits that it is reliable only under certain conditions: the information gleaned from a scientific experiment, for example, does not consist of the bare result, but a collection of conditions and assumptions relating to the details of the experiment, which, if found to be false, would undermine the result. Such a situation is best represented by an ordered algebra which contains not only the sentence expressing the result of the experiment, say \( \phi \), but sentences expressing the appropriate conditions. In such an algebra, \( \phi \) is not a local tautology, since conditions are envisaged in which the result of the experiment would be revoked. It is, however, the most preferred or natural option: it is in the centre of the ordered algebra (cf. the case of beliefs in Section 2.2.1). To be more pedantic, what is learnt from the experiment is characterised precisely by any sentence which is true only in the minimal (or most preferred) worlds of the ordered algebra – that is, by any generator of the ordered algebra (Definition 9). To capture revision properly, such a level of precision shall prove necessary: incoming information shall modelled by an ordered algebra, where the sentence learnt can be thought of as a generator of the algebra.

Finally, this representation permits – thanks to the locality of the framework – that

\(^2\) As shall be shown in Theorem 2, and as is already clear from Example 2 and the ensuing discussion, the notion of irrevocable revision proposed in [25] is thus recovered as a particular case.
not all sentences of some fixed global language figure in the ordered algebra representing the new information; such sentences are out of play in the context in which this information is acquired. As has been noted in Section 1.2, this aspect endows the model with the flexibility to capture accurately both the “standard” cases of inputs figuring single sentences (cf. the examples of point and simple algebras discussed above), and the more complicated cases featuring details on the conditions under which the information was acquired, possible revisions of the new information after subsequent discoveries, and so on. Citing the latter sorts of cases, the idea of representing new information by the same type of structure as that used to model the belief state has been proposed in the literature [19, 16]. However, it has always been done in the framework of a fixed global language, where the idea seems less palatable: for example, such structures determine revisions of the result of scientific experiment carried out in Tübingen in the light of weather reports from Tahiti. The locality of the current framework, and notably the fact that certain sentences may be out of play, allows the current proposal to avoid such counterintuitive consequences.\footnote{The best that can be done in a framework that insists on ordering worlds of a fixed global language is to represent the weather in Tahiti as being independent from the experimental result (for example, at each run of the order, there are worlds where the weather is good, and others where the weather is bad). However independence is not out of play: the former encompasses a certain affirmation concerning the relationship between the Tübingen experiment and the Tahiti sunshine, the latter does not. This delicate point, tangential to the present discussion, is discussed in [13, §4.2.4] and shall be further developed in a forthcoming paper.}

Technically, the representation of new information by a structure of the same sort as that which represents the belief states (an ordered algebra) implies that the representation of belief revision should consist of an operation which takes two ordered algebras to a third. Such a fusion operation on ordered algebras can be defined from simple operations on interpreted algebras and on orders.

### 2.2.3 Revision operation

Under the current proposal, both the belief state and the new information are represented by interpreted algebras with appropriate orders on them (ordered algebras); the revision operation will somehow combine these algebras. The operation which combines the interpreted algebras (local logical structures) has already been defined and motivated: it is the fusion operation $*$ of Section 1.2. The following definition and observation state that the orders on the initial interpreted algebras can be mapped canonically into the fusion interpreted algebra; furthermore, the images of the orders are still Grove orders.

**Definition 10.** For $\preceq_i$ an order on the atoms of $B_i$, for $i \in \{1, 2\}$, let the image of $\preceq_i$ in $B_1 \ast B_2$, also called $\preceq_{i'}$, be defined as follows:

$$x \otimes y \preceq_{i'} x' \otimes y' \quad \text{iff} \quad \begin{cases} x \preceq_1 x' & \text{if } i = 1 \\ y \preceq_2 y' & \text{if } i = 2 \end{cases}$$

**Observation 3.** For $i \in \{1, 2\}$, $j \in \{1, 2, 3\}$, if $\preceq_i$ satisfies $(S \preceq j)$, as an order on (the atoms of) $B_1$, then its image in $B_1 \ast B_2$ satisfies $(S \preceq j)$.

It remains to specify how these orders, coming from the respective algebras, are to be combined. There is a selection of operations which may be employed here, several of which have been discussed in some form or another in the literature. For the purposes of this paper, where the general framework is at issue, it would not be appropriate to enter into detailed considerations and debates; it will suffice to pick a natural
candidate and develop a revision operation built on this operation on orders. Although this candidate, and the revision operation constructed from it, has several interesting, attractive and useful properties, let it be emphasised that other operations on orders may prove equally useful, and may result, using a similar procedure to that carried out below, in equally interesting revision operations. The operation on orders used here is the lexicographic product.

**Definition 11 (Lexicographic product).** Given two orders \( \preceq_1 \) and \( \preceq_2 \) on a set \( S \), the lexicographic product \( \preceq_1 \times_L \preceq_2 \) is an order on \( S \) with, for all \( a, b, c, d \in S \),
\[
(a, b) \preceq_1 \times_L \preceq_2 (c, d) \text{ iff } \left\{ \begin{array}{l}
\text{either } b \prec_2 d \\
\text{or } b \succeq_2 d \text{ & } a \preceq_1 c
\end{array} \right.
\]

This product has the following two useful properties. Firstly, it is non commutative: indeed, it gives priority to one of the orders over the other. This fits well with the idea that new information should have priority over previous beliefs. Secondly, as an operation on Grove orders, it yields a Grove order, and thus, when used to combine ordered algebras, it guarantees that the resultant structure will be an ordered algebra.\(^{28}\)

**Observation 4.** For \( j \in \{1, 2, 3\} \), if \( \preceq_1 \) and \( \preceq_2 \) satisfy \((S \leq j)\), then \( \preceq_1 \times_L \preceq_2 \) satisfies \((S \leq j)\).

Therefore the fusion operation on ordered algebras, consisting of the fusion of interpreted algebras (Definition 7) and the lexicographic product of the images of the initial orders in this fusion, is well-defined.

**Definition 12 (Fusion \(*\) of ordered algebras).** Let \((B_1, \preceq_1)\) and \((B_2, \preceq_2)\) be ordered algebras, and \(\simeq\) a presupposed identification relation between them. The fusion
\[
(B_1, \preceq_1) *_{\simeq} (B_2, \preceq_2) = (B_1 *_{\simeq} B_2, \preceq_1 \times_L \preceq_2)
\]

The subscript \(\simeq\) shall be omitted when clear from the context.

\((B_1, \preceq_1)\) represents the initial belief state: its centre is the set of beliefs (Section 2.2.1). \((B_2, \preceq_2)\) represents the new information: the sentence learnt is a generator (Section 2.2.2). \((B_1, \preceq_1) *_{\simeq} (B_2, \preceq_2)\) represents the resulting belief state: its centre is the new set of beliefs. Note that, by the definition of lexicographic order given above (Definition 11), priority is accorded to new information over previous beliefs.

### 2.3 Properties of the model

The operator \(*\), with the interpretation of ordered algebras as representations of belief states and incoming information, provides a model of belief revision in so far as it satisfies an appropriate translation of the well-known Gärdenfors postulates for belief revision into the proposed framework. Since it satisfies the property of **categorial matching** [11] – the representation of the belief state after revision (ordered algebra) is of the same format as the representation before revision (and thus appropriate for further revision) – it is automatically an **iterated** revision operator; furthermore, two important iterated revision operators proposed in the literature [25, 19, 16] can be recovered in the proposed framework as **special cases** corresponding to particular constraints placed on the incoming information. Finally, the proposed framework supports an analysis of Rott’s counterexample to two of the Gärdenfors postulates [23]. These points shall be dealt with successively.

\(^{28}\) Other interesting properties of this order include associativity (see Remark 4 in the proof of Theorem 2), and the fact that it commutes with quotients on interpreted algebras.
2.3.1 Gärdenfors postulates

In order to state the theorem detailing the postulates satisfied by $\ast$, the following preliminary definition is required.

Definition 13. Let $(B, \preceq)$ be an ordered algebra with generator $\phi$. For $\psi \in \mathcal{B}$, the refinement of $(B, \preceq)$ with generator $\phi \land \psi$ is the ordered algebra $(B, \preceq')$, with

$$x \preceq' y \iff \begin{cases} x \leq q(\psi), y \not\leq q(\psi) & \text{if } x, y \leq q(\phi) \\ x \prec y & \text{otherwise} \end{cases}$$

Where $\prec$ is the strict order obtained from $\preceq$ in the ordinary way.\footnote{The ordering $\prec$ is defined as the reflexive and transitive closure of $\preceq$.} This ordered algebra has generator $\phi \land \psi$.

Using the model of belief states and new information proposed in Section 2.2, with the interpretation of the set of beliefs and sentences learnt as the centre and the generator of the respective ordered algebras, the operator $\ast$ satisfies the Gärdenfors postulates in the following sense.

Theorem 1. Let $(B_1, \preceq_1)$ be an non trivial ordered algebra with centre $K$, let $(B_2, \preceq_2)$ contain sentences $\phi$ and $\psi$ and have generator $\phi$, and let $(B_2, \preceq_3)$ be its refinement with generator $\phi \land \psi$. Let $K \ast \phi$ (resp. $K \ast (\phi \land \psi)$) be the centre of $(B_1, \preceq_1) \ast (B_2, \preceq_2)$ (resp. $(B_1, \preceq_1) \ast (B_2, \preceq_3)$). Then\footnote{The postulates here have intentionally been expressed in a form analogous to those which are common in the literature; possible strengthenings of certain postulates have been ignored.}

\begin{enumerate}
  \item[(K * 1)] $K \ast \phi = Cn_{12}(K \ast \phi)$;
  \item[(K * 2)] If either, for each $\chi \not\vDash_2 \phi$, $\top_1 \not\vDash \neg \chi$, or $\top_2 \not\vDash \phi$, then $\phi \in K \ast \phi$;
  \item[(K * 3)] $K \ast \phi \subseteq Cn_{12}(K \cup \{\phi\})$;
  \item[(K * 4)] If, for each $\chi \not\vDash_2 \phi$, $\neg \chi \in K$, then $Cn_{12}(K \cup \{\phi\}) \subseteq K \ast \phi$;
  \item[(K * 5)] If, for each $\chi \not\vDash_2 \phi$, $\top_1 \not\vDash \neg \chi$, then $K \ast \phi$ consistent under $Cn_{12}$;
  \item[(K * 6)] If $\phi \not\vDash_2 \chi$, then $K \ast \phi = K \ast \chi$;
  \item[(K * 7)] $K \ast (\phi \land \psi) \subseteq Cn_{12}(K \ast \phi \cup \{\psi\})$;
  \item[(K * 8)] If, for each $\chi \not\vDash_2 \psi$, $\neg \chi \not\in K \ast \phi$, then $Cn_{12}(K \ast \phi \cup \{\psi\}) \subseteq K \ast (\phi \land \psi)$.
\end{enumerate}

where $\vDash_1$, $\vDash_2$, $\vDash_{12}$, $Cn_{12}$ and so on, are the consequence relations and sets of consequences, in $B_1$, $B_2$, and $B_1 \ast B_2$ respectively.

The Gärdenfors postulates are normally expressed in terms of sentences and sets of sentences (see, for example [10, §3.3]), whereas the basic notion here is that of ordered algebra. As discussed above, ordered algebras offer a more flexible and general representation of beliefs and new information; accordingly, the subtlety in Theorem 1 is required to curb this flexibility and recover the simpler case dealt with traditionally.

The most important aspect is the translation of the traditional notions of sets of beliefs and sentences learnt as centres and generators of the ordered algebras representing the belief state and the new information $(B_1, \preceq_1)$ and $(B_2, \preceq_2)$. As discussed...
in Section 2.2, the interpretation of “belief” and “sentence learnt” as “most preferred sentence” is only one of the interpretations which could be afforded to these terms by a model consisting of ordered algebra, albeit a particularly natural and intuitive one. It is thus important to emphasise that Theorem 1 only states that the Gärdenfors postulates are (more or less) satisfied when applied to the centres and generators of ordered algebras. In particular, the postulates do not necessarily apply to the local tautologies of the respective algebra, that is, to sentences which have been called “commitments”, “presuppositions”, or “irrevocable sentences” in Section 2.2. This is one concrete sense in which the model of belief revision proposed here recovers the traditional theory as an idealisation: the Gärdenfors postulates hold, but only in the special cases when centres and generators (and the corresponding notions of “belief” or “information”) are being used. Thus, this theorem shows that one of the desiderata of a more realistic model of belief dynamics — namely, to exhibit in which sense previous theories are idealisations — is fulfilled.

Several divergencies from the standard formulation of the Gärdenfors postulates [10, §3.3] arising from the increased generality of the proposed framework are worth recording.

Remark 2. (i) Since all the consequence relations involved here are local — relative to particular interpreted algebras — each of the Gärdenfors postulates, in their original form, have to be modified to specify which consequence relation is involved. This factor requires particular care when specifying the ordered algebra used to represent the new information \( \phi \land \psi \) in axioms \((K * 7 - 8)\) — thus the appeal to the notion of refinement.

(ii) Given that, in the representation of new information by ordered algebras, there may be several sentences in play, the condition stating the \( \phi \) is non contradictory (in the ordered algebra representing the belief state) concerns not only whether or not \( \phi \) itself is contradictory, but whether any sentence equivalent to it (in the model of the new information) is contradictory. \((K * 2), (K * 4), (K * 5)\) and \((K * 8)\) contain clauses to this effect. Furthermore, since the “sentence learnt” \( \phi \) may not be a local tautology of the ordered algebra representing the new information (that is, \( \neg \phi \) is envisaged), \( \ast \) allows cases of revision by \( \phi \), where \( \neg \phi \) is a local tautology of the belief state, with a non trivial result. This is exactly the case ruled out by the condition in \((K * 2)\).

(iii) As a supplementary demonstration that the complications discussed in point (ii) arise from the richer possibilities afforded by the liberalised representation of the new information, note that they dissolve when these liberalisations are removed. More specifically, when a point algebra is used to represent the new information (Example 3), the conditions discussed above are automatically satisfied, and the postulates stated in Theorem 1 reduce to the standard Gärdenfors postulates.

(iv) Recall (Observation 2) that the order on an ordered algebra defines a revision operation, with respect to sentences belonging to the algebra, which satisfies the Gärdenfors postulates. As one would expect, in the case of revision by a sentence already belonging to the ordered algebra representing the belief state, this revision operation coincides with the more general revision operation \( \ast \); this result is stated as Proposition 1 in the Appendix. The precise statement requires a condition relating the interpreting algebras modelling the local logical structures of the

\[31\] The original form of \((K * 2)\) is \( \neg \phi \in K * \phi \).
belief state and the new information (demanding, approximately, that they have compatible consequence relations), which is automatically satisfied when the new information is represented by a point algebra. This result thus not only exhibits to what extent the envisaged revisions agree with actual revisions, but shows once again that, when the complexities afforded by the representation of new information are absent, the two notions of revision coincide, and indeed reduce to the ordinary notion known in the literature.

2.3.2 Iterated Revision

As noted above, * is an iterated revision operator, in that it yields a structure (ordered algebra) fit for subsequent revision (using *). As in the discussion of the Gärdenfors postulates above, the realistic credentials of this model can be further bolstered by showing in what sense several notions of iterated revision proposed in the literature can be recovered as special cases of revision with * − indeed, as special cases where the new information comes in a particular format. The two iterated revision operators considered, called “radical” and “moderate” revision in [22], and of which versions have been suggested and defended in [25] and [19, 16] respectively, are characterised by the following postulates.

Definition 14. (Rad) \((K * \phi) * \psi = K * (\phi \land \psi)\)

(Mod)

\[
(K * \phi) * \psi = \begin{cases} 
K * (\phi \land \psi) & \text{if } \psi \text{ is consistent with } \phi \\
K * \psi & \text{otherwise}
\end{cases}
\]

By requiring that the sentence with respect to which revision is made be represented as coming in a particular form − respectively as a point or a simple algebra (Example 3) − one or other of the postulates are automatically satisfied.

Theorem 2. Let K be the centre of the ordered algebra \((B, \preceq)\), and let consistency in the above definition be understood as if not being the case that \(\phi \simeq \neg \psi\), where \(\simeq\) is the identification relation between ordered algebras representing the information \(\phi\) and \(\psi\) respectively. Then

(Rad) If \(\phi\) and \(\psi\) are modelled by \(B_{\phi} \preceq B_{\psi}\) respectively, then (Rad) is satisfied.

(Mod) If \(\phi\) and \(\psi\) are modelled by \(B_{\phi} \preceq B_{\psi}\) respectively, then (Mod) is satisfied.

Remark 3. As per usual in the more general localist framework adopted here, there are several possible interpretations which can be given to the notion of “consistency”: one might ask for consistency in so far as the sentences appear as elements of the algebra representing the initial belief state \(K\) (a condition which assumes they do appear in this algebra), or as a relation between the sentences as they appear in the algebras representing the new information. The second option is adopted in the statement of the theorem; however, as can be noted from the proof, a similar result holds if the first notion of consistency is used (see Remark 5 in the Appendix).

This theorem counts as a further illustration of the fruitfulness of this model of belief revision, and more generally of the framework of which it is an instance (the framework proposed in Section 1). Iterated revision operations proposed in the literature are apparently recovered as special cases of the form of the input information. In the sense in which they suppose that the input information takes a particular form,
they are *idealisations*; in the sense in which the model proposed here does not make
this supposition, and indeed can accommodate a multiplicity of possible formats for
the incoming information, it is more *realistic*. Although further discussion will take
us too far from the motivational purpose to which this example has been put here, it
is worth noting that the idea that the plurality of belief revision operators might boil
down to a single general operation with a plurality of formats for incoming information
is not only philosophically (and technically) interesting, but may find support, if
not a precursor, from unexpected quarters. Friedman and Halpern [7] argue for a plu-
rality of belief revision operators, each appropriate to a different "ontology"; however,
given that their "ontologies" are differentiated largely by properties of the incoming
information (whether it is absolutely certain, or only has a certain plausibility relative
to other beliefs), one is justified in hoping that the plethora of revision operators may
reduce to a single, general, revision operator, and a range of possible formats for the
new information.

2.3.3 Rott's example

The final sort of challenge for an alleged realistic model of belief revision is to account
for the apparent counterexamples to properties or postulates proposed by existing theo-
ries. Given that, as has been underlined above, the existing theories can be understood
to be idealised precisely in the sense that they only apply in special cases, a more realis-
tic model should exhibit to what extent the situations involved in the counterexamples
do not fall within the remit of the theories. A counterexample to a postulate is not
considered to be a rejection of the postulate, but rather a symptom of its limited range
of validity. This comes out clearly in the analyses of such counterexamples which are
suggested by the model proposed above.

**The counterexample** The counterexample which shall be taken as a test case here
was proposed by Rott [23]. It is a counterexample to the Gärdenfors postulates (K
$\approx 7$) and (K $\approx 8$), which consists of a story concerning an agent who considers the
candidates for a metaphysics post in a philosophy department. Four candidates appear
in the story: $d$, who is by far the best, $a$, who excels in metaphysics (but has no pedigree
in logic), $b$, who is good at metaphysics and logic, and $c$, who is outstanding in logic
but mediocre in metaphysics. The agent is thus originally of the opinion that $d$ will get
the post. However, $d$ cannot take the post, and so the agent's beliefs will need to be
revised. Rott considers two alternative revision scenarios.

I. The agent is told (by a reliable source) that $a$ or $b$ will get the post; he according
revises his belief, and takes the opinion that $a$ will get it.

II. The agent is told (by a reliable source) that $a$ or $b$ or $c$ will get the post. This "trig-
ggers off a rather subtle line of reasoning" [23, p230]. Given $c$'s pedigree in logic,
and lack of pedigree in metaphysics, the agent figures that not only competence
in metaphysics, but also competence in logic, are criteria for the post. Therefore,
he reasons, $b$'s competence in both domains, and $a$'s restriction to metaphysics
alone, give $b$ the advantage: he comes to believe that $b$ will get the post.

---

32 It is in fact a counterexample to a slightly weaker postulate than (K $\approx 8$), but that shall not matter for
present purposes.
This contradicts Gärdenfors’ \((K * 7)\) in the following way.\(^{33}\) Without risk of confusion, let \(a, b, c, d\) be the propositions expressing that \(a, b, c\) and \(d\) respectively get the post. Let \(K\) be the agent’s prior set of beliefs. The described revision patterns yield:

\[
K * ((a \lor b \lor c) \land (a \lor b)) = K * (a \lor b) = Cn(a) \tag{Stn I.}
\]

\[
Cn((K * (a \lor b \lor c)) \cup \{a \lor b\}) = Cn(b) \tag{Stn II.}
\]

Whence \(K * ((a \lor b \lor c) \land (a \lor b)) \not\in Cn((K * (a \lor b \lor c)) \cup \{a \lor b\})\), \textit{contra} \((K * 7)\), which states that this inequality holds.

\textbf{Analysis}  
As the setup is described, there are basically four sentences in play—\(a, b, c,\) \(d\)—which cover all possibilities and are mutually exclusive (one and only one is true). The (minimal) interpreted algebra representing the local logical structure will contain (essentially) these four sentences, and four small worlds, such that each sentence shall be true in one and only one world. Without risk of confusion, the four worlds shall also be called \(a, b, c,\) and \(d\), according to which of the sentences is true in that world. This interpreted algebra, call it \(B\), is the local logical structure in which the agent’s initial belief state is couched. As the initial belief state of the agent is described (he thinks only metaphysics is important), it should be represented by the order \(\preceq\) on \(B\), where \(d \preceq a \preceq b \preceq c\).

Note that, if the information learnt in situation I. \((a \lor b)\) is represented as a point algebra, the revision operation \(*\) yields the desired result: \(a\) becomes the new belief (minimal state). If, on the other hand, a point algebra representation of the information is used in situation II., revision with \(*\) yields \(a\) and not the result \(b\) given in the story. In the proposed framework, the counterexample arises because the point algebra is not the appropriate representation of the incoming information, or to put it another way, the information with respect to which the agent revises is not \textit{just} the bare sentence \(a \lor b \lor c\). In fact, a different representation is more appropriate, and under revision with this representation, the Gärdenfors postulates do not necessarily hold.

To see that the point algebra is not an accurate of the incoming information, compare situation II. with situation III.\(^{34}\)

III. The agent is told (by a reliable source) quite simply that \(d\) will not take the job. No “subtle line of reasoning” is triggered off by this information: the agent alters his opinion and plumps for the next best candidate. He comes to believe that \(a\) will get the job.

Situations II. and III. yield different revisions of belief, despite the fact that the information acquired in the two cases are equivalent in the context of the example.\(^{35}\) This state of affairs is reminiscent of a well-known phenomenon in choice theory, the \textit{framing effect} [27], where equivalent choice problems\(^{36}\) extract different choices from agents. Here, equivalent formulations of the same input information yield different revisions of belief. And, just as Rott’s example apparently contradicts postulates of belief revision, the framing effect apparently invalidates axioms of rational choice.

\(^{33}\) The case of \((K * 8)\) is similar and shall not be repeated here.

\(^{34}\) In the conclusion of his paper, Rott alludes to a comparison with a case similar to the one described below. A detailed discussion of the differences, and similarities, between Rott’s remarks and the analysis proposed here lies beyond the limits of this paper.

\(^{35}\) More rigorously, \(a \lor b \lor c\) and \(\neg d\) are equivalent sentences of the interpreted algebra representing the agent’s belief state.

\(^{36}\) Or, if you prefer: different formulations of the same choice problem.
Given that the problem in the framing effect is that the presentation of the choice problem plays a decisive role, a natural approach is to explicitly model the effect of the presentation on the choice problem which the agent sees himself as faced with. Due to differences in presentation, two choice problems which are logically equivalent will result, after processing of the presentation, in logically non-equivalent choice problems which the agent sees himself as solving. This is the sort of tactic adopted in Kahneman & Tversky’s prospect theory [15, 27], where the choice process involves a “phase of framing and editing” which proceeds evaluation and indeed yields the choice problem which the agent effectively evaluates. According to this sort of theory, the basic evaluation of a choice problem is always the same — and indeed it satisfies all the traditional rationality postulates — but differences in the presentation of a choice problem may yield logically non-equivalent results of the “phase of framing and editing”, and thus different choices by the agent.

Given the considerations above, which extend beyond well-known similarities between choice theory and belief revision, Rott’s example seems ripe for this sort of treatment.\footnote{The tight relationship between belief revision and choice theory has already been explored in the literature, notably in [21], indeed Rott’s counterexample is in a certain sense the fruit of such work. However, to the knowledge of this author, little work has been done on the relationships between the “realism” of the postulates of choice theory and those of belief revision. This section can be seen as a first step in this direction; further development must be left for a later work.} Comparison of the simple story in situation III. and the intricate story in situation II. seem to suggest that, while in situation III. the incoming information (¬d) is directly employed to revise belief’s, in situation II., the information seems subject to a preliminary treatment before being used for revision. How else should one understand the “subtle line of reasoning” involved in situation II., if not as a phase of pre-processing of the incoming information?

Understood in this way, the information by which the agent effectively revises in situation II. is not the simple sentence a ∨ b ∨ c; in other words, the new information post-processing is thus not properly represented by the point algebra B_{a ∨ b ∨ c}. Rather, by his “subtle line of reasoning”, the agent establishes a preference among the options a, b, and c. A better representation of the information with respect to which he revises is an ordered algebra whose interpreted algebra, B’, contains the sentences a, b, c, and three worlds (also called a, b and c), each validating one and only one of the sentences and whose order, ≤’, is such that b ≤’ a ≤’ c.\footnote{This ordered algebra typically represents the result of his pre-processing/subtle line of reasoning. Other options, containing d, or according a different preference between a and c, are equally appropriate, and yield similar results.} The agent accepts as certain that a ∨ b ∨ c, but, on reflection, accords a preference to b over a and c. On the other hand, in situation III., which involves less pre-processing, the sentence ¬d is directly and simply used to revise: the incoming information can thus be represented by the point ordered algebra (B_{¬d} B_{¬d}).

In the context of the story, (B_{¬d} B_{¬d}) and (B’, ≤’) have equivalent sets of local tautologies. In particular, both have ¬d, or equivalently a ∨ b ∨ c, as local tautologies: this information is irrevocably learnt in both cases. This is the precise sense in which the two situations involve revision by equivalent information: the local tautologies are equivalent. However, as was underlined in Section 2.3.1, the Gärdenfors postulates do not apply to local tautologies but only to centres and generators of the ordered algebra (Theorem 1). So it is not expected that revision by these two algebras yield the same result. Indeed, it is straightforward to show that revision by these two algebras yield the results described in the story: a for (B_{¬d} B_{¬d}) (Situation III.); b for (B’, ≤’) (Situation II.). Furthermore, no Gärdenfors postulates are violated.

©Brian Hill
Towards a “sophisticated” model of belief dynamics
in this revision.

This analysis calls for further discussion. There are things to be said about the relationship between the notions of local tautology and generator, and the certainty and irrevocability of the sentence learnt. Similarly, work is required on the pre-processing phase, in which the appropriate representation of the incoming information (ordered algebra) is formed. The analysis invites an extension of the framework to more complex forms of the incoming information, and perhaps a larger range of operators on ordered algebra. Such discussion must be left for another time. Without it, a full theory of belief revision has not been completed. It is however important to understand just what has been already achieved by the model proposed here. It recovers traditional postulates of belief revision and iterated belief revision and reveals in what sense they are idealisations (they apply in special conditions). It provides an analysis of apparent counterexamples to these postulates, in the sense that it exhibits in what sense the cases do not satisfy the conditions required for the postulates to apply. In other words, where traditional theories and models of belief revision cannot cope with these examples, the proposed framework provides an understanding of them. This is an indication of the strength of framework in so far as it is used to represent belief revision — it captures phenomena that simpler frameworks miss. Moreover, it is an indication of the framework’s strength as a conceptual tool for studying belief revision: the analysis of the counterexample suggests an approach towards a full theory of belief revision, according to which the question is factorised into two parts — the pre-processing part and the revision proper — of which the latter is captured by the operation $*$ and the former may benefit from expression in terms of the framework proposed here.

One can conclude that the model of belief revision proposed in the second part of this paper is “realistic” and “sophisticated”, and indeed seems to open up fruitful possibilities of development into a full theory of belief revision. The general framework on which this model rests, and which was presented in the first part of this paper, has proved promising in the case of belief revision; it would not be exaggerated to expect similar success when applied to other questions where belief is involved and “realism” is an issue.

Proofs

Proof of Observation 1. Interpreted algebra By properties of quotients on Boolean algebras, and the fact that $\simeq$ is a congruence relation, $B_1 / \simeq$ is a Boolean algebra. Furthermore, since $\simeq$ is generated by an equivalence relation $\sim$ on $I, B_1 / \simeq$ is isomorphic to $B_1 / \sim$, so it is a free Boolean algebra.

Base algebra $\simeq^q$ is a congruence relation, therefore $B / \simeq^q$ is a Boolean algebra.

Surjective homomorphism By the definition of $\simeq^q$, if $[\phi]_\simeq = [\psi]_\simeq$, then $[q(\phi)]_{\simeq^q} = [q(\psi)]_{\simeq^q}$, so $q_\simeq$ is well-defined. It is straightforward to show that it is a Boolean

---

39 For example, a similar analysis can be carried out when the situation III. is represented by a simple algebra instead of a point algebra, if one employs a wider range of operations on, and relations between, ordered algebra. See [13.Ch5, Appendix A] for examples of such operations and relations.

40 In the concluding discussion of his counterexample, Rott suggests that a formalisation is needed which is sufficiently rich to take account of examples such as the one he proposed, whilst remaining operational (or “processable”) and eschewing gratuitous addition of information. The analysis proffered of Rott’s counterexample, the simple definition of the revision operation $*$, and the solid motivations and interpretations given in Sections 1 and 2.2 seem to suggest that the model proposed here satisfies all of these requirements.
homomorphism. Finally, for any \([x]_{\simeq} \in \mathcal{B} / \simeq\), there is a \(x \in [x]_{\simeq}\) and \(\phi \in B_1\) such that \(q(\phi) = x\), by the surjectivity of \(q\). \(q_{\simeq}(\phi) = [x]_{\simeq}\). So \(q_{\simeq}\) is surjective.

\((B_1 / \simeq, B / \simeq, q_{\simeq})\) is an interpreted algebra; furthermore, by uniqueness of quotients on Boolean algebra (and of the appropriate closures of relations), it is the only algebra resulting from these operations. The quotient is thus well-defined.

\[\square\]

Proof of Theorem 1. Let \(\preceq_{12}\) stand for \(\preceq_1 \times_L \preceq_2\) and similarly for \(\preceq_{13}\). Without risk of confusion, the image of the element \(\phi \in [B_1]\) in \([B_1 * B_2]\) shall also be called \(\phi\).

(K * 1) For any element \(X\) in the base algebra of an interpreted algebra \(B\), \(|X|\) is closed under the consequence relation on \(B\), by Definitions 1 and 9. \(K * \phi\) is such a \(|X|\) in \(B_1 * B_2\), so it is closed under the consequence relation for \(B_1 * B_2\).

(K * 2) There are two cases:

For all \(\chi \leftrightarrow_2 \phi\), \(T_1 \not\models_1 \neg \chi\). If \(B_2\) is trivial, the condition is trivially satisfied. If not, it follows from the hypothesis that \(\phi \not\models_{12} \bot_{12}\). By definition of the lexicographic order, \(\{x \mid x \preceq_{12} \text{min}\} \subseteq q(\phi)\), so \(\phi\) is in \(K * \phi\).

There is \(\chi \equiv_2 \phi\) with \(T_1 \models_1 \neg \chi\), and \(T_2 \models_2 \phi\). \(B_1 * B_2\) is trivial, and so, by Definition 9, the centre \(K * \phi\) contains \(\phi\).

(K * 3) Let \(\theta\) be a generator of \((B_1, \preceq_1)\), and consider the images of \(\theta\) and \(\phi\) in \(B_1 * B_2\). \(C_{n12}(K \cup \{\phi\}) = \{\psi \in B_1 * B_2 \mid \theta \land \phi \models \psi\} = q(\theta) \cap q(\phi)_x\). There are two cases:

\[\theta \land \phi \not\models_{12} \bot_{12}\]. Since \(q(\theta)\) (resp. \(q(\phi)\)) is the (non-empty) set of \(\preceq_1\)-minimal (resp. \(\preceq_2\)-minimal) worlds in \(B_1 * B_2\), \(q(\theta) \cap q(\phi)\) is the set of \(\preceq_{12}\)-minimal worlds in \(B_1 * B_2\). So \(K * \phi = |q(\theta) \cap q(\phi)| = C_{n12}(K \cup \{\phi\})\); \(\theta \land \phi \not\models_{12} \bot_{12}\). Therefore \(C_{n12}(K \cup \{\phi\}) = |B_1 * B_2| \subseteq K * \phi\).

(K * 4) Immediate consequence of the reasoning in the first case of axiom (K * 3).

(K * 5) Since \(B_1\) is non trivial, it follows from the hypothesis that \(B_1 * B_2\) is non trivial. So there is a non empty set of \(\preceq_{12}\)-minimal worlds; since \(q_{12}(\bot) = \bot\) the set of sentences true in these worlds, \(K * \phi\) is consistent under \(C_{n12}\).

(K * 6) By Definition 9, if \(\phi\) is a generator of \((B_2, \preceq_2)\), and \(\phi \equiv_2 \chi\), then \(\chi\) is a generator of \((B_2, \preceq_2)\). So both \(K * \phi\) and \(K * \chi\) are the centre of \((B_1 * B_2, \preceq_{12})\), and hence they are equal (by Definition 9, there is a unique centre).

(K * 7) There are two cases:

There is \(\chi \equiv_2 \phi\) with \(T_1 \models_1 \neg \chi\). If \(B_1 * B_2\) is trivial, then the condition trivially holds. If not, let \(\theta\) be a generator of \((B_1 * B_2, \preceq_{12})\). Since \(\phi\) is not a consequence of \(\theta\), and since \(\preceq_2\) and \(\preceq_3\) coincide on \(\neg \phi\) (Definition 13), \(\theta\) is also a generator of \((B_1 * B_2, \preceq_{13})\). Thus \(K * (\phi \land \psi) = |q(\theta) \cap q(\phi)| \subseteq |q(\theta) \cap q(\psi)| = C_{n12}(K * \phi \cup \{\psi\})\).

For all \(\chi \equiv_2 \phi\), \(T_1 \not\models_1 \neg \chi\). Let \(\theta\) be a generator of \((B_1 * B_2, \preceq_{12})\). There are two cases:
\[ \theta \land \psi \equiv_{12} \perp_{12}, \quad q(\theta) \cap q(\psi) \text{ is thus the (non empty) set of } \preceq_{12} \text{-minimal worlds in } B_1 \ast B_2 \text{ where } \psi \text{ holds. However, by Definition 13, this is exactly the set of } \preceq_{12} \text{-minimal worlds. Therefore, } K \ast (\phi \land \psi) = |q(\theta) \cap q(\psi)| = Cn_{12}(K \ast \phi \cup \{ \psi \}); \]

\[ \theta \land \psi \equiv_{12} \perp_{12}. \text{ Therefore } Cn_{12}(K \ast \phi \cup \{ \psi \}) = |B_1 \ast B_2| \geq K \ast (\phi \land \psi). \]

(K \ast 8) Immediate consequence of the reasoning in the first case of the second case in axiom (K \ast 7).

\[ \Box \]

**Proposition 1.** Let \((B_1, \preceq_1)\) be an ordered algebra with centre \(K\), \((B_2, \preceq_2)\) an ordered algebra with generator \(\phi\), such that the identification relation \(\preceq \) between \(B_1\) and \(B_2\) is total on \(|B_2|\) (to every element of \(B_2\), it associates an element of \(B_1\)), and the consequence relation is preserved under this relation.\(^{41}\) Let \(K \ast \phi = K \ast \psi\) be the centre of \((B_1, \preceq_1) \ast (B_2, \preceq_2)\) (the result of revising according to \(\ast\)), and \(K \ast e \phi = \{ x \in B_1 \mid \exists \preceq_1 \text{-minimal } x, \exists q_1(\phi) \}\) (the result of revising by the revision operation induced by \(\preceq_1\) on \(B_1\)). Then, \(K \ast \phi = K \ast e \phi\).

**Proof of Proposition 1.** By the condition on \(\preceq\), \(B_1 \ast B_2 = B_1\). By the definition lexicographic product, the \(\preceq_{12}\)-minimal worlds are exactly the \(\preceq_{1}\)-minimal worlds where \(\phi\). These are just those worlds relevant for \(K \ast e \phi\).

\[ \Box \]

**Proof of Theorem 2.** Since all the operations used to define \(\ast\), the fusion operator on ordered algebras, are associative, \(\ast\) is associative. Let \((B, \preceq), (B_2, \preceq_2), (B_3, \preceq_3)\) be ordered algebras with centre \(K\) and generators \(\phi\) and \(\psi\) respectively. Let \(\phi \ast \psi\) be a generator of \((B_2, \preceq_2) \ast (B_3, \preceq_3)\), and \(K \ast \ast\) denote the centre of the appropriate fusion of algebras. The associativity of \(\ast\) implies that\(^{42}\)

\[ (\text{Ass}) \quad (K \ast \phi) \ast \psi = K \ast (\phi \ast \psi) \]

**Remark 4.** It follows that all iterated revision operators which are supported by this \(\ast\) (as defined with the lexicographic product) satisfy a formula of the form \((K \ast A) \ast B = K \ast F(A, B)\), for some function taking pairs of sentences to sentences. This property is called "right-associativity" in \([22, \S 8.3]\).

Both the radical and moderate iterated revision operators are characterized by conditions of this form. It remains to show that, when \((B_2, \preceq_2)\) (resp. \((B_3, \preceq_3)\)) are point (resp. simple) algebras for \(\phi\) and \(\psi\), \(\text{(Rad)}\) (resp. \(\text{(Mod)}\)) is satisfied.

**Rad** By the definition of point algebras (Example 3):

\[
(B_{\phi p}, \preceq_{\phi p}) \ast (B_{\psi p}, \preceq_{\psi p}) = \left\{
\begin{array}{ll}
(B_{\phi p}, \preceq_{\phi p}) & \text{if } \psi \simeq \phi \\
(B_{\phi p}, \psi, q_0, \preceq_0) & \text{if } \psi \simeq \neg \phi \\
(B_{\phi p, \psi}, 1, q, \preceq) & \text{otherwise}
\end{array}
\right.
\]

where \(q : \phi \land \psi \mapsto \top\) and \(\preceq\) is the only possible order. In all cases, \(\ast\) yields an ordered algebra with generator \(\phi \land \psi\); using this ordered algebra to revise by \(\phi \land \psi\), \((\text{Ass})\) ensures that \(\text{(Rad)}\) is satisfied.

---

\(^{41}\) For \(\phi \simeq \phi', \psi \simeq \psi', \phi \Rightarrow \psi \Leftrightarrow \phi' \Rightarrow \psi'\)

\(^{42}\) As noted in Remark 1, the definition of interpretation relation (Definition 2) extends naturally to the case of more than two algebras.
(Mod) By the definition of simple algebras (Example 3):

\[(B_\phi, \preceq_\phi) \ast (B_\psi, \preceq_\psi) = \begin{cases} (B_\phi, \preceq_\phi) & \text{if } \psi \simeq \phi \\ (B_\phi, \preceq_{\neg \phi}) & \text{if } \psi \simeq \neg \phi \\ ((B_{(\phi, \psi)}, 4, q), \preceq) & \text{otherwise} \end{cases}\]

where 4 has four atoms, the images, under \(q\), of \(\phi \land \psi\), \(\neg \phi \land \psi\), \(\phi \land \neg \psi\) and \(\neg \phi \land \neg \psi\), and the order imposed by \(\preceq\) is that in which they are listed. \(\ast\)  yields an ordered algebra with generator \(\phi \land \psi\) when \(\phi\) and \(\psi\) are not contradictories, and \(\psi\) when they are. Using this ordered algebra to revise by \(\phi \land \psi\), (Ass) ensures that (Mod) is satisfied.

\[\square\]

Remark 5. Examination of the proof shows that, if \(\psi\) is not a contradictory element of \(B\), then (Rad) and (Mod) are satisfied when the consistency in Definition 14 is understood relative to \(B\).\(^{43}\) This is another sense in which the two notions of iterated revision are captured by this representation of the new information.

References


\(^{43}\)The restriction on \(\psi\) is required given the flexibility of the representation of new information. It is not required in the case of (Rad), since point algebras are being used (see (iii) of Remark 2).


Brian Hill
GREGHEC / IHIST
17 rue Racine
75006 Paris
hill.brian@wanadoo.fr