

ECONOMICS AND ECONOMY IN THE THEORY OF BELIEF REVISION

Hans Rott, Regensburg

1. Introduction

Plato laboured hard to characterize the difference between genuine knowledge and mere belief. He may be read as having claimed the knowledge is justified true belief (*Meno* 97e–98a, *Theaetetus* 201c–202d). This account became the standard analysis of knowledge for more than two millennia. Since knowledge entails belief on the standard account, any attempt to fathom out the foundations of knowledge must include an analysis of belief. After Edmund Gettier (1963) shook the standard account, many philosophers argued that the standard tripartite definition needs to be supplemented by a fourth condition. One particular prominent avenue was to take the stability or indefeasibility of a belief under criticism as a necessary condition for it to qualify as knowledge – an idea that can also be traced back to Plato’s *Meno*. This suggests that it is not only the set of current beliefs but also the potential development of the agent’s beliefs that is relevant to knowledge.¹

Aristotle opened his *Metaphysics* with the statement that all men by nature desire to know (*Met.* 980a). How can this objective be pursued? If knowledge is (or implies) justified true belief then the agent has to check for justification, truth and belief. Ideally, from a first-person perspective, everything one believes seems justified, so this criterion is not helpful for the agent. Truth, on the other hand, is not transparent to agent (it can only be judged from a third-person perspective). Again ideally, belief is transparent to the agent. As a matter of internal control, there is one thing that an agent can do in the pursuit of truth. Since knowledge entails truth and since contradictions cannot be true, the agent has to eliminate contradictions from his beliefs in order to avoid falsehoods.²

¹There are philosophers, however, who deny that knowledge entails belief. In another famous little paper of the 1960s, Colin Radford (1966) presented a case of an agent who reliably answers questions without being conscious of the truth of his answers. This, Radford argued, is a case of knowledge without belief.

²The maintenance of consistency may also be regarded as a problem not for belief as consciously experienced or expressed by the agent, but for belief as *ascribed* to the agent by a third person. Such a view can indeed take inspiration from Aristotle: “For it is impossible for any one to believe the same thing to be and not to be, as some think Heraclitus says. For what a man says, he does not necessarily believe; and if it is impossible that contrary attributes should belong at the same time to the same subject . . . , and if an opinion which contradicts another is contrary to it, obviously it is impossible for the same man at the same time to believe the same thing to be and not to be; for if a man were mistaken on this point he would have contrary opinions at the same time.” (*Met.* 1005b, transl. W.D. Ross)

The 1980s saw the development of a variety of logical models that explicitly addressed the problem of maintaining consistency in beliefs. The prototypical problem dealt with is the one when an agent meets with new information that contradicts what he has believed to be true, as, e.g., when a scientific theory is falsified by an experiment. It was often proclaimed that the particular way of choosing among consistent successor belief sets was guided by an economy principle. I am referring to *belief revision theory*, which is also known as the *theory of theory change*.³ The theory has been limited to some extent, because it stipulates that contradictions are always resolved in such a way that the new piece of information is accepted. However, this restriction seems to me of minor importance, one that does not invalidate the discussion below and that has recently been rectified anyway.⁴

What is the role played by logic (deduction and induction) in the acquisition of knowledge? Deductive logic serves as a standard against which to measure whether potential belief sets are free of contradictions or not. Inductive reasoning always involves decisions as to which sentences to adopt. *Logic in a narrow sense* refers to formal models of deductive reasoning, axiomatic systems which should preferably come together with some semantic underpinning. *Logic in a wider sense* is the theory of good reasoning, providing us not only with the means to check candidate belief sets for consistency but also with a methodology for deciding which of the candidate belief sets to adopt. The choice of a belief set is based on *extralogical* considerations, but only if we refer to the narrow conception of logic. I will in this paper proceed on the assumption that logic does include rules that are relevant for processes of belief formation and transformation (a kind of inductive reasoning, see Spohn 2003). We shall see that we can clearly distinguish an idea of *economical* behaviour (behaviour dictated by considerations of *economy*) from *economic* behaviour (behaviour as recommended by *economics*). We will ask to what extent the two ideas have as a matter of historical fact become embodied in formal models for belief formation and transformation, and we will address the question to what extent they *should* be respected in these models.

In this paper, I want to have a look at the role that the science of economics may play in logic broadly conceived. In order to do this, I make extensive use of material that is discussed in more technical detail in other publications of mine (Rott 2000, 2001, 2003a–c). While the present paper may afford a convenient survey of previous work, it may unfortunately not be easily accessible to people without prior knowledge of the belief revision literature. Such readers are advised to check with the literature to which I refer. The first thing to do for us now, however, is to get an idea of what economics and economy are all about.

³The landmark paper is Alchourrón, Gärdenfors and Makinson (1985), for book-length treatments see Gärdenfors (1988), Hansson (1999) and Rott (2001).

⁴See the papers on “non-prioritised belief revision” collected in Hansson (1997).

2. What is economics?

One way of finding out what a term means is to look at the science that is supposed to study it. For the word ‘economic’, this seems to be an easy task. We just have to look at what economics is about. Dictionaries define ‘economics’ as the scientific study of the production, distribution and consumption of goods, services and wealth, or more concretely, the study of the system of trade, industry, money etc. But we want to dig deeper. According to Francis Y. Edgeworth (1881, p. 16), “[t]he first principle of Economics is that every agent is actuated only by self-interest.” This restriction to a completely selfish attitude has long been removed. The formal part of Edgeworth’s idea, however, remains valid. In the words of Herbert Simon: “The rational man of economics is a maximiser, who will settle for nothing less than the best.”⁵ This still seems to be the dominant view, as is shown by a few more recent statements. Hausman (1998, Sections 1 and 2) gives the following summary of the dominant school in the 20th century:

The main ‘orthodox’, ‘neoclassical’, or ‘neo-Walrasian’ school models economic outcomes as equilibria in which individuals have done as well for themselves as they could given their preferences and the constraints on their choices. . . . Agents are rational in the sense that their choices are determined by their preferences, which are complete and transitive. . . . contemporary theoretical economics is largely a theory of rational choice. This may seem surprising, since economics is supposed to be an explanatory and predictive science of the actual interactions among people rather than a normative discipline studying how people ought rationally to choose, but it is indeed a fact.

In another handbook article, Rosenberg (1995) tells us about the “assumptions of the ‘economic man’: that all agents have complete and transitive cardinal or ordinal utility rankings or preference orders and that they always choose that available option which maximises their utility or preferences”. Earlier the same author characterised economics as guided (or misguided) by an “extremal intentional research programme” (Rosenberg 1983).⁶ And what is being maximised is utility. According to Broome (1999, p. 21–22): “‘utility’ acquired the meaning: *the value of a function that represents a person’s preferences*. . . . The first principle of economics is . . . utility theory, . . . modern, axiomatic utility theory . . .”

What is common to these very abstract formulations of the basic tenets of modern economics is that economic agents are viewed as having definite preferences, and that when choosing actions or commodities, they aim at satisfying their

⁵The quotation is taken from McFadden (1999, p. 73). Simon himself famously advocated an alternative model of rationality, viz., that of *satisficing* instead of maximising. In this model, economic agents set out for achieving a certain level of aspiration, without investing any extra efforts to find out whether they could obtain something even more valuable.

⁶According to Rosenberg (1983), economics is not an empirical science, but a branch of applied mathematics – but Rosenberg does not say applied to *what*. My picture below is more in line with the view that economics is a normative discipline, a theory of rationality.

preferences as well as the circumstances allow. It is important to note that not just any kind of preference is considered to be appropriate. Preferences have to be transitive and complete (technically speaking, they must be *pre-ordering* or *weak orderings*), in order to be representable by a suitably chosen utility function. Ties in preferences are permitted, of course, but incomparabilities are ruled out.⁷ With a little exaggeration, one can say that economics is based on (or: is an elaboration of) the theory of rational choice. Here, *choice* is called *rational* or *coherent* if and only if it is representable by a preference relation, and a *preference relation* in turn is considered to be *rational* if and only if it is representable by a *utility function*.⁸

Economics, then, is about *rational* rather than the *actual* behaviour of individuals.⁹ The laws of economics should not be expected to be empirically adequate, they are valid only as idealisations, or as norms. Economics is based on a formalised variant of common-place folk psychology with individual-level explanation of free agents. It is essential to the research program of economics proper that processes pertaining to whole societies or economies are to be explained by, or reduced to, the behaviour of egocentric, maximising agents.¹⁰ Usually, it is assumed that comparability of preferences or aggregability of utilities across different individuals does not make sense. It is important, however, that the possibility of comparison and aggregation is taken for granted across different “attributes” or “criteria” of goods, as well as across different situations in which the same goods are available (in varying amounts). This is part of ordinal and expected utility theory, and I suspect that this may ultimately be the reason for the economists’ insistence on a person’s preferences being representable by a utility function.

In order to find out whether there is anything economic (referring to economics in this abstract standard sense) about belief revision, we will have to look for patterns of rational choice, i.e., choice that is maximising with respect to some underlying preferences, or more exactly, to some weak ordering which can in

⁷Bernard Walliser (personal communication) has told me that the widespread insistence on transitive and complete preference relations has been due to the fact that for a long time, economists simply did not know how to handle intransitive and/or incomplete relations. Now that they know how to treat them, the insistence has gone. For some state-of-the-art work in this area, see Ok (2002) and Eliaz and Ok (2003).

⁸It takes more than transitivity and completeness for a preference relation to be representable by a utility function. An additional technical continuity property is needed (Debreu 1959, pp. 54–59). A concise presentation of ordinal and expected utility theory is given by Hausman and McPherson (1996, Chapter 3: ‘Rationality’).

⁹This is economics in a narrow sense. Economics in a wider sense has a social component and presumes interpersonal comparability or aggregability of preferences. Blending with social science, political philosophy and ethics, it includes considerations of justice, fairness, solidarity, liberty and equity in the distribution of scarce goods. All this supplements, or rather corrects, the rather restricted focus on maximization in economics in the narrow sense. The present paper does not address any social component of rationality, nor any social phenomena in doxastic matters (like common or distributed belief, information exchange, or multi-agent belief revision).

¹⁰“Egocentric” is not meant to imply “selfish” here. Egocentric agents only look at their personal preference, but the preferences themselves may embody all kinds of (possibly altruistic) thoughts and feelings.

turn be represented by a utility function.

3. Acting economically, a second view: “Informational economy”

There is an alternative and perhaps more intuitive concept of “acting economically” that has played an important role in the development of the research on belief revision. Belief revision is often thought to be *economical* behaviour rather than *economic* behaviour. In the English language, there is a division of labour between the adjectives “economic” and “economical”. The former is closely tied to the noun “economics” and means either “of or referring to economics”, and has “profitable”, “remunerative” and “gainful” as potential synonyms. The adjective “economical”, on the other hand, is closely tied to the noun “economy” and means essentially the same as “thrifty”, “frugal” or “not wasteful”. This particular difference of meanings turns out to be useful for our discussion.

The perspective of *economy* (rather than *economics*) was forcefully taken in Peter Gärdenfors’s influential book *Knowledge in Flux* (1988). A glance at the index of the book makes it immediately clear that the *criterion of informational economy* is employed to motivate the essential parts of the formal modellings of Gärdenfors and his collaborators Carlos Alchourrón and David Makinson. Gärdenfors refers to this criterion for the motivation of belief expansions (1988, p. 49), belief revisions (pp. 53, 58) and belief contractions (p. 61).¹¹ Basically, the criterion is taken to be identical with the idea of *minimal change* (p. 53) and the *conservativity principle* (p. 67). According to Gärdenfors,

The key idea is that, when we change our beliefs, we want to retain as much as possible of our old beliefs – information is in general not gratuitous, and unnecessary losses of information are therefore to be avoided. (Gärdenfors 1988, p. 49, similarly on pp. 16, 157)

Ever since the appearance of Gärdenfors’s book, the criterion of informational economy has been taken to be a “hallmark” of the research paradigm created by Alchourrón, Gärdenfors and Makinson (henceforth, AGM).¹²

There is, however, reason for asking why exactly a rational person should be conservative. Gärdenfors’s argument that information is not gratuitous does not seem to be sufficient, because information, even if costly, may be wrong, and even if it is correct, it may be misleading. Spelt out a little more explicitly, the argument for conservatism seems to be this:

You shouldn’t give away what is valuable
 What you have is valuable
 —————
 You shouldn’t give away what you have

¹¹Throughout this paper, it is presupposed that *revisions* have to be *successful* in the sense that they efficiently incorporate the specified new piece of information into the current belief set. *Contractions* are called *successful* if they efficiently remove some specified sentence from the belief set (unless that sentence happens to be a logical truth).

¹²Compare, for instance, Boutilier (1996, pp. 264–265) and Darwiche and Pearl (1997, p. 2).

The first premise is hardly controversial, it might even be called analytically true as a prescription of economic rationality.¹³ The second premise, on the other hand, is much harder to justify. Its truth (or at least our feeling that it is true) may have evolutionary reasons – probably the human species would long have been extinct if too many of the sentences we accept as true were wrong. Therefore, it is at least likely that many of the sentences that we happen to hold true have some survival value.¹⁴

But the argument for conservatism is a far cry from waterproof. Problems are not hard to come by. Isn't it all too obvious that what we have is not always best? So why should we care to preserve it? It is true that we lose *information* or *content* when we give up some sentences of our belief set, but it is not clear whether we lose some *truths*, and no-one should object to losing falsities. Notice that there is a basic tension here between the economical and the economic precept for belief dynamics: The former tends to recommend leaving everything as it is, while the latter recommends striving for the best.

4. Economic and economical considerations in belief revision theory

In asking what is economic about belief change, we have to keep in mind two different aspects. Besides the choice-preference-utility line of thinking that we sketched in Section 2, we have found a second type of idea in the thrifty clinging to the sentences one has accepted. (Notice that thrift in itself does not play a prominent role in the science of economics.)

In the work of AGM, it is comparatively easy to recognise the criterion of informational economy at work in expansions of belief sets by sentences that do not contradict the prior belief set. In such cases, AGM recommend simply to add the new sentence to the prior beliefs and take the deductive closure of everything taken together.¹⁵ However, as we shall see, there are no traces of this criterion for the belief-contravening case which, after all, is the case for which logical models of belief change have primarily been devised.

But Gärdenfors's argument for conservatism can be generalised. In my view, it is one of the most important philosophical insights of belief revision theory in the 1990s that *belief states* cannot be represented properly by *belief sets* only. Something else has to be added, namely, some structure that encodes how the agent is to revise his belief set in response to surprising information. Typically there is a mechanism exploiting some sort of selection function or preference

¹³However, it abstracts from the case that one may be *forced* to give away what is valuable, or that one may *invest* something at a given time in order to make profit later on.

¹⁴I am ready to grant that this train of thought is not very compelling. Alternatively, the second premise might be replaced by another one which, however, is at least as dubious: "It is always better to have something than to have nothing."

¹⁵In the following, a *belief set* is meant to be a set of sentences that contains its own logical consequences. Thus we endorse the idealisation or stipulation that the beliefs of an agent be deductively closed.

relation.¹⁶ Let us use a neutral name and call the structure exploited by the mechanism a *belief-revision guiding structure*.¹⁷ It turns out that if the new information is inconsistent with the presently accepted belief set, it is impossible to apply the pure idea of informational economy on the level of sentences. However, the idea can be applied on the level of revision-guiding structures. This at the same time defines a form of conservatism that is usable for iterated belief change.

Fig. 1 gives a representation of the various senses that “economic” and “economical” can take in belief revision contexts. The branch at the left indicates the

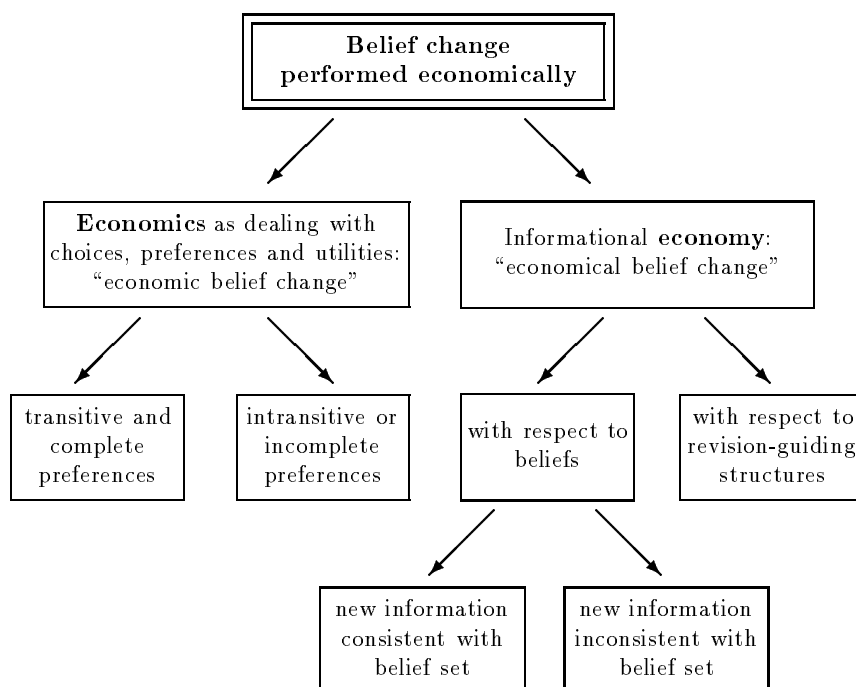


Figure 1: Belief change performed economically

idea that economic agents may be conceived of as rational or coherent choosers. Though this has not been the principal motivation of belief revision models, we will later see that exactly this idea casts a long shadow in the realm of belief formation.¹⁸ On the right-hand branch, we represent the idea of informational economy that has always been advertised as the prime driving force of belief

¹⁶Sometimes, as in approaches working with belief *bases*, use is made of a partitioning or lumping together of the informational contents of the beliefs .

¹⁷Belief-revision guiding structures typically encode *more* information than the set of current beliefs. Since such structures in general allow to retrieve the set of current beliefs, there is no need for a second component specifying the belief set. Therefore, it is possible to formally identify the agent’s *belief state* with a belief-revision guiding structure.

¹⁸As pointed out by Olsson (2003), there is a link with the theory of choice in the early history of belief revision theory in Gärdenfors’s (1978/79) article on Ramsey test conditionals. Here Gärdenfors tried to simulate Lewis’s (1973) logic for counterfactuals that is based on a

revision theory. In the next four sections, we will trace a few important distinctions within the principle of informational economy, also known as principle of minimal change or conservatism. The principle of informational economy *with respect to sentences* tells us: “Don’t give away your beliefs beyond necessity!” The principle of informational economy *with respect to revision-guiding structures* tells us: “Don’t change your doxastic preferences beyond necessity!” These two maxims can further be distinguished as to whether they apply to the belief-contravening case (the one for which belief revision models have primarily been invented) or only to the case where the new information is consistent with the agent’s belief set.

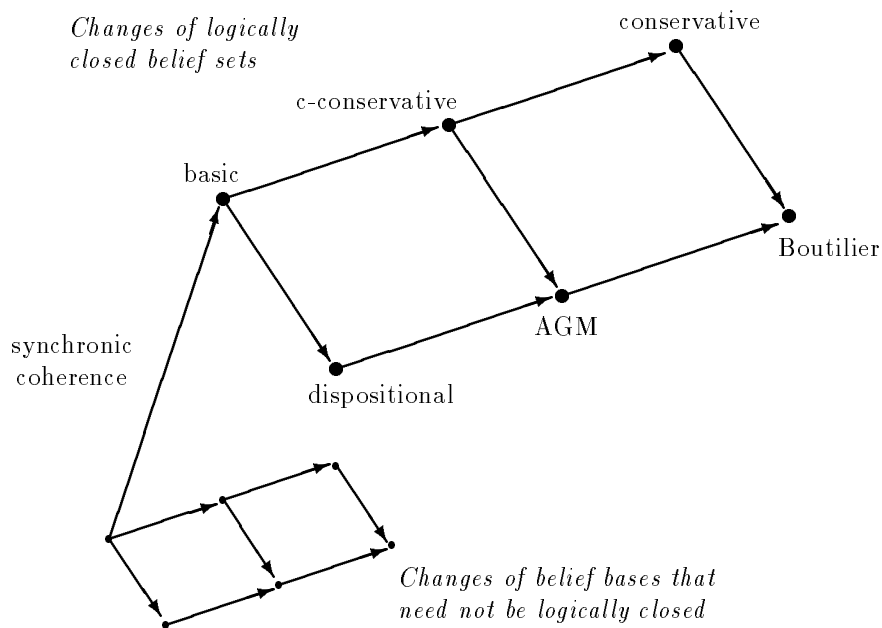


Figure 2: Three dimensions of coherence

Fig. 2 gives a different representation of the situation in terms of various dimensions of coherence. At the bottom there are six dots standing for options of belief base revision, where *belief bases* are sets of sentences that need not obey the static coherence constraint of logical closure. We will not be concerned with this approach in this paper,¹⁹ but instead focus on methods of changing belief sets that are presumed to be logically closed – i.e., on the upper six dots. “Basic changes” represent options that neither recognise the economical constraints

possible worlds semantics with choice functions. Philosophically, however, Gärdenfors set out to *avoid* possible worlds semantics and *replace* it by a belief revision semantics for conditionals (for some serious formal problems of this undertaking, see Gärdenfors 1986, Fuhrmann 1993 and Arló-Costa and Levi 1996). In hindsight, it is somewhat ironic that 10 years after Gärdenfors’s early article, Grove (1988) showed how closely AGM’s belief change model in terms of partial meet operations relates to the systems-of-spheres modelling of Lewis after all.

¹⁹For a thorough treatment of this topic, see Hansson (1999).

of minimal change nor the economic constraints concerning the rationality of choices. Moving from left to right in the lattice of the six dots adds economical constraints of minimal change, moving downwards adds economic constraints on the rationality of choices. The label “c-conservative” denotes conservativity in the case where the input is consistent with the current belief set, without a similar commitment for the belief-contravening case.

Before entering the systematic discussion, let us add a few historical remarks. As the talk of “dimensions” meant to suggest, it is possible to add to the basic form of belief revision elements of conservatism and elements of rational choice independently from one another. In their seminal work of the 1980s, AGM explored pure conservatism with respect to sentences (so-called *maxichoice* contraction and revision functions) and completely unconstrained choices (the case where only AGM’s six *basic* rationality postulates hold).²⁰ However, they rejected the idea of maxichoice functions as intuitively inadequate, and I think it is fair to say that the main elegance and force of their theories derives precisely from the *supplementary* postulates that go *beyond the basic case*. So what makes the work of AGM distinctive is, on the one hand, a rather strong concept of rational choice generated by transitive and connected preferences, and on the other hand a rather weak concept of conservatism. They provided for conservatism with respect to beliefs in the consistent case (where revision reduces to set-theoretic addition plus logical closure), but they provided neither for conservatism in the belief-contravening case nor for conservatism with respect to revision-guiding structures (they provided no change mechanisms for revision-guiding structures at all). So contrary to wide-spread folklore, AGM paid a lot more respect to ideas found in economics than to the idea of informational economy.

Systematic variations of the strength of the relevant ideas were investigated only in the 1990s. On the one hand, *weakenings* of AGM’s strong presuppositions with respect to the rationalisability of choices are suggested by Lindström (1991) and Rott (1993, 2001).²¹ On the other hand, *strengthenings* of the very weak concept of conservatism in AGM are investigated by Boutilier (1993, 1996) and Rott (2003a). Interestingly, it has turned out that the most efficient way – and perhaps the only way – of implementing conservatism with respect to beliefs in the belief-contravening case is at the same time a form of conservatism with respect to revision-guiding structures. This variant of conservative belief revision has actually first been discussed as a particular strategy of extending the classical AGM model in order to equip it with means for performing iterated belief revisions. While Boutilier did that in a context that presupposes the full strength of AGM’s requirements for rational choice, Rott lifts this restriction and shows that the conservative method can be brought to bear without any assumptions about the coherence of the choices involved.²²

²⁰See Alchourrón, Gärdenfors and Makinson (1985).

²¹For similar projects in the related field of non-monotonic reasoning, see Schlechta (1996) and Lehmann (2001).

²²But see footnote 34 below.

This little sketch of some developments in the belief change literature shows that the concept of *economical* belief change is independent of the concept of *economic* belief change. In the sense specified, AGM belief revision (minus two axioms for revisions by inputs that are consistent with the prior belief set²³) is not at all economical but, thanks to AGM’s “supplementary” postulates, fully economic; the conservative approach investigated in Rott (2003a) on the other hand is very economical, but not at all economic.

In the following sections, we recapitulate relevant parts of the actual development of belief revision theory, and then discuss the merits and problems of changing beliefs economically. I will first detail my reasons for claiming that conservatism with respect to *beliefs* has hardly been followed in classical AGM-style belief revision models, and that it is doubtful indeed whether it would be a good idea to follow it. Conservatism with respect to belief-revision guiding structures has been studied as a particular strategy for iterable belief change, but it has turned out that it should not be followed either. On the economics side, I briefly survey the (severe) constraints on rational choices have been endorsed in classical AGM-style belief revision and its iterable extensions. Although the AGM postulates can be liberalised systematically according to one’s wishes, I will finally argue that even modest economic postulates for belief change are problematic – just as problematic as even the fundamental constraints on rational choices are. We shall find that doxastic preferences appear to be context-dependent, a fact that gives rise in particular to a formidable problem sometimes labelled “the informational value of the menu”.

5. Informational economy with respect to beliefs: What has been done?

In discussing informational economy with respect to beliefs, we will keep on making two important idealisations: We stipulate that the set of sentences accepted by an agent be logically consistent and closed. This condition is, of course, wrong as a description of the set of sentences that a real agent would assent to when queried in an interview. However, if we consider a belief set to be the set of sentences that we *ascribe* to an agent from a third person perspective, or the set of sentences that the agent is *committed to*, then the ideas of consistency and closure lose their implausible appearance. Another simplification we make is that when a belief set is revised in response so some new piece of information, the revision process successfully incorporates the new information, so that it is in fact an element of the revised belief set. This is not always sound strategy in realistic belief change situations, but I think we can safely disregard the complications for the discussion to follow. So let us call a consistent and logically closed belief set that includes a new piece of information ϕ a *candidate revision* of a belief set B by ϕ . A (candidate) revision of B by ϕ is called *belief-contravening*, if ϕ is inconsistent with B .

²³In the common numbering of axioms, these are the third and fourth AGM postulates; taken together they state that if $\neg\phi$ is not in B , then $B * \phi$ equals $Cn(B \cup \{\phi\})$.

Let us discuss two attempts at capturing the idea of informational economy on the level of beliefs:

- (1) When accepting a new piece of information, an agent should aim at a minimal change of his old belief set.
- (2) If there are different ways of effecting the belief revision, the agent should give up those sentences that are least entrenched.

These two maxims have frequently been appealed to as the principal motivation for logical models of belief change. However, in their most straightforward readings, they are a caricature of what has really been done in the development of the standard models of belief revision. I have argued for this in Rott (2000), and briefly recall the results presented in that paper.

As regards maxim (1), one can show that no two distinct belief-contravening candidate revisions of a consistent and logically closed belief set by a sentence ϕ can be set-theoretically compared in terms of the sets of sentences on which they differ with the prior belief set.

For the discussion of maxim (2), we need a little bit of terminological preparation. A sentence ϕ is *more entrenched* in a belief set than another sentence ψ if and only if the agent holds on to ϕ and gives up ψ upon learning that (it may be the case that) not both ϕ and ψ are true.²⁴ A new piece of information ϕ is called *moderately surprising* if $\neg\phi$ is a non-minimal element of the prior belief set with respect to epistemic entrenchment. Let us call a revision by ϕ *amnesic* if the revised belief set consists of nothing else but $Cn(\phi)$ (where Cn is some ordinary well-behaved Tarskian logic); otherwise we call the revision *anamnesic*. Now suppose we want to revise a belief set by a sentence ϕ and identify two elements of the initial belief set that non-redundantly entail $\neg\phi$. Then it may well happen, when performing an ordinary AGM-style revision by ϕ , that the agent removes the *more* entrenched and retains the *less* entrenched sentence. An instance of such a situation can indeed *always* be identified when ϕ is moderately surprising and the revision by ϕ is anamnestic. Thus neither of the two maxims that have always belonged to the core rhetoric about AGM-style belief change models is actually obeyed in these very models. This observation may seem too baffling at first sight, but it has turned out to be rather robust. We briefly present four objections and give rejoinders to each of them. For a more extensive treatment, the reader is referred to (Rott 2000).

First, one may hold that it is not belief *revision*, but belief *contraction* that is the right kind of operation to be judged in terms of minimal change. As was mentioned above, AGM first thought of focussing on so-called “maxi-choice contractions”, but these were immediately seen to yield counterintuitive consequences.²⁵ On a more general level, the postulate of “Recovery” says that

²⁴See Gärdenfors and Makinson (1988) and Rott (2001).

²⁵Alchourrón and Makinson (1982) proved a result that may be viewed as a strong argument against maxi-choice revisions as applied to belief sets.

inserting back again a belief that had just been withdrawn should recover the whole of the original theory. Recovery was explicitly introduced as a codification of the idea of minimal change. However, it fulfils this function only partially; its effects are destroyed if the contraction is part of a revision using the so-called Levi identity; and, most importantly, the recovery condition has been forcefully and severely criticized on intuitive grounds by many authors (see Hansson 1999, Section 2.3).

Second, the well-known representation theorems of AGM (and their possible-worlds reinterpretation by Grove) seem to show that “rational” belief contraction and revision operations can be represented as being generated by a minimization process with respect to some underlying preference relation. However, the interpretation of such a preference relation is completely open (it might, for instance, mean remoteness rather than closeness); the minimization process is compromised by overriding principles of preference and indifference;²⁶ and finally, since the AGM postulates do not encode any notion of minimal change in the belief-contravening case, we should not even expect to find substantial traces of this idea in a semantics that can be proved adequate for the AGM axiomatization.

Third, it may be pointed out that we need not aim at the preservation of all of our old beliefs, but only of those that are true. So even if the idea of informational economy is not effective when applied to the whole of the prior belief set, it may still do good work if we restrict our attention to the beliefs that we really treasure, viz., the true beliefs. Unfortunately, however, that move does not help either, since essentially the same results can be reproduced for the conservation of true beliefs as for the conservation of beliefs *tout court*: No two belief-contravening candidate revisions of a consistent belief set that contain different sets of true beliefs can be set-theoretically compared in terms of the *true* beliefs on which they differ with the prior belief set.

A fourth route to saving the idea of minimal change can be taken by applying the idea of informational economy not to belief sets, but to richer representations of doxastic states: to belief-change dispositions, or equivalently, to structures suitable for guiding (iterated) belief revision. This is a point that we will address presently in some detail. Before doing that, however, we have a look at potential norms for belief change.

6. Informational economy with respect to beliefs: What should be done?

First of all, we have to be aware of the fact that the demand for informational economy conflicts with other desiderata. For instance, it competes with the *synchronic* or *static coherence* constraints of *logical consistency* and *closure*. If we find ourselves caught in an inconsistency, we should give up *something* and we typically have to give up not only a single culprit sentence, but also many

²⁶Principles of Preference and Indifference are discussed in Pagnucco and Rott (1999).

sentences that are deductively related with the latter. This is because we want to maintain the logical closure of our belief sets. But of course, a change that is minimal subject to the constraints of consistency and closure will in general be a bigger change than one that is minimal when no constraints are to be respected. Conservativity may itself be viewed as a criterion of *diachronic* or *dynamic coherence* (Rott 1999). There are more concepts of coherence that we will have reason to consider below, viz., *dispositional coherence* and *temporal coherence*. At this juncture it is not at all clear whether the latter two concepts give rise to more requirements that compete with the requirement of informational economy.

Secondly, it is instructive to contrast the ideas that are advocated in the logical literature on belief revision with ideas recommended by philosophers of science who usually think of belief change as arising in the (r)evolution of scientific theories or research programmes. For the latter point of view, it may suffice here to rely on the elementary but thought-provoking little book *The Web of Belief* by Quine and Ullian (1978). In Fig. 3, the ‘virtues’ of hypotheses that these authors mention are contrasted with the criteria advocated in the belief revision literature.

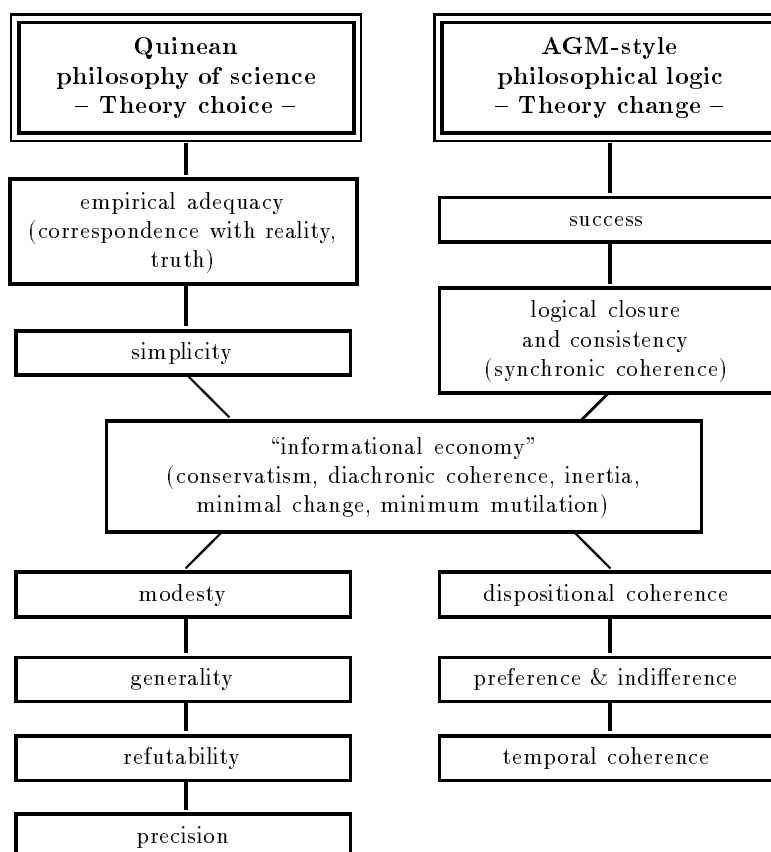


Figure 3: Criteria for theory choice and theory change

These terms will be explained in due course. What we can immediately see from Fig. 3 is that informational economy is the only criterion that is endorsed both by Quine and Ullian and the logical modellings of belief revision. A moment's thought makes it clear that at least some of the virtues listed here compete with one another. Simplicity often requires deviations from informational economy, modesty conflicts with refutability. We don't have to commit ourselves to one or the other set of criteria here, nor do we need to specify some ranking or weighting of the criteria. It is enough at this juncture to note that an account of how to integrate various criteria of *theory choice* (the perspective dominant in the philosophy of science) and *theory change* (the main perspective of philosophical logic) is badly needed, but has never been offered. There is no obvious reason for according informational economy a privileged status among the many contenders that we have identified. It seems fair to say that informational economy can only claim a very restricted normative force.

7. Conservatism with respect to belief-revision guiding structures: What has been done?

We now turn to a second interpretation of the idea that agents should aim at preserving what they have. The propositional content encoded in sentences of the agent's language is not the only kind of information that may be deemed valuable. We might also be interested in preserving the non-propositional information encoded in belief-revision guiding structures, i.e., in richer representations of belief states. Two such representations that have gained some currency in the literature are two kinds of doxastic *preference relations* to which we shall now turn.

First, we consider *plausibility orderings* \preceq of the set W of possible worlds, with the understanding that $u \preceq v$ means that u is *at least as plausible* as v in the belief state represented by \preceq .²⁷ Given such an ordering \preceq , the core set $\min_{\preceq} W = \{u \in W : \text{there is no } v \in W \text{ such that } v \prec u\}$ of the \preceq -minimal worlds contains exactly those worlds that are consistent with the current belief set B , i.e., those that could be the real world given what the agent believes. The ordering of the remaining worlds reflects their relative distance from this core set. Belief revision prompted by a new piece of information ϕ then proceeds by manipulating the ordering of worlds in such a way that all the minimal worlds in the revised ordering \preceq_{ϕ}^* satisfy ϕ . This constraint is mild and leaves a lot of leeway for the exact specification of a coherent revision mechanisms. The most *conservative* or *economical* way of changing the plausibility ordering that respects the constraint was first defined and investigated by Boutilier (1993, 1996):

$$u \preceq_{\phi}^* v \text{ iff } \begin{cases} u \in \min_{\preceq}[\phi] & \text{or} \\ v \notin \min_{\preceq}[\phi] \text{ and } u \preceq v \end{cases}$$

²⁷This reading is perhaps that reverse of what the reader has expected. An explanation for having more plausible theories smaller under \preceq is that they are less distant from the agent's beliefs and expectations.

Here $[\phi]$ denotes the set of all possible worlds satisfying ϕ . Given the basic constraint that the set $\min_{\leq^*} W$ of worlds defining the revised belief set $B * \phi$ should be identical with the set $\min_{\leq} [\phi]$,²⁸ the ordering \leq_{ϕ}^* preserves as much of the ordering \leq as possible.²⁹

Another way of richly representing belief states consists in *entrenchment orderings* \leq of the set L of sentences (phrased in a given language), with the understanding that $\phi \leq \psi$ means that sentence ψ is *at least as entrenched as* sentence ϕ in the belief state represented by \leq (as already indicated in Section 5). Given an entrenchment ordering \leq , the set $\min_{\leq} L = \{\phi \in L : \text{there is no } \psi \in L \text{ such that } \psi < \phi\}$ of sentences that are minimally entrenched contain exactly the the agent's non-beliefs, i.e., the complement of his belief set B . The ordering of the sentences within B reflects how firmly they are endorsed by the agent, the relative tenacity with which the agent is determined to cling to his beliefs. In this model, belief revision prompted by a new piece of information ϕ proceeds by manipulating the ordering of sentences in such a way that ϕ is not minimally entrenched under the revised ordering \leq_{ϕ}^* . A constraint following from basic AGM-theory is that the revised belief set $B * \phi = L - \min_{\leq^*} L$ should be the set $\{\psi \in L : \neg\phi < \phi \rightarrow \psi\}$.³⁰ Again, this constraint is mild and leaves a lot of leeway for the exact specification of the appropriate revision mechanism.

The most *conservative* or *economical* way of changing the entrenchment ordering that respects the constraint is investigated in Rott (2003a):

$$\psi \leq_{\phi}^* \chi \text{ iff } \begin{cases} \psi \notin B * \phi & \text{or} \\ \chi \in B * \phi \text{ and } \psi \leq \chi \end{cases}$$

$$\text{iff } \begin{cases} \phi \rightarrow \psi \leq \neg\phi & \text{or} \\ \neg\phi < \phi \rightarrow \chi \text{ and } \psi \leq \chi \end{cases}$$

Given the above-mentioned constraint, this ordering \leq_{ϕ}^* preserves as much of the prior entrenchment ordering \leq as possible. To see this, we look at the pairs for which \leq_{ϕ}^* reverses the ordering of \leq . When do we have $\psi \leq \chi$ but not $\psi \leq_{\phi}^* \chi$? Inspection of the definition shows that this can only happen if ψ is in, but χ is not in $B * \phi$, and this deviation is well motivated by the fact that non-beliefs can never be as entrenched as beliefs. Conversely, when do we have $\psi \leq_{\phi}^* \chi$ but not $\psi \leq \chi$? Inspecting the definition once more, we find that this can only happen if ψ is not in $B * \phi$, and the same motivation applies: non-beliefs are the least entrenched sentences. We see that there is no unforced

²⁸More formally, the constraint says that the revised belief set $B * \phi$ should be identified with the set of all sentences ψ that are satisfied by every world in $\min_{\leq} [\phi]$.

²⁹This claim is true if \leq is connected (i.e., "fully economical" in our sense) which is what Boutilier presupposes. The claim becomes problematic if the restriction to connected preference relations is lifted. See footnote 34 below.

³⁰See for instance Rott (2003b).

deviation of \leq_ϕ^* from \leq . This is why this recipe defines the most *conservative* or *economical* way of changing an entrenchment ordering (cf. Rott 2002a).³¹

It has turned out (Rott 2003b) that the account based on entrenchment orderings is essentially a generalisation of the possible worlds account proposed by Boutilier. In the specific context of the full comparability assumption made by Boutilier, however, the methods are equivalent. Both methods satisfy an axiom for iterated belief revision which is sufficient to characterise conservative revisions of richer representations of belief states. It is sufficient to deal with the case of two subsequent revisions. Any finite number of further revisions can easily be constructed from this case by induction. As shown by Boutilier (1993, 1996) and, in a more general setting, by Rott (2003a), a repeated conservative change of B first by a sentence ϕ and then by a second sentence ψ leads to the same result as a single conservative revision of B by ψ , if ψ is inconsistent with the result of the revision of B by ϕ . Taken together with basic AGM theory, this amounts to the following recipe for iterated revision:

$$(B * \phi) * \psi = \begin{cases} (B * \phi) + \psi & \text{if } \neg\psi \notin B * \phi \\ B * \psi & \text{if } \neg\psi \in B * \phi \end{cases}$$

Why would we want to call this recipe “conservative”? Because the upper case is just AGM’s c-conservativity generalised to the iterated case. And the lower line suggests that if ψ cannot be accommodated consistently, the way of handling it in the revised belief set $B * \phi$ is just the same as it was in the original belief set B . Loosely speaking, the structure of the old belief set is stronger than the new piece of evidence ϕ , making it seem as if the agent had never learnt about ϕ .³²

8. Conservatism with respect to belief-revision guiding structures: What should be done?

Almost immediately after Boutilier had suggested conservative belief revision as a natural extension of the AGM model, Goldszmidt and Pearl (1994, 1997) discovered that the behaviour defined by this model is queer. They gave the following example. A person who we may for the sake of argument take to be in a state of complete ignorance observes an animal that she takes to be a bird (b). As the animal comes closer, the person perceives that the animal is red (r). A few moments later, she realises (perhaps informed by an ornithologist) that the animal is not a bird after all ($\neg b$). If we use Boutilier’s method of conservative belief revision, the result of these three subsequent revisions is $Cn(\emptyset) * b * r * \neg b = Cn(\neg b)$.

Goldszmidt and Pearl rightly argue that this is counterintuitive. Why “forget” the colour of the animal just because it turns out that it has misclassified as

³¹Notice that this procedure tends to introduce new comparabilities: Any new non-belief is comparable to every other sentence, even if it was an ‘isolated’ belief before.

³²The appearance is deceptive, though, since in general $(\leq_\phi^*)_\psi^* \neq \leq_\psi^*$ and $(\leq_\phi^*)_\psi^* \neq \leq_\psi^*$.

a bird? In more general terms, it can be shown that this model is *temporally incoherent*. The AGM-postulate of “success” for revisions says that the most recent piece of information should always be included in the revised belief set. Thus, at the moment of receipt a piece of information is being treated as the most important one. But this privilege is immediately lost when another, new piece of information happens to come in. To see this, let ϕ , ψ and χ stand for sentences that are pairwise consistent, but jointly inconsistent. Then iterated conservative change of the trivial belief set $Cn(\emptyset)$ first by ϕ , then by ψ and finally by χ results in the belief set

$$Cn(\emptyset) * \phi * \psi * \chi = Cn(\phi, \chi)$$

The first and the last piece of information are stronger than the one that comes in between. Conservatism with respect to revision-guiding structures thus has unacceptable consequences when applied as a method for iterated belief revision. It violates the requirement that a good method of belief revision be *temporally coherent*, i.e., coherent in its attitude towards the value of the recency of information.

Both AGM’s and Boutilier’s models assume that the orderings of worlds or sentences involved are complete pre-orderings, i.e., that all worlds and all sentences are comparable as regards their plausibility or entrenchment, respectively. It might be suspected that the problem of conservative revision is (at least in part) due to the strong requirements of “dispositional coherence” inherent in the AGM model upon which Boutilier’s model is built. However, the unwelcome effects of temporal incoherence remain present in exactly the same way even if *all* of the dispositional requirements of the AGM model are dropped (Rott 2003a).

As long as one decides invariably to accept new information (i.e., to regard the last piece of information as the most important one), the only coherent attitude towards the recency of information is to regard the second-last piece of information as the second-most important one, and so on. Instead of $Cn(\phi, \chi)$ as above, the desired result would thus be

$$Cn(\emptyset) * \phi * \psi * \chi = Cn(\psi, \chi)$$

There is an alternative model of iterated belief revision that yields precisely this result. This less conservative, more *moderate* model has been mentioned and used quite a number of times in the literature, without there being a canonical paper where the model was first endorsed.³³ I take the opportunity to give definitions paralleling those that characterise conservative belief change. Like the latter definitions, the definitions of moderate belief change make good sense also for the case when dispositional coherence is not presupposed.

³³To my knowledge, the model was first studied systematically by Abhaya Nayak (1994, also see Nayak *et al.* 1996, 2003), but many people have had the idea independently of Nayak. See Liberatore (1997), Glaister (1998), Kelly (1999), Konieczny and Pino Perez (2000), Papini (2001), Nayak, Pagnucco and Peppas (2003). I discuss the merits of this model and explain my label “moderate belief revision” for it in Rott (2003a).

Using the representation of belief states by means of orderings between models, the moderate way of changing such states is this:

$$u \preceq_{\phi}^* v \text{ iff } \begin{cases} u \in [\phi] \text{ and } v \notin [\phi] & \text{or} \\ (u \in [\phi] \text{ or } v \notin [\phi]) \text{ and } u \preceq v \end{cases}$$

Since here the whole $[\phi]$ -area (not just the set $\min_{\preceq}[\phi]$) is shifted, this definition is intuitively less conservative than Boutilier's model. However, it can be shown that if Boutilier's method is stripped of dispositional coherence (of its *economical* features), it ceases to be conservative (or *economic*) in a sense that can be given a precise mathematical definition.³⁴

Using the representation of belief states by means of entrenchment orderings, the moderate way of revising one's belief state can be represented thus:

$$\psi \leq_{\phi}^* \chi \text{ iff } \begin{cases} \psi \wedge \chi \notin Cn(\phi) \text{ and } \phi \rightarrow \psi \leq \phi \rightarrow \chi & \text{or} \\ \psi \wedge \chi \in Cn(\phi) \text{ and } \psi \leq \chi \end{cases}$$

Let us finally have a look at the properties of the iterated revision functions that result from this approach. The moderate idea can be shown to amount to the following recipe:

$$(B * \phi) * \psi = \begin{cases} B * (\phi \wedge \psi) & \text{if } \neg\psi \notin Cn(\phi) \\ B * \psi & \text{if } \neg\psi \in Cn(\phi) \end{cases}$$

In marked contrast to conservative belief change, the case distinction for moderate belief change tests the consistency of ψ with respect to the previous input sentence ϕ , and not with respect to the intermediate belief set $B * \phi$ (which includes ϕ). That the most recent piece of information still is the preferred one is evident from the lower line of this definition. Otherwise, however, the method of moderate revision goes some way towards treating ϕ and ψ symmetrically. Philosophically, this seems to be on the right track, since here the two pieces of evidence are grouped together and distinguished from the initial theory B (which may now be regarded as playing the role of a general "background theory"). In conservative belief change the first piece of evidence, ϕ , is merged

³⁴If a preference relation is not connected, i.e. not fully *economic*, its conservative revision cannot strictly speaking be called fully *economical*. For any two relations R_1 and R_2 over a given domain, we can define the *difference* between R_1 and R_2 to be the set of pairs in the domain that are related by R_1 but not by R_2 , or vice versa. The difference between an ordering of possible worlds \preceq and its revision \preceq_{ϕ}^* is not strictly smaller for conservative revision as it is for, say, moderate revision. Consider three ϕ -worlds w_1 , w_2 and w_3 . Let $w_1 \prec w_2$, let w_3 be unrelated by \preceq to both w_1 and w_2 , and suppose that both w_1 and w_3 are minimal in $[\phi]$. Then conservative revision with respect to ϕ introduces the new comparison $w_3 \preceq_{\phi}^* w_2$ which moderate revision does not. Thus the difference between \preceq and its conservative revision by ϕ is not strictly smaller (in terms of set inclusion) than between \preceq and its moderate revision by ϕ . So for the concept of informational *economy* to make perfectly well-defined sense, it seems that fully *economical* behaviour is presupposed. This suggests that the ideas of economy and economics in belief change cannot be neatly separated from each other – *pace* Rott (2003a).

with B , and only after the merging is done, the second piece of evidence, ψ , gets processed. At this second stage, there is no reliable way of knowing that ϕ had been the same kind of thing as ψ .

Having gone thus far, it is natural to argue in favour of a *perfectly* symmetrical treatment of the pieces of evidence and thus call in question the primacy of the most recent piece of information as expressed by AGM's success postulate and respected by all the methods we have been discussing so far. Being altogether indifferent towards the time at which a certain piece of information is received is certainly a form of temporal coherence, too. This is an interesting topic for further research, but it cannot be dealt with here.

We now leave the field of "economical reasoning" and turn to "economic reasoning", i.e., to the left branch of Fig. 1 which refers us to the realm of choices, preferences and utilities.

9. Rational choices and logical properties: What has been done?

As pointed out above, the classical AGM model of belief revision embodies a strong idea of dispositional coherence. More precisely, the agent's dispositions to change his belief set in potential revisions are reflected in choice functions that can be rationalised by a complete pre-ordering. That is, the agent acts *as if* he was a maximiser with respect to such an ordering. That the ordering is complete means that all possible worlds are presumed to be comparable with each other in terms of plausibility (Alchourrón, Gärdenfors and Makinson 1985, Grove 1988), and all sentences are presumed to be comparable with each other in terms of entrenchment (Gärdenfors and Makinson 1988). These facts find expression in the seventh and eighth postulates of AGM which constrain the agent's disposition to change his belief set: (*7) which is sometimes called *Disjunction in the premises* and (*8) which is sometimes called *Rational monotonicity*.

So far we have been discussing preference relations between worlds and sentences. We will now adopt a related, but slightly general approach using choice functions. Preferences are used in the selection of elements from a given 'menu' of options open to the agent, and the economic man will usually select those that are 'best' according to his preferences. In the converse direction, suppose that we have given a certain choice function that selects, for each potential menu, the elements that are 'best' in some unspecified sense. It is a sensible question to ask whether there exists a preference relation \leq such that the choices as determined by the choice function can be rationalised as picking the best elements according to \leq . It is an important fact that not just any choice function can be rationalised in this way, and in it is in this sense that the approach using choice functions is more general than the approach based on preference maximisation. The question is how to characterise, in purely choice-theoretic terms those choice functions that are rationalizable by some preference relation.

Rott (2001, Chapter 7) describes how exactly one can use semantic choice func-

tions (for the selection of most plausible worlds) and syntactic choice functions (for the selection of least entrenched sentences) in the construction of belief revisions, and also how postulates for belief revision correspond to rationality requirements for semantic and syntactic choice functions.³⁵

Table 1 shows how a number of postulates for belief revision correspond to requirements for the choice functions that govern the selection of most plausible models or the selection of least entrenched sentences. In the table, σ refers to a choice functions which selects for any menu S the (typically non-empty) choice set $\sigma(S)$ of ‘best’ elements of S .

POSTULATES FOR REVISIONS	POSTULATES FOR CHOICES
(*7c) If $\psi \in B * \phi$, then $B * (\phi \wedge \psi) \subseteq B * \phi$	If $S \subseteq S'$ and $\sigma(S') \subseteq S$, then $\sigma(S') \subseteq \sigma(S)$
(*7) $(B * \phi) \cap (B * \psi) \subseteq B * (\phi \vee \psi)$ (*7') $B * (\phi \wedge \psi) \subseteq (B * \phi) + \psi$	(α) If $S \subseteq S'$, then $S \cap \sigma(S') \subseteq \sigma(S)$
(*8c) If $\psi \in B * \phi$, then $B * \phi \subseteq B * (\phi \wedge \psi)$	If $S \subseteq S'$ and $\sigma(S') \subseteq S$, then $\sigma(S) \subseteq \sigma(S')$
(*8) If $\neg\psi \notin B * \phi$, then $(B * \phi) + \psi \subseteq B * (\phi \wedge \psi)$	(β^+) If $S \subseteq S'$ and $\gamma(S') \cap S \neq \emptyset$, then $\sigma(S) \subseteq \sigma(S')$
(*8d) $B * (\phi \vee \psi) \subseteq (B * \phi) \cup (B * \psi)$	(Aiz) If $x \in \sigma(S)$ and $y \in \sigma(S')$, then $x \in \sigma(S \cup S')$ or $y \in \sigma(S \cup S')$
(*8wd) $B * (\phi \vee \psi) \subseteq$ $(B * \phi) + \psi \cup (B * \psi) + \phi$	(γ) $\sigma(S) \cap \sigma(S') \subseteq \sigma(S \cup S')$ for syntactic choices
(*8vwd) $B * (\phi \vee \psi) \subseteq$ $Cn((B * \phi) \cup (B * \psi))$	(γ) $\sigma(S) \cap \sigma(S') \subseteq \sigma(S \cup S')$ for semantic choices
(*8n) $B * \phi \subseteq$ $B * (\phi \wedge \psi) \cup B * (\phi \wedge \neg\psi)$	If $S \cap S' = \emptyset$, then $\sigma(S) \subseteq$ $\sigma(S \cup S')$ or $\sigma(S') \subseteq \sigma(S \cup S')$ for semantic choices If $S \cap S' \subseteq Cn(\emptyset)$, then $\sigma(S) \subseteq$ $\sigma(S \cup S')$ or $\sigma(S') \subseteq \sigma(S \cup S')$ for syntactic choices
(*8m) $B * \phi \subseteq B * (\phi \wedge \psi)$	If $S \subseteq S'$, then $\sigma(S) \subseteq \sigma(S')$

Table 1: Correspondences between revisions and choices

(*7) and (*8) have turned out to correspond to conditions on semantic or syntactic choice functions known as *Sen’s Properties* α and β^+ , which in effect require that the revision function be rationalizable by a complete and transitive preference relation.

In order to illustrate how central the conditions α and β^+ are, we have a look at a number of equivalent conditions:

Observation. The following conditions are all equivalent.

- (i) Sen’s properties α and β^+ taken together.

³⁵Important earlier work was done by Lindström (1991).

- (ii) If $S \subseteq S'$ and $\sigma(S') \cap S \neq \emptyset$, then $\sigma(S') \cap S = \sigma(S)$.
(‘*Arrow’s axiom*’; see Moulin 1985, p. 153; Suzumura 1983, p. 25; used for the semantics of counterfactuals by Lewis 1973, p. 58)
- (iii) Either $\sigma(S \cup S') = \sigma(S)$ or $\sigma(S \cup S') = \sigma(S')$ or $\sigma(S \cup S') = \sigma(S) \cup \sigma(S')$.
(‘*Ventilation*’; used for deontic logic by Alchourrón 1993, p. 71, and for a logical reconstruction of linguistic ‘optimality theory’ by Besnard, Fanselow and Schaub 2003, pp. 159–160)
- (iv) If $S \cap S' \neq \emptyset$, then
either $\sigma(S \cup S') = \sigma(S)$ or $\sigma(S \cup S') = \sigma(S')$ or $\sigma(S \cup S') = \sigma(S) \cup \sigma(S')$.
(‘*Restricted Ventilation*’)

Assuming that the domain of σ is closed under finite unions and differences, the proof of this observation is straightforward. I take it that the multiplicity of conditions that have been used by researchers in different context testifies to the importance of these conditions, or to the representability in terms of a complete and transitive relation.

Many concrete systems of belief revision, however, do not satisfy (*7) and (*8) – a fact that shows that these AGM axioms are very strong. Knowing this, it is reassuring to find that one can draw on the rich resources of the theory of rational choice in order to introduce appropriate weakenings of the belief revision postulates. Both Property α and Property β^+ can be weakened in various interesting and reasonable ways, some of which are represented in the table. Postulate (*7c) is a weakening of (*7) that corresponds to the condition *Cut* in non-monotonic reasoning; the parallel weakening of Property α does not seem to play any significant role in the theory of rational choice. The weakening (*8c) of (*8) corresponds to the condition of *cumulative monotonicity* in non-monotonic reasoning; the parallel weakening of Property β^+ is known as *Aizerman’s axiom* in the theory of rational choice (see Moulin 1985). Postulates (*8d), (*8wd) and (*8vwd) are known as variants of *Disjunctive rationality*. The latter two conditions have a well-established counterpart in the theory of choice, viz., Sen’s *Property γ* . Postulate (*8n) is known as *negation rationality* in non-monotonic reasoning.

The theory of rational choice has turned out to be a powerful instrument suitable for analysing and constructing revision operations that are much more flexible than the original AGM ones. Seen from this perspective, belief revision theory can indeed be interpreted as being based on economical principles. We are now going, however, to cast a shadow over this neat picture.

10. Rational choices and logical properties: What should be done?

Almost from its beginning the classical theory of rational choice has been subjected to serious criticism. In this section I will present an argument to the effect that a fundamental problem for the theory of rational choice transfers directly to belief revision theories.

Consider the following example. A well-known philosophy department has announced a position in metaphysics. Among the applicants for the job there are a few persons we happen to know. First, there is Amanda Anderson, a young but already distinguished, excellent metaphysician. Second, we have Bernice Becker, who is also definitely very good, though not quite as accomplished, in metaphysics as Andrews. Becker has also done some substantial research in logic. A third applicant is Carlos Cortez. He has a comparatively slim record in metaphysics, but he is widely recognised as one of the most brilliant logicians of his generation.

Suppose that our initial set of beliefs and expectations about the case includes that neither Anderson nor Becker nor Cortez will get the job (say, because we think that Derek Davidson, an outstanding metaphysician, is the obvious candidate who everyone expects to be appointed anyway). Let us also be clear about the fact that there is only one job available.

Consider now three hypothetical scenarios, each of which describes a *potential* development of the selection procedure (the scenarios do *not* describe a *sequence* of stages of the procedure). In each of these alternative scenarios we are genuinely taken by surprise, because we learn that one of the candidates we had believed to be losing will be offered the position. To make things shorter, we introduce some abbreviations. Let the letters a , b and c stand for the statements that Anderson, Becker and Cortez, respectively, will be offered the position.

Scenario 1. The dean tells us in confidence that it has been decided that either Anderson or Becker will be appointed.³⁶ This message comes down to supplying us with the premise $a \vee b$. Given this premise, we conclude that Anderson, being the better metaphysician, will get the job. We also infer that the other candidates will return empty-handed.

Scenario 2. This is a very unexpected scenario in which we are told by the dean that Cortez is actually the only serious candidate left in the competition. Fortunately, there is no need to invest a lot of thinking here. We accept c in this case.

Scenario 3. In this scenario the dean tells us that it has been decided that either Anderson or Becker or Cortez will get the job, thus supplying us with the premise $a \vee b \vee c$. This piece of information triggers off a rather subtle line of reasoning. Knowing that Cortez is a splendid logician, but that he can hardly be regarded as a metaphysician, we realise that competence in logic is considered to be a non-negligible asset by the selection committee. Still we keep on believing that Cortez will not make it, because his credentials in metaphysics are just too weak. Since, however, logic appears to contribute positively to a candidate's profile, we conclude that Becker, and not Anderson, will get the job.

This qualitative description should do for our purposes, but for readers who prefer more precision, the following story may help. The selection committee

³⁶We take it for granted in this example that the dean is not playing games with us, that she is not lying, that she has the relevant knowledge etc.

has decided to assign points to evaluate the candidates' work. Anderson scores 97 out of 100 in metaphysics, but as she has done no logic whatsoever, she scores 0 here. Becker scores 92 in metaphysics and a respectable 50 in logic. Cortez scores only 40 in metaphysics, but boasts of 99 in logic. In scenario 1, we take it that metaphysics is the only desideratum, so clearly Anderson must be the winner. Scenario 2 is trivial. In scenario 3, we gather that, rather unexpectedly, logic matters. As can easily be verified, any weight we attach to logic between $1/10$ and $1/2$ (with metaphysics taking the rest) will see Becker end up in front of both Anderson and Cortez.

Let us now summarise our conclusions from the various premises that the dean supplies us with. Our initial belief set B contains $\neg a$, $\neg b$, $\neg c$ and d among other things. In scenario 1, the new piece of information $a \vee b$ leads us to accept a and $\neg b$ (along with $\neg c$ as well as $\neg d$ which we will not mention any more). In scenario 2, accepting c simply makes us retain $\neg a$ and $\neg b$. In scenario 3, the new piece of information $a \vee b \vee c$ leads us to accept that $\neg a$ and b . We can now show that this situation refutes some of the basic logical principles of "economic" belief revision.

First, the example shows that Disjunction in the premises, (*7), does not hold. Take (*7) and substitute $a \vee b$ for ϕ and c for ψ . Then notice that $\neg b$ is believed if the input is $a \vee b$, and also if the input is c . But $\neg b$ is not believed if the input is $a \vee b \vee c$. Thus the revised belief set $B * (a \vee b \vee c)$ does not include what is common to $B * (a \vee b)$ and $B * c$, and (*7) is violated.

Secondly, we find that the situation does not conform to the weakened monotonicity postulate (*8c). Take (*8c) and substitute $a \vee b \vee c$ for ϕ and $a \vee b$ for ψ . Even though we believe that $a \vee b$ is true if we are given the information $a \vee b \vee c$, it is not the case that everything believed on the basis of the latter is also believed on the basis of $(a \vee b \vee c) \wedge (a \vee b)$ which is equivalent with $a \vee b$. Sentences $\neg a$ and b are counterexamples. Thus the revised belief set $B * (a \vee b \vee c)$ is not a subset of the belief set $B * ((a \vee b \vee c) \wedge (a \vee b)) = B * (a \vee b)$, and (*8c) is violated. *A fortiori*, (*8) is violated as well.

What do these problems derive from? We said that principles of belief revision can be systematically interpreted in terms of rational choice. On this interpretation, Disjunction in the premises, (*7), turns out to be an instantiation of one of the most fundamental conditions – perhaps *the* most fundamental condition – of the theory of rational choice: Sen's Property α . This condition, also called *Independence of Irrelevant Alternatives* or *Chernoff property*, says that any element which is optimal in a certain menu remains an optimal element after some other elements have been cancelled from the menu. The three scenarios in our example are modelled after well-known choice situations in which Property α is violated – cases that also happen to infringe Aizerman's axiom. Both properties may fail to be satisfied if the very 'menu' from which the agent is invited to choose carries important information. This phenomenon which Sen calls the 'epistemic value' or the 'epistemic relevance of the menu'³⁷

³⁷Sen (1993, pp. 500–503; 1995, pp. 24–26) has brought the problem to wide attention.

suggests that the context of choice has a decisive influence upon the shape of the preference relation of the agent – an idea strongly opposed to the idea of context-independent preferences that underlies the classical theory of rational choice. The *locus classicus* for the problem is a passage in Luce and Raiffa (1957, p. 288) who chose to avoid the problem of the epistemic value of the menu by *fiat*:

This illustrates the important assumption implicit in axiom 6 [essentially Property α , H.R.], namely, that adding new acts to a decision problem under uncertainty *does not alter one's a priori information as to which is the true state of nature*. In what follows, we shall suppose that this proviso is satisfied. In practice this means that, if a problem is first formulated so that the availability of certain acts influences the plausibility of certain states of nature, then it must be reformulated by redefining the states of nature so that the interaction is eliminated.

This may make good sense as a rejoinder in the context of the general theory of choice and decision. An explanation of how information is surreptitiously conveyed through the particular contents of the menu and how it affects the chooser's preferences is simply not this theory's business. Unfortunately, the same defence is not available for the problem highlighted by our example above. It *is* the task of the theory of belief formation to model how one's prior belief set is affected by information received from external sources. This is precisely what this theory has been devised to explain, and therefore the anomaly cannot be pushed to a neighbouring research field. We cannot find fault with the dean's message for the very fact that it conveys information!

The question raised by our example is a general one. It is hard to get rid of the feeling that the dean's remark about the final candidates tells us more than meets the ear. The fact that a logician gets *mentioned* as a top-ranking contender or that logic becomes a *topic* seems to carry surplus information, over and above the propositional content of the corresponding statement. Does the very fact that a sentence is offered in a menu for acceptance have a special relevance for processes of belief revision that has been overlooked so far? Or are there other ways out of the predicament?

11. Conclusion

We have reviewed work in the tradition of the AGM approach to belief revision, arguably the most prominent logical paradigm for the purification of belief sets from contradictions. Our overarching questions were to what extent economic(al) principles have played a role in the *actual* development of this paradigm, and to what extent such considerations *should* have been followed. Our conclusions are mostly negative. Informational economy (conservatism) with respect to beliefs, although widely advertised as *the* central motivation of belief revision models, turns out not to have played anything like a dominant role in the development of such models, and we have found no reason why it

should. Conservatism with respect to revision-guiding preferences has fact been suggested as a strategy for iterated belief revision, but it soon turned out to have unwelcome consequences. So belief revision theory has as a matter of fact not focused on *economy*, and the idea of *economical* belief revisions has very limited normative force, too. Regarding *economic* belief revision, our findings are more encouraging. It is possible to reconstruct large parts of belief revision in terms of rational choice theory. As a matter of fact, ideas coming from *economics* have prevailed in the AGM paradigm and related approaches. However, at the end of the paper we have found that a fundamental problem of the general theory of choice seriously infects the specific application area of belief revision. Thus the use of rational choice theory has helped us to spot a new puzzle rather than to solve old problems. There is a lot of work that waits to be done in cognitive economics.³⁸

References

- Alchourrón, Carlos: 1993, ‘Philosophical Foundations of Deontic Logic and the Logic of Defeasible Conditionals’, in J.-J. Ch. Meyer and R.J. Wieringa (eds.): *Deontic Logic in Computer Science: Normative Systems Specifications*, Wiley, pp. 43–84.
- Alchourrón, Carlos, Peter Gärdenfors and David Makinson: 1985, ‘On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions’, *Journal of Symbolic Logic* 50, 510–530.
- Arló-Costa, Horacio, and Isaac Levi: 1996, ‘Two Notions of Epistemic Validity: Epistemic Models for Ramsey’s Conditionals’, *Synthese* 109, 217–262.
- Besnard, Philippe, Gisbert Fanselow and Torsten Schaub: 2003, ‘Optimality Theory as a Family of Cumulative Logics’, *Journal of Logic, Language and Information* 12, 153–182.
- Boutilier, Craig: 1993, ‘Revision Sequences and Nested Conditionals’, in R. Bajcsy (ed.), *IJCAI-93 – Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 519–525.
- Boutilier, Craig: 1996, ‘Iterated Revision and Minimal Change of Conditional Beliefs’, *Journal of Philosophical Logic* 25, 263–305.
- Broome, John: 1999, *Ethics Out of Economics*, Cambridge University Press, Cambridge.
- Darwiche, Adnan, and Judea Pearl: 1994, ‘On the Logic of Iterated Belief Revision’, in Ronald Fagin, ed., *TARK ’94 – Proceedings of the Fifth Conference on Theoretical Aspects of Reasoning About Knowledge*, Pacific Grove, Cal.: Morgan Kaufmann, 5–23.
- Darwiche, Adnan, and Judea Pearl: 1997, ‘On the Logic of Iterated Belief Revision’, *Artificial Intelligence* 89, 1–29.
- Debreu, Gerard: 1959, *Theory of Value*, Wiley, New York.
- Edgeworth, Francis Ysidro: 1881, *Mathematical Psychics*, Kegan Paul, London.
- Eliasz, Kfir, and Efe A. Ok: 2003, ‘Indifference or Indecisiveness? Choice-Theoretic Foundations of Incomplete Preferences’, manuscript, Department of Economics, New York University, April 2003.

³⁸Thanks go to Vincent F. Hendricks, Eckehart Köhler, Carsten Köllmann, Maria Kronfeldner, Hannes Leitgeb, Nils-Eric Sahlin, Gerhard Schurz, Krister Segerberg, Wolfgang Spohn and Bernard Walliser for helpful comments and discussions.

- Fuhrmann, André: 1993, ‘Observations on Validity and Conditionals in Belief Revision Systems’, *Journal of Applied Non-Classical Logics* 3, 225–238.
- Gärdenfors, Peter: 1986, ‘Belief Revisions and the Ramsey Test for Conditionals’, *Philosophical Review* 95, 81–93.
- Gärdenfors, Peter: 1979, ‘Conditionals and Changes of Belief’, in Ilkka Niiniluoto and Raimo Tuomela (eds.), ‘The Logic and Epistemology of Scientific Change’, *Acta Philosophica Fennica* 30 (1978), nos. 2–4, 381–404.
- Gärdenfors, Peter: 1986, ‘Belief Revisions and the Ramsey Test for Conditionals’, *Philosophical Review* 95, 81–93.
- Gärdenfors, Peter: 1988, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge, Mass.
- Gärdenfors, Peter, and David Makinson: 1988, ‘Revisions of Knowledge Systems Using Epistemic Entrenchment’, in Moshe Vardi (ed.), *TARK’88 – Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, Los Altos: Morgan Kaufmann, 83–95.
- Gettier, Edmund: 1963, ‘Is Justified True Belief Knowledge?’, *Analysis* 23, 121–123.
- Glaister, Stephen: 1998, ‘Symmetry and Belief Revision’, *Erkenntnis* 49, 21–56.
- Grove, Adam: 1988, ‘Two Modellings for Theory Change’, *Journal of Philosophical Logic* 17, 157–170.
- Hansson, Sven Ove: 1999, *A Textbook of Belief Dynamics: Theory Change and Database Updating*, Kluwer, Dordrecht.
- Hansson, Sven Ove (ed.): 1997, Special Issue on ‘Non-Prioritized Belief Revision’, *Theoria* 63, 1–134.
- Hausman, Daniel: 1998, ‘Economics, Philosophy of’, *Routledge Encyclopedia of Philosophy*, electronic edition.
- Hausman, Daniel M., and Michael S. McPherson: 1996, *Economic Analysis and Moral Philosophy*, Cambridge University Press, Cambridge.
- Kelly, Kevin: 1999, ‘Iterated Belief Revision, Reliability, and Inductive Amnesia’, *Erkenntnis* 50, 11–58.
- Konieczny, Sébastien, and Ramón Pino Pérez: 2000, ‘A Framework for Iterated Revision’, *Journal of Applied Non-Classical Logics* 10, 339–367.
- Lehmann, Daniel: 2001, ‘Nonmonotonic Logics and Semantics’, *Journal of Logic and Computation* 11, 229–256.
- Lewis, David: 1973, *Counterfactuals*, Blackwell, Oxford.
- Liberatore, Paolo: 1997, ‘The Complexity of Iterated Belief Revision’, in *Proceedings of the Sixth International Conference on Database Theory (ICDT’97)*, Lecture Notes in Computer Science 1186, Berlin: Springer, 276–290.
- Lindström, Sten: 1991, ‘A Semantic Approach to Nonmonotonic Reasoning: Inference Operations and Choice’, Uppsala Prints and Preprints in Philosophy, Department of Philosophy, University of Uppsala, 1991:6.
- Luce, R. Duncan, and Howard Raiffa: 1957, *Games and Decisions*, Wiley, New York.
- McFadden, Daniel: 1999, ‘Rationality for Economists?’, *Journal of Risk and Uncertainty* 19, 73–105.
- Moulin, Hervé: 1985, ‘Choice Functions over a Finite Set: A Summary’, *Social Choice and Welfare* 2, 147–160.
- Nayak, Abhaya: 1994, ‘Iterated Belief Change Based on Epistemic Entrenchment’, *Erkenntnis* 41, 353–390.

- Nayak, Abhaya, Norman Foo, Maurice Pagnucco and Abdul Sattar: 1996, ‘Changing Conditional Beliefs Unconditionally’, in Yoav Shoham (ed.), *TARK’96 – Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, pp. 119–135.
- Nayak, Abhaya, Maurice Pagnucco and Pavlos Peppas: 2003, ‘Dynamic Belief Revision Operations’, *Artificial Intelligence*, forthcoming.
- Ok, Efe A.: 2002, ‘Utility Representation of an Incomplete Preference Relation’, *Journal of Economic Theory* 104, 429–449.
- Olsson, Erik: 2002, ‘Belief Revision, Rational Choice and the Unity of Reason’, *Studia Logica* 73, 219–240.
- Pagnucco, Maurice, and Hans Rott: 1999, ‘Severe Withdrawal (and Recovery)’, *Journal of Philosophical Logic* 28, 501–547. (Corrected reprint in the *JPL* issue of February 2000.)
- Papini, Odile: 2001, ‘Iterated Revision Operations Stemming from the History of an Agent’s Observations’, in Mary-Anne Williams and Hans Rott (eds.), *Frontiers in Belief Revision*, Dordrecht: Kluwer, 279–301.
- Quine, Willard V.O., and Joseph S. Ullian: 1978, *The Web of Belief*, second edition, Random House, New York.
- Radford, Colin: 1966, ‘Knowledge – By Examples’, *Analysis* 27, 1–11.
- Rosenberg, Alexander: 1983, ‘If Economics Isn’t Science, What Is It?’, *Philosophical Forum* 14, 296–314. Reprinted in E.D. Klemke et al. (eds.), *Introductory Readings in the Philosophy of Science*, 3rd edition, Prometheus Books, Amherst 1998, 154–170.
- Rosenberg, Alexander: 1995, ‘Philosophy of Economics’, in Robert Audi (ed.), *Cambridge Dictionary of Philosophy*, Cambridge University Press, Cambridge, pp. 582–583.
- Rott, Hans: 1993, ‘Belief Contraction in the Context of the General Theory of Rational Choice’, *Journal of Symbolic Logic* 58, 1426–1450.
- Rott, Hans: 1999, ‘Coherence and Conservatism in the Dynamics of Belief. Part I: Finding the Right Framework’, *Erkenntnis* 50, 387–412.
- Rott, Hans: 2000, ‘Two Dogmas of Belief Revision’, *Journal of Philosophy* 97, 503–522.
- Rott, Hans: 2001, *Change, Choice and Inference*, Oxford University Press.
- Rott, Hans: 2003a, ‘Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence’, *Journal of Logic and Computation* 13, 111–145.
- Rott, Hans: 2003b, ‘Basic Entrenchment’, *Studia Logica* 73, 257–280.
- Rott, Hans: 2003c, ‘A Counterexample to Six Fundamental Principles of Belief Formation’, manuscript, January 2003.
- Schlechta, Karl: 1996, ‘Some Completeness Results for Stopped and Ranked Classical Preferential Models’, *Journal of Logic and Computation* 6, 599–622.
- Sen, Amartya K.: 1993, ‘Internal Consistency of Choice’, *Econometrica* 61, 495–521.
- Sen, Amartya K.: 1995, ‘Is the Idea of Purely Internal Consistency of Choice Bizarre?’, in J. E. J. Altham and R. Harrison (eds.), *World, Mind, and Ethics. Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge, 19–31.
- Spohn, Wolfgang: 2003, ‘Die Logik und das Induktionsproblem’, in Peter Schroeder-Heister und Wolfgang Spohn (eds.), *Logik in der Philosophie*, Synchron-Verlag, Hei-

delberg, forthcoming.