PARIS-
AMSTERDAM
LOGIC MEETINGS
OF YOUNG
RESEARCHERS

Towards a ‘sophisticated’ model of belief dynamics

Brian Hill
(IHPST)

It has been often noted that classical models of belief do not accurately represent the limits of human doxastic capacity, at least if belief is understood in a sufficiently *internal* sense. This is equally true of models of belief revision (Hansson, 2003). Furthermore, Rott (2004) has recently raised doubts about several of the most important of Gärdenfors postulates relating to belief revision; he called for a “more sophisticated model of belief formation”. This paper is attempt to tackle some of these challenges. It proposes a model of belief states which (1) is faithful to a certain *finiteness* of human doxastic capabilities, whilst at the same time (2) permitting a general model for *iterated* belief revision which (2a) satisfies the basic G¨ardenfors postulates for belief revision and (2b) reproduces several of the suggested notions of iterated belief revision as ‘special cases’. Finally (3), it provides a natural analysis of Rott’s supposed counterexample, in terms of a ‘framing effect’ which is easily modeled in the proposed framework.

The fundamental observation behind the proposal is the importance of the notion of a sentence or an issue which is ‘in play’ at a particular moment. This observation seems pertinent not only to certain cases of failure of logical omniscience, but equally to cases of ‘overlooked’ beliefs. For example, if the agent forgets to go to his meeting at 10.00, 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 ‘enter into his mind’*, or it *doesn’t enter ‘into play’*.

A simple way of accommodating the notion of being (or not) ‘in play’ is to explicitly specify the sentences involved at a given moment, and permit this set of ‘pertinent’ sentences to change with time. A ‘*local* ’ language, or at least a ‘*local fragment*’ of language, with its own ‘local’ logical structure (notion of logical equivalence between its sentences, and so on), will be ‘operational’ or ‘relevant’ at a given moment; the only beliefs which are explicit or active at that moment are those towards the sentences of this local language.

However, we will not follow the proposals of Fagin and Halpern (1988), which effectively consist in taking the ordinary possible worlds belief framework and adding appropriate restrictions on the belief operator relating to a set of sentences, since, for one thing, this suggestion proves a little clumsy and inoperable when it comes to studying belief revision. Rather, we replace the whole possible worlds structure by a set of “small possible worlds”, if you like, where only the sentences of the ‘local language’ are assigned truth values. Technically (in the simple propositional case considered here) a model of the *local* logical structure at a given moment will comprise of a local language consisting of Boolean combinations of a set of atomic sentences, and an interpretation of this language, that is, an appropriate function from the language into a set of ‘states’ or ‘*small* worlds’ (these worlds are ‘*small* ’ in the sense that *only* sentences of the local language receive an interpretation in them).

The belief state at a given moment is modeled by a transitive, connected, finitarily stoppered order on the states of the logical structure at this moment (the sentences which are believed are those which are true in all states which are minimal with respect to this order). Such an order provides a semantics for belief revision which validates the Gärdenfors postulates (Gärdenfors, 1988; Grove, 1988). Hence the Gärdenfors postulates are satisfied *with respect to the local logical structure*: for example, the only logical consequences whose belief is implied by the belief in a sentence *A* are the consequences of *A* in the local logical structure, and the only revisions of belief which are accounted for are revisions by sentences in the local language. Note furthermore that the sort of structure described is algebraic: the set of sentences of the local language and the set of sets of states each form a Boolean algebra, and the interpretation function is a quotient homomorphism between the two algebras. We shall call this sort of algebraic structure equipped with the appropriate order *ordered algebra*.

The question of belief dynamics is how new information – and possibly sentences not figuring in the previous local language – come into play. New information shall be modelled by an ordered algebra: the information 'learnt' consists of the sentences true in all the states which are minimal according to the order; the order on the rest of the states represents anticipated revisions of the new information in the light of possible subsequent information. This permits the representation not only of the new information, but equally of the details of how it was learnt, and of certain conditions under which it may be weakened or overturned. The 'standard' case, where no account is given of possible revision of the new information, is a special case of this type of model.

Once one represents the new information as an ordered algebra, belief revision by this new information corresponds to a sort of 'fusion' of two ordered algebra. To model this 'fusion', one may call upon familiar algebraic operations (especially given that the structures involved are algebraic), namely product and quotient operations on algebras and order relations. We shall define an operation which provides a *general model of iterated belief revision*. That is, firstly, it satisfies the G¨ardenfors postulates for 'one shot' belief revision. Secondly, since the result of a belief revision has the form of a *belief state* (it is an ordered algebra), which can undergo further revisions, this is a model of *iterated* belief revision. Thirdly, other proposed models of iterated belied revision, such as those of Segerberg (1998) and Konieczny and Pérez (2000) are reproduced as special cases, corresponding to restrictions on the ordered algebra representing the new information.

Finally, using this formalism, an analysis shall be given of Rott's (2004) supposed counterexample to several of the G¨ardenfors postulates, identifying the apparent problem with a "framing effect" accounted for by an appropriate modeling of the new information with which the belief is to be updated. Under this analysis, Rott's example does not invalidate the fact that the Gärdenfors postulates are satisfied by this model, but only underlines the fact that they are satisfied only in a particular sense, or in rather special cases. Ordered algebras are thus a good candidate for the more 'sophisticated' or 'realistic' models of belief revision which have recently been called for.

References

Fagin, R. and Halpern, J. Y. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76.

Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170.

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

Hansson, S. O. (2003). Ten philosophical problems in belief revision. *Journal of Logic and Computution*, 13:37–49.

Konieczny, S. and P´erez, R. P. (2000). A framework for iterated revision. *Journal of Applied Non-Classical Logics*, 10(3-4):339–367.

Rott, H. (2004). A counterexample to six fundamental principles of belief formation. *Synthese*, 139:225–240.

Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39:287–306.