

**Belief revision** edited by Peter Gärdenfors, Cambridge University Press, 1992, pp 277, £22.95, ISBN 0-521-41260-9.

Suppose that we have a small knowledge-base containing information about animals:

$\alpha$ : All mammals with pointed canine teeth and sharp molars are carnivores.

$\beta$  All carnivores are meat-eaters.

$\gamma$ : Skeletons of Giant Pandas have large pointed canine teeth and sharp molars.

Without having seen a live Panda, we might accept the inference from this database that Pandas are carnivores, and hence meat-eaters. Suppose now we find ourselves in the enviable position of being able to observe live Pandas in the wild in the mountains of China, and make the following observation:

$\epsilon$ : Giant Pandas are not meat-eaters.

Were we merely to add this information to our earlier database, the logical theory as a whole would become inconsistent. So how might we reconcile this new information with what we previously knew in such a way that the knowledge-base does remain consistent?

We might, of course, take the database as sacred and refuse to update it with any information which would make it inconsistent; essentially the approach taken with the use of integrity constraints in the deductive database community. That would mean us throwing away our observation, in this case. That may be acceptable if the knowledge-base represents a well established theory. But it so happens that it does not in this case, and the new observation requires that the previous theory about animals be revised in some way to incorporate the new finding and remain consistent.

One possible strategy to the revision of the database might be to “throw away” one or more of the facts which led to the derivation of the contrary to the new data. In this case, we could choose any one from  $\alpha$ ,  $\beta$  or  $\gamma$ . However,  $\gamma$  is a confirmed fact, and we would be inclined to give it a similar “epistemic” status to the new data  $\epsilon$  which we are considering. So it is hard to see why we should reject  $\gamma$  rather than  $\epsilon$ . Placental mammals with pointed canine teeth and sharp molars are by definition carnivores. So we would not really want to reject  $\alpha$  either. That leaves  $\beta$ , which is really an extrapolation from a series of observations of known carnivores, and so perhaps has a lower epistemic status than any of the other knowledge.

So, we intend to throw away the fact that all carnivores are meat-eaters. But any database has logical consequences, as well as facts. In particular,  $\beta$  has consequences: some carnivores are meat-eaters; or, all carnivores except Giant Pandas are meat-eaters. Do we want to throw away those too? They seem to be quite useful conclusions, even in the light of  $\epsilon$ . So perhaps we want to make the revision in such a way that it incurs the minimal possible change to the original knowledge-base *and* all its consequences? That seems much more scientific. And how to do this in a coherent way is, in a nutshell, the central topic of *Belief Revision*.

This book is a collection of up to date research papers on the subject, and is edited by one of the foremost researchers in this field. Peter Gärdenfors has provided an introductory chapter which covers the basic concepts associated with belief states and belief revision, before moving on to an overview of the alternative models of belief revision which are described in the papers which constitute the bulk of the book. Consequently, this makes for a self-contained volume which is as effective as a text book as it is as a research monograph; although admittedly it is only really suitable for advanced students.

There is, of course, a long tradition of work on belief revision in Artificial Intelligence which stems from the work of Jon Doyle and Truth Maintenance Systems (TMS). The preference now is towards using the more appropriate term “Reason Maintenance System” rather than TMS, and they epitomise one approach to the problem of belief revision. This is the *foundational* approach, which holds that one should keep track of the *justifications* of one’s beliefs. The other approach, that taken by Gärdenfors and his co-workers for example (AGM), is the *coherence* approach. Here

the “focus is instead on the *logical* structure of the beliefs—what matters is how a belief coheres with the other beliefs that are accepted in the present state” (p. 8). Both approaches are given a full airing here, starting with a discussion of their relative merits by Jon Doyle himself.

With respect to the foundational approach, further distinctions may be made. One is to differentiate between *revision*, in which new information is obtained about a *static* world and *updating*, in which new information is obtained about a *changing* world. A paper by Katsuno and Mendelzon argues that there are important differences between the approach which one should take to updating as opposed to revision. Their argument that revision in the AGM sense is inadequate as a model of rational belief change caused by the second kind of information, is continued in a paper on “Planning from first principles” by Michael Morreau.

A second distinction is between those who consider revision should be applied to belief *sets*, and those who consider belief *bases* as the focal point. The AGM approach considers belief revision from the perspective of the axioms explicitly represented in a knowledge-base *and* the logical consequences thereof; a belief set. Papers by Hansson and Nebel in this volume consider that the postulates for revision operations should focus on the *bases* for belief sets (the explicit axioms) rather than on the sets themselves. This model is based “on the intuition that some of our beliefs have no independent standing but arise only as inferences from our more basic beliefs” (p. 89).

Making connections between apparently disparate branches of a discipline can often provide deep insights. This is exemplified here with a paper by Dubois and Prade in which possibility theory is linked to the epistemic entrenchment orderings of AGM. This enables them to explore the links between rational updating and the handling of inconsistencies in their possibilistic logic. They also demonstrate how their model of a belief state can be used to describe updating with uncertain pieces of information.

Further papers cover belief revision and communication (Julia Galliers), conditionals and knowledge-base update (Cross and Richmond) and a detailed analysis of the logic of theory change (Hans Rott). In all, the book provides a comprehensive and thoroughly up to date survey of the state of the art in formal models of belief revision. It is well and thoughtfully produced, and can be highly recommended both for those wishing to gain an initial understanding of the field and for those already familiar with it who wish to broaden their outlook.

Reviewed by Paul J Krause, Advanced Computation Laboratory, Imperial Cancer Research Fund, London, UK

### **Textbooks for artificial intelligence**

**Essentials of artificial intelligence** by Matt Ginsburg, Morgan Kaufmann, 1993, £26.50, pp 430, ISBN 1-55860-221-6

**Formal concepts in artificial intelligence—fundamentals** by Rajjan Shinghal, Chapman & Hall, 1992, £19.95, pp 666, ISBN 0-412-40790-6

**Artificial intelligence—structures and strategies for complex problem solving** by George F. Luger and William A. Stubblefield, Benjamin/Cummings, 1993, pp 740, ISBN 0-8053-4780-1

**Knowledge engineering, Volume 1, Fundamentals** edited by Hojjat Adeli, McGraw Hill, 1990, £21.95, pp 345, ISBN 0-07-000355-6

**Knowledge engineering, Volume 2, Applications** edited by Hojjat Adeli, McGraw Hill, 1990, £31.95, pp 352, ISBN 0-07-000355-7

To assess any book that professes to be aimed at students of artificial intelligence one must first consider what is the best way to go about teaching the subject. This is no easy matter since it requires some opinion about the best way to go about “doing” artificial intelligence—by which I mean the best way of progressing towards the aims of the discipline. The question as to what constitutes these aims is somewhat controversial but is much less controversial than that of the best way in which one should pursue these aims. On the latter question the world of artificial intelligence research is divided into a number of camps. The first division seems to be between those who