Brian Hill

Towards a "Sophisticated" Model of Belief Dynamics. Part I: The General Framework

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 a 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 some of 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. The companion paper (*Towards a "sophisticated" model of belief dynamics. Part II*) contains an application of this framework to the particular case of belief revision.

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 beliefs proposed by logicians, philosophers, and economists has been the source of anxiety and debate. The realism of models that rely on such representations, such as those models proposed by decision theory, game theory, and, more recently, belief revision, has given rise to similar worries. However, apart from systems developed specifically to avoid particular postulates, in decision theory for example, much of the work in these domains is still being carried out without taking account of these issues. On the other hand, those working on the problem of the realistic representation of beliefs, in particular bounded rationality, are only beginning to discuss the application of their models in domains where beliefs play an important role, such as economics and game theory (for example [11]). The study of doxastic attitudes and the processes involving them is in need of a general framework that provides a realistic model of beliefs *and* is easily and widely applicable to situations where beliefs are involved. This framework should be powerful and fruitful enough to permit the development of realistic models of the changes they undergo, as well as their role in decision, action and communication.

The purpose of this paper is to propose and motivate such a framework. The tools that shall be proposed for modelling beliefs are designed to be applicable in a vast range of domains. An extended illustration of the framework is contained in

Presented by **Hannes Leitgeb**; *Received* October 26, 2006

DOI: 10.1007/s11225-008-9119-4

a companion paper ([17], henceforth just ‘Part II’), where the case of belief revision is developed in detail.

The choice of belief revision as an illustration has much to recommend it. First of all, the problem of realism has not as yet been treated in the domain, though it has been raised by several theorists. More importantly, the problem of belief revision exemplifies two characteristics which are of central importance to any study which aims to develop realistic models of activities involving belief.

The first characteristic is that there are two facets to 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 models of belief revision [8] as elsewhere, 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 [6, 18]. The second facet concerns the models of *operations involving beliefs*. In decision theory, there has been much discussion about the “realism” of the postulates posed by the principal models of decision. 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 [20]. 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” [21, p. 12].

The second characteristic which is neatly exemplified by the case of belief revision concerns the relationship between these two facets of the problem. In a word, they have to be treated *together*. The point of modelling belief is to use such models to capture what happens to, and with, belief: a model of a belief state will therefore not be “sophisticated” unless it can also support an equally “sophisticated” model of belief dynamics, or of the role of belief in decision, action and communication. Furthermore, every model of what happens to, and with, belief relies on a model of the belief states: a “sophisticated” model of belief change or of the role of belief in action will have to rely on a “sophisticated” model of the belief states at particular moments. Traditionally at least, the question of representing belief realistically and the questions regarding accurate representation of operations involving beliefs have been treated separately; this seems to be a mistake. The challenge is to propose a representation of belief which *not only* accurately models the doxastic state of an agent at a particular moment, *but also* permits a realistic model of the changes in these beliefs, and of their role in action, decision and communication. The proposed framework shall have the machinery to deal with *both* of these facets of the problem posed by realism. In the application to the case of belief revision, this means that it provides not just "sophisticated models of belief formation", but equally "sophisticated models of belief states". Indeed, it will turn out that interesting properties of the revision can be naturally captured relying, at least partially, on a subtle, realistic, representation of the belief states.

Here is a toy example that illustrates the main issues which should be captured by the framework in its application to the case of belief revision.

TOY EXAMPLE. Consider Leon, the head of the centre-left party in the run-up to a general election. He is worried about the far-left parties taking some of his vote; indeed, he is convinced that if his support decreases, it is to them that he is losing out. Several days before the election, he believes that the support for the far-left is high, and that his support is (relatively) low. Then one of his aides tells him, as he speeds past on his way to a rally, that, according to a recent poll, support for the far-left parties has fallen. This cheers him up, because he concludes that he will have gained support. This is a simple case of belief revision — in learning that support for the far-left is low, he revises his belief and comes to believe that support for his party is high. Accordingly, he gives an upbeat speech, further pandering to the far-left vote.

However, in changing his belief in this way, Leon has overlooked the possibility that the far-left support could have fallen because of a general strengthening of the right. Had he kept this possibility in mind, he would have been less certain about the positive way the new information bodes for his party. He would not immediately have jumped to the conclusion that his party was fairing well, and his speech would have been more measured.

This oversight is corrected that evening, when he reads the daily papers and in particular the report on the morning's poll. This report essentially conveys the same information as his aide had earlier that day, namely, that support for the far-left parties has fallen. However, in the discussion of the poll, the report argues that this fall is due to a general strengthening of the right. This possibility, previously overlooked, now comes into play: indeed, taking this possibility and the newspaper's argument into account, and realising that if the right has taken support away from the far-left, it has certainly taken support away from his own party, Leon again revises his belief, becoming more pessimistic about the performance of his party. At first glance, this is not a classic case of belief revision, because it does not seem to involve any new information being learnt: he already knew that the far-left's support had fallen. Indeed, perhaps the only thing that could count as "new" — the general strengthening of the right — cannot really be said to give rise to a revision in the traditional sense, since he was not even aware of this possibility, so it becomes tricky to say to what extent he had a prior attitude to it which was revised.

This example poses at least four problems for the project of realistically modelling belief and belief change. Firstly, there is the fact of overlooking or being unaware of the possibility of losing votes to the right: bounded rationality phenomena such as these pose a challenge to the realism of classical representations of belief states. To the knowledge of the author, there are no models of belief revision which can account for this phenomenon. Secondly, there is the phenomenon of *change* in Leon's awareness (he becomes aware of the possibility of a strengthening of the right on reading the paper): this is an issue involving both belief revision and bounded rationality theory. To the knowledge of the author, not only do existing models of belief revision not deal with such changes, but there is little discussion in the awareness literature of the question of change in awareness. Thirdly, there is the phenomenon of repeated — or, as it is often called, iterated — belief revision (he revises his belief once before the rally and again in the evening). This question has received significant attention in recent literature (see for example [22] and the references therein). Finally, there is the problem of actual changes of belief which do not appear to satisfy the classic postulates of belief revision (the so-called Gärdenfors postulates): the second revision seems to involve the learning of information which Leon already had, and which, according to the Gärdenfors postulates, would yield no change. It is safe to say that most major models of belief revision, and notably all those that satisfy the Gärdenfors postulates (on "factual" sentences, see Literature Remark II.3[1]), cannot capture such phenomena. The model of belief revision developed in the second half of this paper captures all these aspects in a single, articulated framework: awareness and other "bounded" aspects of belief states, changes in these aspects, iterated revision, and apparent violations of the classical postulates.

The example also gives a clearer idea of the general project: to provide a framework for *modelling* agents' *actual* belief states and the *actual* operations that they are involved in. Such a project is not incompatible with the project of furnishing rules or constraints for "rational" behaviour or belief change, but it does not reduce to it; in particular, it cannot restrict itself to belief states and operations conforming to a predetermined notion of rationality. That Leon has overlooked the possibility of losing votes to the right may perhaps count as irrational — that is one question. That it nevertheless has *important* consequences for his decisions and actions (for example, for the speech he makes at the rally) and for the way he changes his beliefs (for example, for the first belief change) — this is beyond doubt. Unawareness does not always imply irrelevance; hence the fact that an agent is unaware of a sentence may lead to a difference in his actions. A realistic model must capture such "im-

[1] Henceforth, the prefix II indicates that the reference is to be found in Part II.

perfections" in the belief states, as well as their consequences. Traditional theories of belief and belief revision, considered as models, fail to capture such behaviour. This is because they have strong rationality constraints — coherence constraints on belief states, for example, or belief revision postulates — built into their technical apparatus, and so must make the assumption that any agent they represent satisfies these stringent standards of rationality. Evidently, a *modelling* project cannot disqualify certain behaviour and particular belief states from consideration just because they do not live up to predetermined standards of rationality, on pain of undermining the project itself; therefore, such assumptions cannot be made. Indeed, the challenge is to find a way of weakening these assumptions without overly weakening the model. Note that, in such a model, it is still possible to speak of rationality in the traditional sense, but the term will only apply to a subset of all the states or operations which can be represented; namely those which satisfy the conditions deemed sufficient for rationality. The question of rationality is however tangential to the concerns of this paper, and shall not be discussed further.

The modelling framework proposed here will be "realistic" in at least two senses. First and foremost, in its motivation: whereas traditional models risk deforming the belief state of the agent by forcing it to conform to the requirements of their technical apparatus (consider the representation of beliefs as sets of possible worlds and the problems it has with logical omniscience; see Section), care will be taken to respect *the agent's point of view*. By taking account of the way the agent sees the world, and the way he understands the changes in his beliefs, the framework, in its very motivation, will be more realistic than most existing ones. Secondly, it will turn out that the traditional models and theories can be recovered as special cases. For example, the traditional models of belief states are special examples of the general model proposed, and the Gärdenfors postulates will apply to certain particular sorts of changes, though not all. In this concrete sense, traditional models appear as *idealisations* of the more general, and realistic, framework proposed here.

The general framework for constructing models of beliefs and of activities involving them shall be developed in two stages. Firstly a representation of the logico-linguistic state at a particular moment shall be proposed and motivated. Then an operation capturing the dynamics of this state shall be defined. This construction will be general and abstract: *no specific model of belief shall be presented*, but just *the basic tools which can be used*, in conjunction with known methods, *to construct realistic models of belief*. In order to illustrate the framework, it shall be applied, in Part II, to the case of belief revision.

1. Interpreted Algebras

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 framework for modelling beliefs proposed here 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 models of beliefs developed with this framework will be more "realistic" or "sophisticated" in that they pay explicit attention to, and indeed represent formally, the local logical structures effective at particular moments, as well as the changes in the structures as new information comes into the fray.

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

Motivation The technical notion used to model local logical structures — *interpreted algebras* — can be motivated by two sorts of failure of logical omniscience. Each of these failures suggests a property that any model of the local logical structure should have. The notion of interpreted algebra will be designed to have these properties; it will thus constitute a single model which can account for both sorts of failure of logical omniscience.

Firstly, consider the notion of a sentence or an issue being *in play* at a particular moment. If, at a 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 take 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"*.

A second important aspect of logical omniscience concerns the failure to recognise the logical equivalence of sentences. For example, 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.

Traditional models of belief cannot account for either of these phenomena. By considering the reasons for this, and the suggestions as to how to represent such phenomena, one can draw important morals for the construction of a model of local logical structures which is capable of dealing with them.

Consider firstly the notion of 'in play'. Traditional models cannot capture this notion 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 both the sentence and its negation believed to be open possibilities — and the notion of 'in play' is orthogonal to this triple distinction. Attempts to account for such phenomena have often 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. Indeed, even models of awareness that take a more "semantic" approach, involving relations between elements of a rich state-space (for example, [10]), generally turn out to be essentially equivalent to these syntactic-styled models [9]. The moral to be taken from the phenomenon of 'in play' is thus the following: 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).

Regarding the failures to recognise logical equivalences, these cannot be represented in traditional frameworks, because logically equivalent sentences express the same proposition, or, to put it another way, they have the same truth values in all possible worlds. Suggestions for dealing with this sort of example often consist in altering the *semantic structure* of the agent's beliefs. The typical example, dating back to [24] 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 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 that he needs to see the ophthalmologist; the fragmentation into these consistent "minds" prevents "logical" conflict between contradictory beliefs. The moral of this sort of phenomenon is thus 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 any *global* logical structure pertaining to some *global* language.

The notion of interpreted algebra will take heed of these two morals, and will therefore be capable of accounting for these two sorts of failure of logical omniscience. It is worth remarking that, to the knowledge of the author, there is no single model of knowledge or belief which is capable of accounting for *both* of these phenomena. On the one hand, models of awareness generally start from a traditional possible worlds model, adding extra structure to deal with awareness, so that logically equivalent propositions are satisfied in the same worlds (or states). Such models cannot capture cases where the agent believes a sentence and not its logical equivalent, even though he is aware of both. On the other hand, "society of minds" models have it that, in each of his "minds", the agent either believes or does not believe every proposition: thus cases of unawareness cannot be captured by such a model. It will thus count as an important advantage of the notion of interpreted algebra that it naturally captures both sorts of phenomena in a single, simple framework.

Definition The preceding considerations indicate that a model of the local logical structure should be composed of two elements: a representation of the local language involved (the set of sentences in play) and a representation of the logical structure on that language. For example, for the classical propositional case that shall be considered here, one could use, following the tradition, a language constructed from a set of propositional letters by closing under Boolean connectives, and a logical consequence relation on the language (or equivalently a set of "worlds"). For reasons to be discussed shortly, algebraic structures shall be preferred as models of the language and the logical structure. More precisely, the following two-levelled structure, called *interpreted algebra*, shall be used.

DEFINITION 1 (Interpreted Algebra). An *interpreted algebra* $\mathbf{B}$ is a triple (B_I, B, q), where:

- B_I is the free Boolean algebra generated by a set I (the *interpreting algebra*);[2]
- B is an atomic Boolean algebra (the *base algebra*);[3]
- $q : B_I \to B$ is a surjective Boolean homomorphism.

An *element* of $\mathbf{B}$ is a pair $(\phi, q(\phi))$, $\phi \in B_I$.

[2]A Boolean algebra is a distributed complemented lattice; the order will be written as $\leqslant$, meet, join, complementation and residuation as $\wedge, \vee, \neg, \to$. The free Boolean algebra generated by a set X shall be noted as B_X for the rest of the paper; details on this and the other notions used in this paper may be found in [19].

[3]Standard terminology is employed here: an atom of a Boolean algebra is an element $a \in B$, such that, for all $x \in B$ with $\bot \leqslant x \leqslant a$, either $x = \bot$ or $x = a$.

The *consequence relation* $\Rightarrow$ is defined as follows: for any elements $(\phi, q(\phi))$, $(\psi, q(\psi))$ of (B_I, B, q), $(\phi, q(\phi)) \Rightarrow (\psi, q(\psi))$ if and only if $q(\phi) \leqslant q(\psi)$. For brevity, this shall often simply be written as $\phi \Rightarrow \psi$. $\Leftrightarrow$ will designate the derived equivalence relation.

The interpreting algebra models the local language effective at the moment in question. I is the set of *locally* atomic sentences in play at that moment (the equivalent of the propositional letters in traditional definitions of languages). These are the sentences that taken as primitive in the local logical structure at the moment in question and from which the other sentences are constructed.

The base algebra is the *local logic* on this language: it is this part of the structure which provides the consequence relation. 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.[4] Elements of the interpreted algebra consist of a sentence and the proposition which it expresses; the consequence relation on elements of the interpreted algebra arises from relations between the propositions they express. Non trivial logical equivalences between sentences 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 base algebra is atomic (see below for a discussion of this property), there is an alternative, "extensional" way of thinking of it. Namely, 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.

An interpreted algebra is thus basically an algebraic representation of a language generated by a set of propositional letters and a set of small worlds where only those sentences of the language get truth values. The important point is that the language (in particular, the set of propositional letters taken as primitive) and the small worlds involved in local logical structures change from one moment to the next: the whole point of representing the language and the logical structure explicitly is to take heed of and account for these changes. It is thus important to capture such changes in terms of relations between local logical structures and operations on them. This is the principal reason for resorting to the algebraic perspective, and in particular to interpreted algebras: the framework offers a rich variety of well-known operations on algebras, and thus turns out to be fruitful for modelling the dynamics of local logical structures, as shall be seen in Section 2.

---

[4]The fact that it is a Boolean homomorphism guarantees that the ordinary conditions on valuations are satisfied.

By the motivation and conception of interpreted algebras, they naturally accommodate 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 is captured by their presence or absence in the interpreting algebra: in the example considered above, the sentence "there are infinitely many primes" is absent from the interpreting algebra at the moment in question. 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, *global* logical relationships between sentences are not respected. In the example given above, the sentences concerning the eye-doctor and the ophthalmologist are mapped to different propositions in the local logical structure at that moment, whereas they express the same proposition when considered as elements of some global logical structure. As anticipated, the ability to account for both sorts of phenomena in a simple fashion in a single framework is an important advantage of the proposed model: most models can deal with either one or the other but not both (see for example the families of models proposed in [6]).

Before discussing the assumptions which are — and which are not — involved in the proposed model, let us give three examples of basic, but important, sorts of interpreted algebra.

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

The *point interpreted algebra* for the sentence ϕ, $\mathbf{B}_{\phi_\mathbf{p}} = (B_{\{\phi\}}, \mathbf{1}, q)$, where $\mathbf{1}$ is the two element Boolean algebra ($\{\top, \bot\}$), and $q : \phi \mapsto \top$.

The *simple interpreted algebra* for the sentence ϕ, $\mathbf{B}_\phi = (B_{\{\phi\}}, \mathbf{2}, q)$, where $\mathbf{2}$ is the four element Boolean algebra ($\{\top, \bot, x, x'\}$), and $q : \phi \mapsto x$.[5]

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 (ϕ) 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 the language (in terms of small worlds, there is one world, where ϕ 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 ϕ holds, the other where $\neg\phi$ holds).

---

[5]Recall (footnote 2) that $B_{\{\phi\}}$ is the free Boolean algebra generated by $\{\phi\}$.

Discussion Interpreted algebras model the local language involved at a given moment by a free Boolean algebra (B_I). It is worth clarifying which assumptions are implied by this choice, and which are not.

The first assumption is that the local language is closed under ordinary linguistic connectives such as 'and', 'or', 'not'. Philosophically, this seems a relatively 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. Indeed, it is accepted by many major logics of awareness (for example [10, 9]).

Secondly, and more importantly, this model of the local logical structure has it that the connectives 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".[6] This is strictly speaking the algebraic equivalent of the traditional definition of a language (as the closure of a set of propositional letters under connectives), since it does not suppose any relations (such as equivalence) between the sentences of the language. However, since a term 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 through the interpreting algebra B_I. So technically speaking, the difference between using term algebras and Boolean algebras is one of ease: 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: not in any overarching language of which the local language may be considered a fragment, for example. Therefore, the model can account for an appearance of a connective which does not obey the Boolean properties by taking the whole sentence featuring the connective, rather than the appropriate clauses, as an element of I. If "it is cold and it is wet" is in play, it does not necessary mean that "it is cold" and "it is wet" are individually implied; indeed,

---

[6]For example, [5, Ch 5].

they may not even be in play as separate linguistic entities. The assumption that the connectives satisfy the ordinary Boolean conditions is weaker than it seems at first blush, since it only applies to those connectives connecting sentences which themselves figure in the local language.

It has already been noted that the use of Boolean algebras will allow a clearer and easier modelling of the dynamics of local logical structures. But, even as concerns the modelling of the local logical structure at a particular moment, the use of a Boolean algebra to model the language may perhaps be argued to be more accurate. In particular, whereas recursive application of Boolean connectives on the elements of (a non-empty set) I yields an infinite set, so that the language defined in the traditional way or the term algebra mentioned above will be infinite, the set of equivalence classes under Boolean equivalence — that is, the set of elements of the interpreting algebra B_I — will be much smaller, and may be finite. Interpreting algebras are intended as tools for modelling local languages at given moments, so the question is whether such languages are faithfully modelled as finite. This certainly does not seem implausible. Indeed, there is a strong intuition according to which the local language at a particular moment is, in practice, finite: not only are there a finite number of "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 such differences as those between '*A and A and A and A*' and '*A*' are naturally discounted. In this paper, where the concern is with realistic modelling, the interpreting algebras shall be taken to be finite. It follows from this assumption that the base algebra B is finite and thus atomic: this may be seen as a justification of the assumption of atomicity of B in Definition 1.

The third point to be made about the model of the local language is the use of a *free* algebra: this 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 (ie. 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 $\Rightarrow$.

A final remark regarding Definition 1 is in order. The aim here is simply to present the general framework. For that reason, only the case of classical propositional logic has been treated (thus the use of Boolean algebras in the definition of interpreted algebra); however, the basic idea can be extended according to taste or requirements. Other propositional logics may be employed, by using different sorts of algebras: for example Heyting algebras instead of Boolean algebras for intuitionnistic logic. A similar point holds for logics with a richer syntax: for example, for first-order logic, cylindrical algebras would be used [12], or for modal logic, the appropriate modal algebras would be relevant [5, Ch. 5]. Generally, the sorts of operations defined in the following section have equivalents for these richer species of interpreted algebra, and the main points made above apply to them (though see Section II.1). From the point of view of this paper, the use of interpreted algebras of the sort given in Definition 1, based on Boolean algebras, receives two different forms of justification. On one hand, the model presented in this section and the next will work on the propositional level, without dealing explicitly with quantification or modal operators such as belief (in particular, there will be no question of beliefs about beliefs; see Section II.1). Cylindrical and modal algebras thus will not be required for the largely expositional and illustrative purposes of this paper, though they would be relevant for more complicated modelling projects. On the other hand, in affirming that the "Boolean" interpreted algebras defined above are accurate models of the local logical structures, it is being assumed that the local logical structure is classical, rather than, say, intuitionnistic. Those who wish to challenge this may use the algebras for the local logic of their choice; for the purposes of this paper it shall be admitted as a modelling assumption without further discussion.[7]

To close this section, let us illustrate how interpreted algebras can be applied to analyse some of the local logical structures involved in the toy example given in the introduction.

TOY EXAMPLE, ANALYSIS. PART 1. At several points in the paper, the notions developed shall be illustrated on the toy example. By the end of the paper, these notions will have provided a full analysis of the example. One preliminary remark is in order. As per normal with analyses of practical examples, there are, at various points in the analysis, several modelling choices. Given the illustrative purpose of the example, and space constraints, at each such point one choice of analysis shall be made and discussed, although it should be emphasised that other choices could be made, and would lead to equally amenable analyses, all of which can be comfortably developed with the sorts of tools and concepts developed here.[8] That said, let us turn to the application of interpreted algebras to the example.

As yet, no equipment has been introduced to model the agent's beliefs (this shall be done in Section II.2), but with interpreted algebras it is already possible to

---

[7]See [16] for some considerations regarding the consequences of this choice.

[8]Perhaps a word of warning is in order for those readers who wish to try for themselves: as shall be noted at the beginning of Section 2, the operations introduced in this article are merely a subset of a larger selection which capture different changes that the local logical structure, and the belief states couched in terms of it, may undergo. Although the most pertinent operations will be discussed here, there are ways of analysing the example that may require, in addition to the techniques presented here, some supplementary operations. A more comprehensive selection of operations on algebras can be found in [13, Ch 5].

model the language and logical structure "in terms of which he is thinking" (and in terms of which his beliefs are couched). At the beginning of the story, Leon has an interpreted algebra containing at least the sentences "support for my party has increased" (call it ϕ) and "support for the far-left has fallen" (ψ): these are in the interpreting algebra. However the possibility of a general strengthening of the right is out of play at this moment: sentences such as χ — "support for the right has increased" — do not figure in the interpreting algebra at this moment. An appropriate interpreted algebra, $\mathbf{B_1^E}$ will have an interpreting algebra generated by ϕ and ψ (ie. $B_{\{\phi,\psi\}}$) and an isomorphic base algebra (ie. with four atoms or small worlds giving the four consistent combinations of ϕ, $\neg\phi$, ψ and $\neg\psi$). The use of this base algebra represents the fact that there is no "logical" relation between the sentences in so far as they appear in Leon's state at that moment. Figure 1 (p106) contains graphical representations of the small worlds of $\mathbf{B_1^E}$, as well as all the other algebras involved in the analysis of the example.

This is not the logical structure which *would* have been appropriate *had* Leon been aware of the threat of the right. Such a structure would contain χ ("support for the right has increased") as well as ψ and ϕ: that is, it could be represented using an interpreting algebra $B_{\{\phi,\psi,\chi\}}$. Furthermore, in such a structure, there would be "logical" relations that are tacitly accepted by Leon: namely, that if the right's support increases (χ) and the far-left's support decreases (ψ), then his party's support has also decreased ($\neg\phi$), and if the far-left's support has fallen (ψ) and the right has not increased its support ($\neg\chi$), then his party has (ϕ).[9] This is the expression of his recognition of the threat the right poses to his party, as well as to the far-left parties. It is represented in the algebra by the fact that $(\phi \wedge \psi \wedge \chi) \wedge (\neg\phi \wedge \psi \wedge \neg\chi) \Leftrightarrow \bot$; in other words, there are no small worlds where $(\phi \wedge \psi \wedge \chi) \wedge (\neg\phi \wedge \psi \wedge \neg\chi)$ (where support for the far-left has fallen and his party and the right have both gained or both lost support). The base algebra is thus the quotient of the interpreting algebra by $(\phi \wedge \psi \wedge \chi) \wedge (\neg\phi \wedge \psi \wedge \neg\chi)$ (the result of removing the small worlds where $(\phi \wedge \psi \wedge \chi) \wedge (\neg\phi \wedge \psi \wedge \neg\chi)$). Let us call this interpreted algebra $\mathbf{B_{1'}^E}$; it is presented in Figure 1. Note that, already at this basic level, the importance of the sentence χ, and the difference which it would make if it were in play, is apparent: in $\mathbf{B_1^E}$, ϕ and ψ are logically independent, whereas in $\mathbf{B_{1'}^E}$, there is a logical dependence between ϕ and ψ (in particular, in the parts of the algebra where χ).

[9]Strictly speaking, $\neg\phi$ is the sentence "support for Leon's party has not increased"; however, for the sake of simplicity, the negation of an increase will be assimilated with a decrease (and vice versa) at several points of the analysis. It is not difficult to carry out a more complicated analysis which acknowledges the difference between a decrease and the negation of an increase with the tools afforded by the framework.

2. Fusion

Investment in a model which captures the logical imperfections of an agent's instantaneous belief state seems worthless if it is not accompanied by an account of how this state changes. Most models of "awareness" are proposed without an account of the frequent changes of awareness from one moment to the next. Models of the agent's "several minds" do not generally include an understanding of how these "minds" interact with each other over time. As discussed in the introduction, this is a mistake: static models of belief cannot be properly evaluated and defended without considering their ability to account for changes of belief, or more generally for the role of beliefs in decision, action and communication. As models of local logical structures, it has already been noted that interpreted algebras have the advantage of being able to deal with different sorts of lack of logical omniscience in a single framework. They have an even more important advantage: they permit a rich and sophisticated theory of their dynamics.

Indeed, one may imagine a family of types of change which the local logical structure may undergo. Interpreted algebras admit a family of relations and operations, based on more or less standard algebraic notions. It turns out that the latter provide adequate models of the former. For the purposes of this paper, it is sufficient to consider only one basic operation: the main one involved in modelling change in the local logical structure as new information comes into play. This is the operation of *fusion*, which shall be defined in this section. The model of belief revision proposed in the Part II will be based on this operation. For an account of other relevant operations on interpreted algebras, see [13, 15].[10]

The sorts of changes to local logical structures which shall be considered here are those brought about by the incoming information. At the abstract level of the discussion in this part of the paper, this incoming information could be information to be accepted from the environment, sentences communicated, or even the form of a decision problem which the agent finds himself in. Very often, and especially in models of belief change, such information is considered as coming in the form of a sentence (or a set of sentences) of the language.[11] However, no global or overarching language is assumed in the current framework; indeed, given that the only

[10] Evidently, although the fusion operation defined and discussed here will be sufficient to capture the main aspects of belief revision, it by no means exhibits the full power and interest of the framework. To give but one example, sentences can fall out of play as well as enter into play. The fusion operation represents a way in which they enter into play (this is, after all, the most pertinent case for belief revision), but it is not appropriate for modelling cases where they fall out of play: for such cases, other, equally natural, operations are appropriate (see the references mentioned).

[11] This is often the case not only in belief revision (for example [8, 25]), but also in Bayesian update theory [23] and in models of communication (for example [7, 26]). Moreover, it is arguable that even the information forming a decision problem can be represented as coming in such a format.

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. In order to model the new information, it is first necessary to represent the fragment of language in which it is couched, 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 context of a local language with a local logic — a local logical structure, to use the term introduced in the previous section. A model has already been proposed for local logical structures such as these: they shall be modelled using interpreted algebras.

Indeed, the flexibility of the notion of interpreted algebra permits it to capture the variety of more or less complicated forms which incoming information might take. At one end of the spectrum, rich local languages (large B_I) with interesting logical structures (B and q) can accurately model an "input" which does not consist of a simple sentence, but comprises a complex of different items of 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). The revisions involved in the toy example illustrate the more or less complex logical contexts in which the new information may be couched, as well as the ease with which interpreted algebras can capture this diversity of forms.

TOY EXAMPLE, ANALYSIS. PART 2. As for the case of the belief state (part 1 of the analysis), not enough structure has yet been introduced to model the information learnt by the agent; only the tools necessary to model the local logical structure in which this information is couched — namely, the notion of interpreted algebra — have been introduced. The incoming information in the first revision in the example (when the aide relates the news regarding the support for the far-left) consists simply of the sentence ψ ("support for the far-left has fallen") with no complicated relevant accompanying structure; it comes in a "null context", so to speak. It can be modelled, standardly, by the appropriate simple or point algebra, depending on whether the possibility of $\neg\psi$ is left open or not (see Example 1). For the rest of the analysis, the point algebra $\mathbf{B}_{\psi_\mathbf{p}}$ shall be used.[12]

Now consider the second revision, where the newspaper report conveys the information that the far-left's support has fallen, with an argument that this is due to a strengthening of the right: the "essential" information — the far-left's support has fallen — comes in a complex "context" of more or less connected reasons, arguments, assumptions and hypotheses. A richer interpreted algebra is required to represent this information: it will at least have to contain the sentences ψ and χ

[12]The case of the simple algebra is similar. See the remarks at the beginning of part 1 of the analysis.

("support for the right has increased"). However, just as a point algebra was used to represent the first update, where ψ was communicated as certain, an algebra which does not contain any $\neg\psi$-small worlds can be used to model this second update, which also conveys the information ψ. A model of the logical structure will thus be the interpreted algebra $\mathbf{B_3^E}$ with interpreting algebra freely generated by ψ and χ (ie. $B_{\{\psi,\chi\}}$), and base algebra the quotient of this algebra by $\neg\psi$ (ie. containing two small worlds where ψ holds). This is shown in Figure 1 (p. 106).

Once the local logical structure in which the new information is couched is modelled by an interpreted algebra, the change in the face of new information takes the form of an operation sending a pair of interpreted algebras — one modelling the current local logical structure, the other that of the new information — to another interpreted algebra — the local logical structure after 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 algebra*. 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, or are equivalent to sentences belonging to this local logical structure. The representation of the prior local logical structure and that accompanying 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 modelled using an appropriate relation — called *identification* — defined as follows.

DEFINITION 2 (Identification between algebras). For $\mathbf{B_1} = (B_{I_1}, B_1, q_1)$, $\mathbf{B_2} = (B_{I_2}, B_2, q_2)$ two interpreted algebras, $\simeq = (\simeq_I, \simeq_B)$ is an *identification* of sentences between the two algebras if:

(i) $\simeq_I \subset |B_{I_1}| \times |B_{I_2}|$ and $\simeq_B \subset |B_1| \times |B_2|$ are both restrictions of congruence relations in the following sense.
For C_1, C_2 Boolean algebras, $\sim \subset |C_1| \times |C_2|$, let $\sim^e \subset |C_1 \otimes C_2| \times |C_1 \otimes C_2|$ be the image of $\sim$ in $|C_1 \otimes C_2| \times |C_1 \otimes C_2|$: that is, $x \sim^e y$ iff $\exists x' \in C_1, y' \in C_2$ such that $x = e_1(x')$, $y = e_2(y')$ and $x' \sim y'$. $\sim$ is a *restriction of a congruence relation* if and only if there exists a congruence relation $\bar{\sim}$ on $C_1 \otimes C_2$, such that $\sim^e = \bar{\sim}|_{|e_1(C_1)| \times |e_2(C_2)|}$.[13]

[13]Following ordinary notation, $|B|$ is the underlying set of the algebra B, $B \otimes C$ is the free product of B and C and e_1 (respectively e_2) is the canonical homomorphism from B (resp. C) into $B \otimes C$. A congruence relation $\sim$ on an algebra B is a relation on the elements of $|B|$ (that is, a subset of $|B| \times |B|$)

(ii) $\simeq_I$ is generated by its restriction to $(I_1 \times |B_{I_2}|) \cup (|B_{I_1}| \times I_2)$: for ϕ any element of B_{I_1} and ψ any element of B_{I_2}, with normal disjunctive conjunctive forms $\phi = \bigvee_{m \in M}(\bigwedge_{j \in J_m} \phi_j)$, $\phi_j \in I_1$ and $\psi = \bigvee_{n \in N}(\bigwedge_{k \in K_n} \psi_k)$, $\psi_k \in I_2$ respectively,[14] if $\phi \simeq_I \psi$ then either: for each $j \in \bigcup_{m \in M} J_m$, there is a $\hat{\psi}_j \in B_{I_2}$ such that $\phi_j \simeq_I \hat{\psi}_j$ and $\psi = \bigvee_{m \in M}(\bigwedge_{j \in J_m} \hat{\psi}_j)$, or: for each $k \in \bigcup_{n \in N} K_n$, there is a $\hat{\phi}_k \in B_{I_1}$ such that $\hat{\phi}_k \simeq_I \psi_k$ and $\phi = \bigvee_{n \in N}(\bigwedge_{k \in K_n} \hat{\phi}_k)$.

(iii) $\simeq_B$ contains $\widehat{\simeq_I} \subset |B_1| \times |B_2|$, where $X \widehat{\simeq_I} Y$ iff $\exists \phi \in B_{I_1}$, $\psi \in B_{I_2}$, $q_1(\phi) = X$, $q_2(\psi) = Y$, and $\phi \simeq_I \psi$.

Explication In accordance with the goal of the article, identification relations are considered as tools for modelling the relations between the elements of local logical structures. They will be thus be specified in the application of these techniques to modelling particular situations; as is usual for such cases, the correctness of the application will depend, partially, on the correctness of the choice of identification relation. The clauses of the definition shall now be motivated, with an aim to suggesting the adequacy of this notion as a modelling tool.

Local logical structures are modelled by interpreted algebras, which have two components: the local language (interpreting algebra) and the logical structure on that language (base algebra). Similarly, an identification relation, relating elements of different interpreted algebras, has two components: a relation between the interpreting algebras (that is, between sentences of the local languages), and a relation between the base algebras (that is, between propositions of the local logical structures).

An interpreting algebra represents not only the sentences of the local language, but also the way these sentences are constructed from locally primitive or atomic sentences (elements of I). Clause (ii) effectively demands that any identification of sentences must respect, to a minimal extent, the way they are constructed from, or parsed into, more primitive sentences. It is permissive enough to allow the following identifications: *A and B*, an atomic sentence of one algebra (ie. in I_1) with $A \wedge B$, where A and B are atomic sentences of the other algebra (ie. in I_2). It however prevents identifications such as the following: $(A$ *and* $B) \vee C$, where $(A$ *and* $B)$ and C are atomic sentences of one algebra, with $(A$ *or* $C) \wedge (B$ *or* $C)$, where $(A$ *or* $C)$ and $(B$ *or* $C)$ are atomic sentences of the other algebra. Intuitively, this seems plausible: how can two sentences be considered identical if they are considered as the conjunction and disjunction respectively of incomparable atomic

which is transitive, symmetric, reflexive and closed under Boolean operators (for example, if $x \sim y$ then $\neg x \sim \neg y$). See [19] for details.

[14] For the existence of disjunctive conjunctive form in free Boolean algebras, see any basic logic textbook.

sentences? Technically, this clause assures that the fusion operation will yield a free interpreting algebra (see the proof of Observation 1).[15]

Note that, for practical modelling applications, this clause is not overly restrictive: in particular, it does not prevent the representation of the *equivalence* of sentences such as $(A\ and\ B) \vee C$ and $(A\ or\ C) \wedge (B\ or\ C)$. Such an equivalence can be captured by the relation $\simeq_B$ between the propositions of the local logical structures (the elements of the base algebra). Clause (iii) demands that one identify *at least* the propositions whose sentences have been identified: this is a minimal consistency constraint on the notion of identification between sentences and between propositions of these double-levelled structures. However, it allows the possibility of further identifications between propositions. This allows the notion of identification relation to represent logical equivalences between propositions in different interpreted algebras which do not correspond to syntactically identical sentences, such as the $(A\ and\ B) \vee C$ and $(A\ or\ C) \wedge (B\ or\ C)$ mentioned above (in this case, by a relation with $q_1((A\ and\ B) \vee C) \simeq_B q_2((A\ or\ C) \wedge (B\ or\ C))$). Another example of such an extended $\simeq_B$ occurs in the toy example; see part 3 of the analysis.

It remains to motivate Clause (i). One expects the relation of identity or equivalence between sentences to be an equivalence relation respecting the (in this case, Boolean) connectives; that is, one expects it to be a congruence relation. However, each component of the identification relation only relates elements belonging to *different* algebras (B_{I_1} and B_{I_2}, for example), and thus cannot be a congruence relation (for example, it is not reflexive). Clause (i) expresses the next best thing: each component of the identification relation is the restriction of a congruence relation, to the set of pairs, one element from each algebra.[16]

Note finally that this definition can be extended canonically to cases of more than two interpreted algebra and to the case of a single interpreted algebra. In particular, by setting $\mathbf{B_1} = \mathbf{B_2} = \mathbf{B}$, one obtains a notion of identification relation on a single interpreted algebra, as a pair of congruence relations (on the interpreting and base algebras respectively) satisfying (ii) and (iii).[17]

---

[15] Issues such as this arise naturally in the proposed framework, though they are not often discussed in the literature, because the point is to explicitly represent and understand the differences between local languages involved at different moments.

[16] An equivalent formulation would be (for the $\simeq_I$ case): 1. for $\phi, \phi' \in |B_{I_1}|$, $\psi, \psi' \in |B_{I_2}|$, if $\{\chi | \phi \simeq_I \chi\} \cap \{\chi | \phi' \simeq_I \chi\} \neq \emptyset$, then $\{\chi | \phi \simeq_I \chi\} = \{\chi | \phi' \simeq_I \chi\}$, and if $\{\pi | \pi \simeq_I \psi\} \cap \{\pi | \pi \simeq_I \psi'\} \neq \emptyset$, then $\{\pi | \pi \simeq_I \psi\} = \{\pi | \pi \simeq_I \psi'\}$; and 2. $\simeq_I$ respects the Boolean connectives: if $\phi \simeq_I \phi'$, $\psi \simeq_I \psi'$, then $\phi \wedge \psi \simeq_I \phi' \wedge \psi'$; if $\phi \simeq_I \phi'$, then $\neg\phi \simeq_I \neg\phi'$.

[17] That is, for (ii): if $\phi \simeq_I \psi$, then, either, for $\phi = \bigvee_{m \in M}(\bigwedge_{j \in J_m} \phi_j)$ the normal disjunctive conjunctive form, for each $j \in \bigcup_{m \in M} J_m$, there is a $\hat{\psi}_j \in B_{I_2}$ such that $\phi_j \simeq_I \hat{\psi}_j$ and $\psi = \bigvee_{m \in M}(\bigwedge_{j \in J_m} \hat{\psi}_j)$, or the equivalent for ψ.

TOY EXAMPLE, ANALYSIS. PART 3. Consider the identification relation involved in Leon's second revision: that is, the identification relation $\simeq$ between $\mathbf{B_2^E}$ and $\mathbf{B_3^E}$ (see Figure 1 and parts 2 and 4 of the analysis). It is composed of two elements. The first, $\simeq_I$, is a relation between the elements of the two interpreting algebra, and will relate "the same" sentences figuring in the different algebras: in this case, it will relate ψ in the former with ψ in the latter. The other element, a relation $\simeq_B$ between the base algebras, relates the "equivalent propositions" (elements of base algebras). This relates the "propositional" elements of the two ψ ($q_2(\psi) \simeq_B q_3(\psi)$), but goes further in this case, because there are non-trivial equivalences between the elements of the different algebra. In particular, given that the far-left's support has fallen (ψ), if the right's support has increased (χ) then Leon's party's support has also fallen ($\neg\phi$), whereas, if the right has not increased its support ($\neg\chi$), then his party has (ϕ). In Leon's consideration of the question at that moment, these relations between the sentences (and thus states of affairs) are assumed to hold; hence they shall be modelled by the identification relation which purports to represent such equivalences between sentences.[18] An appropriate $\simeq_B$ will thus be such that $q_2(\neg\phi \wedge \psi) \simeq_B q_3(\chi \wedge \psi)$. The identification relation can thus be modelled by the $\simeq_I$ above and, for $\simeq_B$, the minimal relation containing $(q_2(\neg\phi \wedge \psi), q_3(\chi \wedge \psi))$ and satisfying the conditions of Definition 2.

The recognition of such identifications between local logical structures has consequences 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 elements between them, and yielding an interpreted algebra which respects the identification of the elements. The operation of *fusion* of interpreted algebras does just this. It can be defined from two simple operations on interpreted algebras.

The first is the operation of free product.

DEFINITION 3 (Free Product of interpreted algebras). The *free product* of interpreted algebras $\mathbf{B_1} = (B_{I_1}, B_1, q_1)$ and $\mathbf{B_2} = (B_{I_2}, B_2, q_2)$ is $\mathbf{B_1} \otimes \mathbf{B_2} = (B_{I_1 \uplus I_2},\ B_1 \otimes B_2,\ q_1 \otimes q_2)$, where $\otimes$ is the free product on Boolean algebras and Boolean homomorphisms respectively.[19]

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. At the level of the logical structure, the set of small worlds or states in the resulting

[18] See the remarks at the beginning of part 1 of the analysis.

[19] $B_{I_1 \uplus I_2} = B_{I_1} \otimes B_{I_2}$, 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 [19].

interpreted algebra is the cartesian product of the sets of small worlds or states of the initial algebras;[20] the valuation on these worlds 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 the algebras is the well-known operation of quotient. In the current case, it is defined as follows.

DEFINITION 4 (Quotient of an interpreted algebra). The *quotient* of an interpreted algebra $\mathbf{B} = (B_I, B, q)$ by an identification relation $\simeq$ on $\mathbf{B}$ is the interpreted algebra $\mathbf{B}/\simeq\ = (B_I/\simeq_I,\ B/\simeq_B,\ q_\simeq)$, where

- $B/\sim$ is the quotient of the ordinary Boolean algebra B by the congruence relation $\sim$;
- $q_\simeq([\phi]_{\simeq_I}) = [q(\phi)]_{\simeq_B}$.

OBSERVATION 1. *Quotient is a well-defined operation.*[21]

Two different elements of a Boolean algebra that 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 $\simeq_I$-equivalent in the initial local language; that is to say, such sentences in B_I have a common image in $B_I/\simeq_I$. In terms of the local logic, the propositions which are $\simeq_B$-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 that are witness to differences between any pair of $\simeq_B$-equivalent elements x and y; that is, worlds where the valuations of x and y differ.[22]

The free product of two interpreted algebras puts them together, without identifying any of the elements between them. The quotient identifies elements which are $\simeq$-equivalent in an interpreted algebra. The operation of *fusing* two interpreted algebra, whilst respecting the identification of elements 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.

[20] For 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 $S_1 \times S_2$.

[21] Proofs of this and other statements are to be found in the Appendix.

[22] Formally: the s with $s \leqslant x \vartriangle y$, where $\vartriangle$ is the symmetric difference, are removed in the quotient operation.

DEFINITION 5 (Fusion $*$). Given two interpreted algebras $\mathbf{B_1}$ and $\mathbf{B_2}$, with an identification relation $\simeq$ between them, the *fusion* of the two algebras respecting this relation is defined as:

$$\mathbf{B_1} \ *_{\simeq} \ \mathbf{B_2} = (\mathbf{B_1} \otimes \mathbf{B_2})/\bar{\simeq}$$

where $\bar{\simeq} = (\bar{\simeq}_I, \bar{\simeq}_B)$ is an identification relation on $\mathbf{B_1} \otimes \mathbf{B_2}$ with $\bar{\simeq}_I$ the smallest congruence relation containing $\simeq_I^e$ (see Definition 2, clause (i)), and similarly for $\bar{\simeq}_B$.[23]

In subsequent discussion, canonical identification relations shall be assumed between interpreted algebras: that is, unless specified, the identification relations shall only make identifications between homophonic sentences, with the generated identifications between propositions. The fusions will then be written simply as $\mathbf{B_1} \ * \ \mathbf{B_2}$.

This operation models the change in the local logical structure in the face of new information coming in its own local logical structure: both the original local logical structure and the structure accompanying the new information are modelled by interpreted algebras; the resulting local logical structure is the resulting interpreted algebra. This model is intuitive: in fusing the fragment of language in which the new information is couched with the existing logical structure, the "sum" of the two logical structures is taken (free product), and then appropriate elements figuring in the different logical structures are identified (quotient). Given that the operation to be modelled is that of "merging" or "combining" two fragments of language (to which the appropriate structure for representing beliefs or information has yet to be added) one would expect it to be commutative: no priority should be given to one over the other. The operation $*$ has this property.

Two examples shall serve to illustrate this sort of operation.

EXAMPLE 2. For ϕ in $\mathbf{B}$, the fusions with the simple and point algebras for ϕ (Example 1) are as follows:[24]

Simple algebra $\mathbf{B} \ * \ \mathbf{B}_\phi$ is isomorphic to $\mathbf{B}$;

Point algebra $\mathbf{B} \ * \ \mathbf{B}_{\phi_\mathbf{p}}$ is isomorphic to $(B_I, B/(q(\neg\phi)), q')$, where $B/(q(\neg\phi))$ is the quotient of B by the smallest congruence relation such that $q(\neg\phi) \simeq \bot$, and q' the composition of q with the quotient homomorphism.

The first example shows that 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.

[23] It is easily seen that $\bar{\simeq}$ is an identification relation on $\mathbf{B_1} \otimes \mathbf{B_2}$.

[24] As specified in Definition 5, the canonical identification relation is assumed.

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 sentence (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 sentence (or alternatively true in all small worlds).

Relation to the Literature, Remark 1. The fusion operation proposed here is technically very similar to the operation involved in models of updates by epistemic programs, used to model knowledge change in situations of communication [1, 2], and, more recently, belief revision [3, 4] (Part II contains further consideration of the belief revision case). Indeed, from a technical point of view, there seem to be two basic differences. The first concerns the richness of the machinery. On the one hand, at this stage of the paper, only the local logical structures have been considered, and no machinery has been added to model beliefs. So, for example, the fusion operation defined above only corresponds to the part of the epistemic program update product which does not involve operations on the elements of their models which model knowledge and belief (the "accessibility relations" and "plausibility relations"). This gap shall be partly filled in Part II, where machinery similar to that employed in [3] is added to represent beliefs (see Section II.2 and Literature Remark II.1). On the other hand, only the propositional case is discussed in this paper, and so there are none of the modal operators present in models for epistemic programs. As remarked in Section , this is but an instance of a general framework that is intended to be naturally applicable to richer languages, such as languages containing knowledge or belief operators (but see the discussion in Section II.1). To the extent that [1] represents a natural way of carrying out this extension (and in particular, one, but not necessarily the only way of defining the fusion of modalities and applying the framework to the case of communication), its success may be counted as extra support for this framework.

The second technical difference with respect to models of epistemic programs is the notion of local language, which is completely absent from the logic of epistemic programs and related logics of belief revision, and indeed, most modern logics (to the knowledge of the author). This difference is related to dissimilarities in the interpretation of the technically similar aspects of the model. In the current proposal, all the algebras involved are considered to be fragments of logical structure, in which attitudes, such as belief or knowledge, and effects on them, such as communicative acts or incoming information, will be couched (via the addition of supplementary structure). By contrast, in the epistemic logic framework developed in [1, 2, 3, 4], there is one fixed language (for a given signature), and this is essentially interpreted on the structure representing the agents' initial states of knowledge or belief (the "epistemic state model" or "epistemic plausibility model"). The other structure (the "action model"), by which one effectuates the update, is not regarded as a linguistic structure at all: its elements are seen as "actions", and the relationships between these actions are supposed to capture the structure of the communicative act [1, §4]. The relationship between these actions and the language is expressed in their "preconditions", which effectively determine which elements of the state model and the action model are "compatible" (that is, the states in which a given action can be carried out), in much the same way that the identification relation defined above determines which elements of the respective interpreted algebras are incompatible (that is, if two elements are to be identified, then the conjunction of one with the negation of the other is "inconsistent" – equivalent to the false element in the fusion).

The difference in the role of the language is important in the context of this paper, since, as argued in Section , the use of a local language is essential for the realistic modelling of beliefs and activities involving them. These concerns are absent from the epistemic programs literature, which accordingly assumes a single global language. Besides this issue of the locality of the languages, it may seem that the difference in the interpretation of the incoming structure as linguistic or as consisting of "actions" is one of terminology rather than of fundamentals. However, it is not clear that it is so. First of all, as shall be seen in the case of belief revision discussed in Part II, this difference results in diverging conceptions of the model of belief revision (see Literature Remark II.4). Furthermore, there may be consequences for the extension of the apparatus proposed here. To take but a simple example, a relation of subalgebra can be defined on interpreted algebras, and used to understand the idea of a logical structure being richer or poorer than another.[25] This operation is more naturally interpreted in the perspective proposed here than in that of [1]: it would seem perverse to check, before deciding on the comparative richness of structures, whether the structures involved are linguistic (or "state") structures or "action" structures.

Let us close by assembling the parts of the discussion of the toy example, to see how the notions can be used to model the changes in Leon's local logical structures.

TOY EXAMPLE, ANALYSIS. PART 4. Consider the first revision. Modelling the initial local logical structure by $\mathbf{B}_{\mathbf{1}}^{\mathbf{E}}$ (see part 1 of the analysis) and the local logical structure in which the new information is couched by the point algebra $\mathbf{B}_{\psi_{\mathbf{p}}}$ (see part 2), fusion results in the interpreted algebra $\mathbf{B}_{\mathbf{2}}^{\mathbf{E}}$, with the same interpreting algebra as $\mathbf{B}_{\mathbf{1}}^{\mathbf{E}}$, but with the base algebra only containing the two states where ψ

[25] See [13, Ch 5] and [15] for more details on this operation and its importance.

(see Example 2 and Figure 1). This represents the local logical structure after the first revision.

On the other hand, had Leon been aware of the possibility of losing votes to the right — had his logical structure at that moment been $\mathbf{B}^{\mathbf{E}}_{\mathbf{1}'}$ instead of $\mathbf{B}^{\mathbf{E}}_{\mathbf{1}}$ (see part 1) — the revision by the point algebra would have yielded an algebra with the same interpreting algebra as $\mathbf{B}^{\mathbf{E}}_{\mathbf{1}'}$ (that is, involving χ, ψ and ϕ), but with only two small worlds: the remaining worlds where ψ holds. This is the algebra $\mathbf{B}^{\mathbf{E}}_{\mathbf{4}}$ shown in Figure 1.

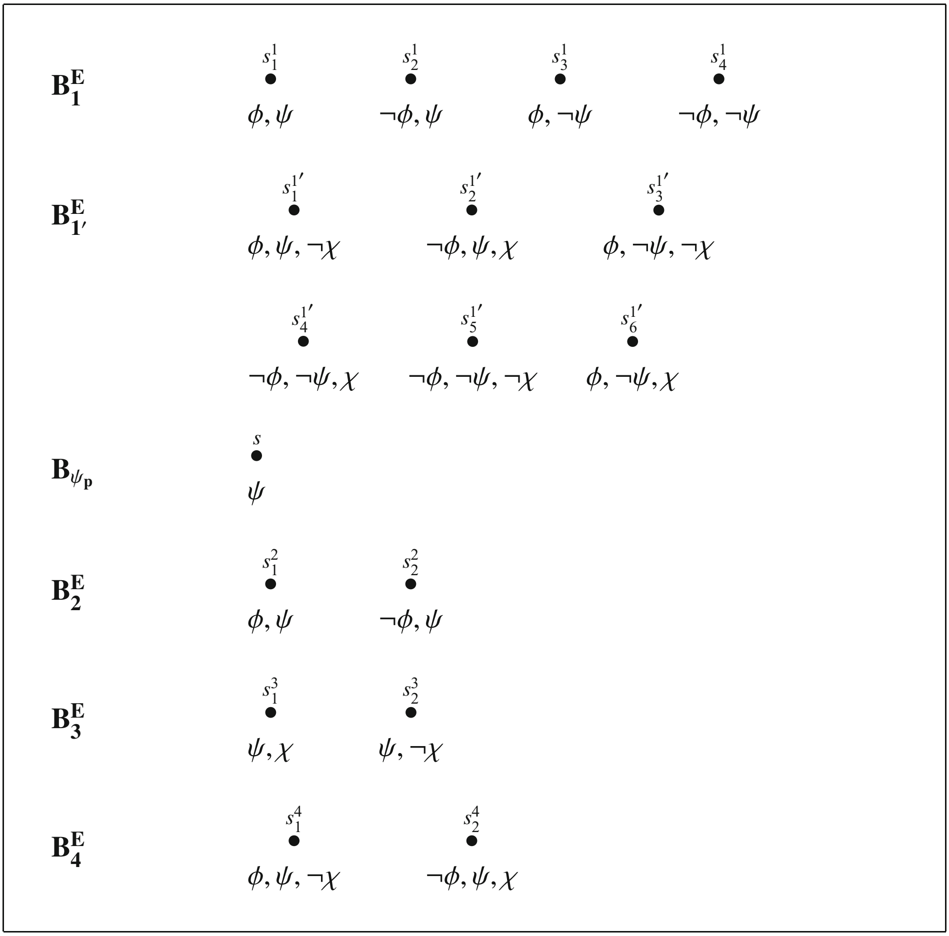
Now consider the second revision, where the local logical structure $\mathbf{B}^{\mathbf{E}}_{\mathbf{2}}$ is revised with a structure represented by $\mathbf{B}^{\mathbf{E}}_{\mathbf{3}}$ (see part 2), with the identification relation $\simeq$ between them (see part 3). Regarding the interpreting algebra, the two instances of ψ (from $\mathbf{B}^{\mathbf{E}}_{\mathbf{2}}$ and $\mathbf{B}^{\mathbf{E}}_{\mathbf{3}}$) are identified by $\simeq_I$, so the new interpreting algebra is $B_{\{\phi,\psi,\chi\}}$: all three sentences are now in play. However, because of the element of the identification relation pertaining to the base algebras — $\simeq_B$ — the base algebra of the resulting algebra will not just be the free product of the two initial base algebras. In other words, because $q_2(\neg\phi \wedge \psi) \simeq_B q_3(\chi \wedge \psi)$, some of the small worlds — those where $(\phi \wedge \psi \wedge \chi) \vee (\neg\phi \wedge \psi \wedge \neg\chi)$ — are removed in the quotienting part of the fusion operation. Fusion thus gives the algebra $\mathbf{B}^{\mathbf{E}}_{\mathbf{4}}$, shown in Figure 1. Note that this is the same algebra as the one Leon would have arrived at, after his first revision, had he been aware of the right's threat. The analysis gets this right: the second revision adds logical or linguistic structure to the situation only because Leon had overlooked the possibility of the right gaining votes (χ).

The task of this paper 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 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 the lack of recognition of intensional equivalence. The operation of fusion of interpreted algebras is the principal operation required to model the change in the local logical structure in the face of input couched in its own local logical structure.

This framework is abstract, and intentionally so. As anticipated at several points in the preceding discussion (see notably the introduction and Literature Remark 1), it can be applied to several different questions in several different fields; in each, the basic notions may assume a different interpretation. Examples of applications may include context and conversation, knowledge and communication, belief and decision, counterfactuals and nonmonotonic logic.[26] And belief and its revision. Part II, intended as an extended example of the application of this sort

[26] For the first case, see [13, Ch 5] for the second, see the references given in Literature Remark 1; for the third, see [14]; for the fourth, see [13, Ch 5, Appendix].

Figure 1. Toy Example: Representations of the interpreted algebras



of framework to phenomena already considered by logicians, proposes a model of belief and belief revision which is based on the general framework introduced here.

Appendix

PROOF OF OBSERVATION 1. **Interpreting algebra** Since $\simeq_I$ is a congruence relation, $B_I/\simeq_I$ is a Boolean algebra. It remains to show that it is a free algebra. To do this, construct a set which freely generates it as follows. Let a set $\Psi \subseteq I$ be called *straight* if no Boolean combination of the elements of Ψ is $\simeq_I$-equivalent to $\perp$; such sets clearly exist (singletons are examples). Define the relation $\leqslant$

on subsets of I as follows: $\Psi \leqslant \Psi'$ iff, for each element $\psi' \in \Psi'$, there exists a $\phi \simeq_I \psi'$ whose disjunctive conjunctive normal form features only elements in Ψ. Note that if $\Psi' \subseteq \Psi$, then $\Psi \leqslant \Psi'$. Now take any $\leqslant$-minimal straight set Ψ. The claim is that this set freely generates $B_I/\simeq_I$. First of all, since Ψ is straight, the algebra generated by it is (mapped under the quotient to) a free subalgebra of $B_I/\simeq_I$. Secondly, the image of this algebra in the quotient is the entirety of $B_I/\simeq_I$: suppose not, that is, suppose that there is a $\psi \in B_I$ such that ψ is not $\simeq_I$-equivalent to any element of B_Ψ. $\Psi' = \Psi \cup \{\psi\}$ cannot be a straight set, since this would contradict the minimality of Ψ. Note that, by Clause (ii) of Definition 2 (see also footnote 17), any $\simeq_I$-equivalence between a Boolean combination of elements of Ψ' and $\perp$ must be generated by $\simeq_I$-equivalences between individual elements of Ψ' and some Boolean combination of other elements of Ψ'. Since Ψ is straight, and since ψ is not $\simeq_I$-equivalent to a combination of elements of Ψ, there must thus be a $\phi \in \Psi$ which is $\simeq_I$-equivalent to a some combination of elements of $\Psi \setminus \{\phi\}$ and necessarily ψ. Thus $\Psi'' < \Psi$, where $\Psi'' = \Psi' \setminus \{\phi\}$. If Ψ'' is straight, this contradicts the minimality of Ψ; if not, repeat the above reasoning to obtain a straight set which contradicts the minimality of Ψ (one necessarily arrives at such a set, since this paper works with the case where I is finite).

Base algebra $\simeq_B$ is a congruence relation, therefore $B/\simeq_B$ is a Boolean algebra. Since B is assumed to be finite in this paper (Section), $B/\simeq_B$ is atomic.

Surjective homomorphism By Clause (iii) of Definition 2, if $[\phi]_{\simeq_I} = [\psi]_{\simeq_I}$, then $[q(\phi)]_{\simeq_B} = [q(\psi)]_{\simeq_B}$, so $q_\simeq$ is well-defined. It is straightforward to show that it is a Boolean homomorphism. Finally, for any $[x]_{\simeq_B} \in B/\simeq_B$, there is a $x \in [x]_{\simeq_B}$ and $\phi \in B_I$ such that $q(\phi) = x$, by the surjectivity of q. $q_\simeq([\phi]_{\simeq_I}) = [x]_{\simeq_B}$. So $q_\simeq$ is surjective.

$(B_I/\simeq_I, B/\simeq_B, q_\simeq)$ is an interpreted algebra; furthermore, by uniqueness of quotients on Boolean algebra, it is the only algebra resulting from these operations. The quotient is thus well-defined. ∎

Remark 1. It is easily seen from the proof that the set generating the interpreting algebra is not necessarily unique; there is thus the option, in the case of fusion where this result is used (Definition 5), to select it such a way as to give priority to the "primitivity" of sentences in the initial local logical structure, or, conversely, to give priority to those in the new local logical structure.

Acknowledgements. Early versions of this paper were presented at Logica 2006 and the Studia Logica Trends in Logic IV 2006 conference. The author wishes to thank the respective audiences, as well as two anonymous referees, for their useful comments.

References

[1] BALTAG, ALEXANDRU, and LAWRENCE S MOSS, 'Logic for epistemic programs', *Synthese*, 60 (2004), 1–59.

[2] BALTAG, ALEXANDRU, LAWRENCE S MOSS, and SLAWOMIR SOLECKI, 'The logic of common knowledge, public announcements and private suspicions', in I Gilboa, (ed.), *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, 1998, pp. 43–56.

[3] BALTAG, ALEXANDRU, and SONJA SMETS, 'Dynamic belief revision over multi-agent plausibility models', in Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, (eds.), *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT06)*, 2006, pp. 11–24.

[4] BALTAG, ALEXANDRU, and SONJA SMETS, 'The logic of conditional doxastic actions: A theory of dynamic multi-agent belief revision', in S Artemov, and R Parikh, (eds.), *Proceedings of the Workshop on Rationality and Knowledge ESSLLI 2006*, 2006, pp. 13–30.

[5] BLACKBURN, PATRICK, MAARTEN DE RIJKE, and YDE VENEMA, *Modal Logic*, Cambridge University Press, Cambridge, 2001.

[6] FAGIN, RONALD, and JOSEPH Y HALPERN, 'Belief, awareness, and limited reasoning', *Artificial Intelligence*, 34 (1988), 39–76.

[7] GERBRANDY, JELLE, and WILLEM GROENEVELD, 'Reasoning about information change', *Journal of Logic, Language, and Information*, 6 (1997), 147–169.

[8] GÄRDENFORS, PETER, *Knowledge in Flux : Modeling the Dynamics of Epistemic States*, MIT Press, Cambridge, MA, 1988.

[9] HALPERN, JOSEPH Y, and LEANDRO CHAVES RÊGO, 'Interactive awareness revisited', in *Proceedings of Tenth Conference on Theoretical Aspects of Rationality and Knowledge*, 2005, pp. 78–91.

[10] HEIFETZ, A., M. MEIER, and B. SCHIPPER, 'Interactive unawareness', *Journal of Economic Theory*, 130 (2006), 78–94.

[11] HEIFETZ, A., M. MEIER, and B. SCHIPPER, 'Awareness, beliefs and games', in *Theoretical Aspects of Rationality and Knowledge XI*, vol. XI, 2007.

[12] HENKIN, LEON, J. DONALD MONK, and ALFRED TARSKI, *Cylindrical Algebras. Part I*, North-Holland, 1985.

[13] HILL, BRIAN, *Jouer avec le Faux. Recherches sur les processus mentaux à l'œuvre dans la lecture des textes de fiction*, Doctorate thesis, University Paris 1 Panthéon-Sorbonne, 2006.

[14] HILL, BRIAN, 'Living without state-independence of utilities', Tech. rep., GREGHEC, 2007. 874/2007.

[15] HILL, BRIAN, 'The logic of awareness change', in *Proceedings of ILCLI International Worskop on Logic and Philosophy of Knowledge, Communication and Action*, University of the Basque Country Press, 2007.

[16] HILL, BRIAN, 'Logicality: from a local point of view', *Yeditepe'de Felsefe Yearbook*, 6 (2007).

[17] HILL, BRIAN, 'Towards a "sophisticated" model of belief dynamics. Part II: Belief revision', *Studia Logica*, 89 (2008).

[18] HINTIKKA, JAAKKO, 'Impossible possible worlds vindicated', *Journal of Philosophical Logic*, 4 (1975), 475–484.

[19] KOPPELBERG, S., 'General theory of boolean algebras', in J. D. Monk, and R. Bonnet, (eds.), *Handbook of Boolean Algebras*, vol. 1, North Holland, 1989.

[20] ROTT, HANS, 'A counterexample to six fundamental principles of belief formation', *Synthese*, 139 (2004), 225–240.

[21] ROTT, HANS, 'A counterexample to six fundamental principles of belief formation', *Synthese*, 139 (2004), 225–240. Retrieved February 20, 2006, from http://www.uni-regensburg.de/Fakultaeten/phil_Fak_I/Philosophie/theo_neu/RottV/Index_HRott.htm.

[22] ROTT, HANS, 'Shifting priorities: Simple representations for twenty-seven iterated theory change operators', in *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, vol. 53, Uppsala Philosophical Studies, 2006, pp. 359–384.

[23] SAVAGE, LEONARD, *The Foundations of Statistics*, Dover, New York, 1954. 2nd edn 1971.

[24] STALNAKER, ROBERT C., *Inquiry*, MIT Press, Cambridge, MA, 1984.

[25] VAN BENTHEM, JOHAN, 'Dynamic logic for belief change', *Journal of Applied Non-classical Logics*, 17 (2007), 2007.

[26] VAN DITMARSCH, HANS P, 'Knowledge games', *Bulletin of Economic Research*, 53 (2001), 249–273.

BRIAN HILL
GREGHEC and IHPST
HEC Paris
1 rue de la Libération
78351 Jouy-en-Josas, France
hill@hec.fr