TWO DOGMAS OF BELIEF REVISION*

Quine's epoch-making paper "Two Dogmas of Empiricism"[1] sealed the fate of logical positivism by undermining two presumptions of classical as well as modern empiricism: The idea that there is a sharp distinction between analytical judgements (knowledge of meanings) and synthetical judgements (knowledge of facts), and the idea that every meaningful sentence can be reduced to a construction upon observation reports. The most interesting feature from the point of view of the present article, however, is that Quine closed his paper with a beautiful section on "empiricism without the dogmas" that deals almost exclusively with a topic that would nowadays be called *theory change*, or *belief revision*. He paints a picture of how we should, and for the most part do, accommodate the scientific lore if we meet with recalcitrant experiences. According to Quine's pragmatist approach, theory revision is a matter of choice, and the choices are to be made in a such a way that (a) the resulting theory squares with the experience, (b) it is simple and (c) the choices disturb the original theory as little as possible.

Quine's picture is mainly metaphorical, and he never intended to turn it into a formal theory of theory change. The situation today is different. At least since the seminal work of Alchourrón, Gärdenfors and Makinson ("AGM" for short) in the early 1980s, there has been a clear-cut paradigm of how to logically formalize theory change.[2] The original

*A preliminary version of this paper was presented at the Third Conference on Logic and the Foundations of the Theory of Games and Decisions (LOFT3, Torino, December 1998). I am grateful to Eduardo Fermé, Oliver Schulte, and in particular Isaac Levi and David Makinson for a number of very valuable comments on earlier versions of this paper.

A slightly different version of this paper has been published in
The Journal of Philosophy **97** (2000), pp. 503–522.

AGM approach has turned out to be limited in various ways, and many extensions and revisions of it have been developed over the past two decades. I do not want to find fault in this paper with any of the formal theories of belief revision that are currently being advocated. My aim is rather to call in question what has been taken to be *the* principal idea behind the current theories of belief revision: The idea of *informational economy.* This idea is basically the same as Quine's criterion (c), often called the *principle of minimum mutilation*, or also *(doxastic or epistemic) conservatism.* I shall try to cast doubt not on the principle itself, but on its being *the* philosophical background that makes AGM style theories of belief revision intelligible in the first place.

The principle on informational economy tells us that we should not give up our beliefs beyond necessity. Now it is not spelt out clearly what "necessity" here means. It might be suggested that we can satisfy the principle of minimal change in an ideal way simply by not changing our beliefs at all.[3] But that is certainly not intended. There are basic requirements for belief revision that have to be satisfied, and only against the background of these conditions can we apply the minimal change principle. One of the basic requirements that we are going to leave unquestioned in this paper is that revisions should be *successful* in accommodating new information: The kind of revision that we consider is such that incoming input gets always accepted. That this is too strict as a general constraint for a truly realistic theory of belief change has been recognized by many writers and is taken account of in numerous recent papers on "non-prioritized belief revision."[4]

The second basic requirement we are going to impose is that agents should be taken to be ideally competent as regards matters of logic. They should accept all the consequences of the beliefs they hold (that is, their set of beliefs should be *logically closed*), and they should rigorously see to it that their beliefs are *consistent* (it is just the task of belief revision theory to give an account of *how* they can manage that their beliefs remain consistent). As stressed by researchers in belief *base* revision and paraconsistent logics respectively, the insistence on closure and consistency may be regarded as unrealistic in general. The case, however, is not as clear as with the success condition, since there are at least two interpretations of belief that make the logical restrictions compelling: beliefs as attitudes ascribed by a third person, and beliefs as commitments.[5]

I shall assume from now on that these *basic requirements*—that belief sets should be closed and that revisions should be successful and lead to consistent belief sets—form the background against which questions of informational economy or minimal change are discussed. Agents who do not change their beliefs at all can perhaps be called very "economical" in the administration of what they currently possess, but they are not able to successfully interact with the world. "Do nothing!" is not a feasible option for our cognitive life.

I

Informational economy has been proclaimed to lie at the basis of belief change from the very beginning of the systematic study of belief revision. The idea comes in two distinct versions:

(1) When accepting a new piece of information, an agent should aim at a minimal change of his old beliefs.

(2) If there are different ways to effect a belief change, the agent should give up those beliefs that are least entrenched.

These maxims have been accompanying belief change theory since its inception, and they were repeated in the literature time and again. Here are some relevant references.

Ad (1).

> "The concept of contraction leads us to the concept of *minimal change of belief*, or briefly, *revision*." (Makinson[6])

> "The criterion of informational economy demands that as few beliefs as possible be given up so that the change is in some sense a *minimal* change of K to accommodate for A." (Gärdenfors[7])

> "The amount of information lost in a belief change should be kept minimal." (Gärdenfors and Rott[8])

> "At the centre of the AGM theory [of theory change] are a number of approaches to giving formal substance to the maxim [of minimal mutilation: Keep incisions into theories as small as possible!]." (Fuhrmann[9])

> "The hallmark of the AGM postulates is the principle of minimal belief change, that is, the need to preserve as much of earlier beliefs as possible and to add only those beliefs that are absolutely compelled by the revision specified." (Darwiche and Pearl[10])

Ad (2).

> "When a belief set K is contracted (or revised), the sentences in K that are given up are those with the *lowest* epistemic entrenchment." (Gärdenfors[11])

> "The guiding idea for the construction is that when a knowledge system K is revised or contracted, the sentences in K that are given up are those having the *lowest* degrees of epistemic entrenchment." (Gärdenfors and Makinson[12])

> "In so far as some beliefs are considered more important or entrenched than others, one should retract the least important ones." (Gärdenfors and Rott[13])

> "... when it comes to choosing between candidates for removal, the least entrenched ones ought to be given up." (Fuhrmann[14])
>
> "A hallmark of the AGM theory is its commitment to the principle of *informational economy*: beliefs are only given up when there are no less entrenched candidates. ... If one of two beliefs must be retracted in order to accommodate some new fact, the less entrenched belief will be relinquished, while the more entrenched persists." (Boutilier[15])

Although (1) and (2) look quite different and it is not immediately clear how they fit together, there is a result that appears to show that they are even "at root identical"[16] and that they can therefore be viewed as two incarnations of a unified idea of informational economy. The result mentioned maps onto one another two prominent types of belief change constructions investigated by Alchourrón, Gärdenfors and Makinson: belief changes obtained by partial meets of maximal non-implying sets of beliefs,[17] and belief changes based on relations of epistemic entrenchment.[18] These two methods are generally taken to be closely associated with (1) and (2), respectively, and so it is both surprising and pleasing to find that they can be proved equivalent in a rather strict sense by directly relating the underlying relations used.[19]

The overall picture that we have now gotten of the AGM approach seems nice and harmonious. But I want to argue in this paper that *maxims (1) and (2) are a travesty of the principles that have actually been followed in the traditional theories of belief revision*. The philosophical underpinnings of the prevailing theories of belief change are not at all what the folklore would like to have them.

As will become clear later on, I call (1) and (2) "dogmas" *not* because almost all researchers actually *kept to* these principles (quite the opposite is true), but because so many authoritative voices have *professed* that these principles are the principal philosophical or methodological rationale for their theories.

II

The following simple fact that will be used to attack principle (1) draws only on the basic requirements for belief revision. Irrespective of the prior belief set K, I shall call any consistent and logically closed belief set including a sentence ϕ a *candidate revision* of K by ϕ.

Observation 1. 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 beliefs on which they differ with the prior belief set.[20]

This observation shows that the most straightforward idea to measure conservativity in terms of a comparison of the sets of beliefs in which original and revised theories differ fails: there are no two candidate revisions such that the symmetric difference between the one and the original belief set K is a proper subset of the symmetric difference between the other and K.[21] If we use this criterion, then *all* (successful and closed) candidate revisions of a consistent belief set are minimally removed from it. Maxim (1) therefore cannot be a good recommendation. In one reading, it can be used to license the choice of any arbitrary (successful and closed) revision. In another reading, one which recommends the cautious strategy of taking the greatest common denominator of all minimally distant belief sets, it has as a result that all belief-contravening belief changes are amnesic.[22] Here I call a revision *amnesic* if the revised belief set consists of nothing else but the logical consequences of the sentence to be revised with; otherwise we call it *anamnestic*.

This seems to be a strong indication that in order to make good sense of the idea of informational economy we need to turn to a refined description of belief states, such as the one afforded by the ordering of beliefs in terms of their epistemic entrenchment. The intuition behind the term 'entrenchment,' in the sense that has been given a formal analysis in the literature on belief revision, is basically that of a relation of comparative retractability. A belief ϕ is *more entrenched* than another belief ψ if and only if the agent holds on to ϕ and gives up ψ after learning (or hypothetically assuming) that either ϕ or ψ is (can be, may be) false.[23]

Having a rudimentary understanding of the term 'entrenchment' as it is used in the AGM tradition, we now turn to principle (2). Like the previous observation, the following one is mathematically trivial, but conceptually striking. We call a new piece of information *moderately surprising* with respect to a belief that K if its negation is an element of K which is more than minimally entrenched in K.

Observation 2. Suppose we want to revise a belief set by a sentence ϕ and find two elements of the belief set that non-redundantly entail the negation of ϕ. Then it may well be rational, according to the standard belief revision constructions, to restore consistency by removing the *more* entrenched and retain the *less* entrenched belief. In fact, such a situation can *always* be identified in an anamnestic revision by a consistent and moderately surprising sentence.[24]

Principle (2) thus turns out to be plainly wrong on the "local" interpretation given to it in this observation (where reference is made to entailment sets with only two elements). Actually, the proof of Observation 2 does not presume that belief revision is effected according to the traditional Gärdenfors-Makinson construction recipe for contractions,[25] but it draws only on the logical structuring of the entrenchment relations these authors introduce. Figure 1 gives a simple illustration of the situation described in Observation 2, using a Grovean possible worlds representation of AGM-style belief change.[26] In the center of such a system of spheres we find the worlds satisfying all current beliefs. The entrenchment of a belief ϕ is measured by the distance of the closest $\neg\phi$-worlds from the center. The result of revising the current belief set K by a new piece of information ψ is the set $K * \psi$ of all sentences satisfied by all closest ψ-worlds.

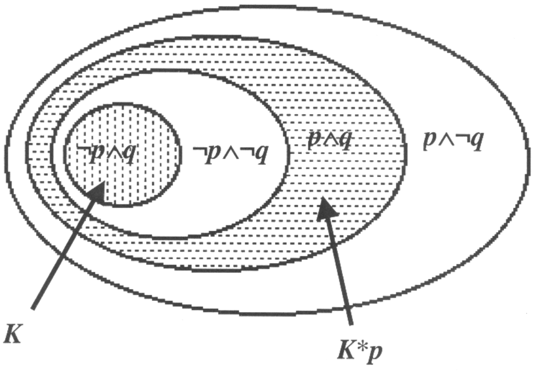

Figure 1: Example for Obs. 2

The situation is this: Two beliefs, q and $q \supset \neg p$, form a set that minimally implies the negation of the incoming information p; q is strictly less entrenched than $q \supset \neg p$. And yet the agent keeps the less entrenched q and abandons the more entrenched $q \supset \neg p$ —and all this according to the widely accepted AGM methodology![27]

The situation sketched here is explained in more detail in the appendix. I believe that the paradoxical appearance of Observation 2 is not due to a shortcoming of the AGM recipe, but due to the fact that principle (2) expresses a wrong ideology (or, perhaps, the fact that principle (2) gives the wrong expression to an unclear idea[28]).

III

In the previous section I have been attacking the almost universally held idea that informational economy *alias* minimum mutilation is the basis of the most promising approaches to belief change. My reason for doing so is that on the one hand the idea of minimal change is difficult to formulate (Observation 1), and that on the other hand its application in the construction of revisions is ill-understood (Observation 2).

I am now going to briefly discuss four ways of defending the minimal change idea against the above criticism. The first defense claims that I have misdirected my attack by leveling it against belief revisions rather than against the more fundamental operations of belief contraction for which even two ways of characterizing minimality seem available. The second defense counters my arguments by saying that I have been forgetful of nothing less than the core of the AGM theory, to wit, the central representation theorems linking the well-known AGM rationality postulates to the existence of orderings that are suitable for determining a reasonable standard of minimality. The third defense advances the claim that if AGM have not been successful in circumscribing minimality, this was so only because their notion of a doxastic state was too simplistic, and as soon as we introduce richer models which in addition to plain beliefs also represent dispositions for belief change, we will find a natural way to make conservatism both comprehensible and defensible. The fourth defense argues that we should take up the connection of beliefs with the real world that has been lost by the usual "internalistic" approaches to belief change. If we respect truth as an aim of inquiry in our characterizations of minimal change, the disturbing results may be expected to disappear.

I shall argue that all these attempts to save the two dogmas, though promising at first sight, are mistaken.

III.A. Contractions. The first attempt to save the idea of informational economy as a cornerstone of the traditional theories of belief change is to recall that most of the time AGM and their followers have written about belief *contractions*, i.e., the problem how to rationally give up a belief without acquiring any new belief. Every contracted belief set is by definition a subset of the prior set, so here the problem with symmetric differences (that made it possible to prove Observation 1) does not arise. Thus we have an obvious standard by which to measure differences. If a candidate contraction[29] K' is a proper subset of another candidate contraction K'', then K' is farther removed from the original set K than K'', since the set-theoretic difference $K - K'$ properly includes $K - K''$. An early suggestion of Isaac Levi's is that each legitimate revision by ϕ must be decomposable into a contraction of K by $\neg\phi$, followed by an addition of ϕ and logical closing-up (this is the so-called Levi identity). In this way distances of revisions from the prior set are indeed expressible in terms of well-defined distances of the corresponding contractions from the prior set. Unfortunately, this is not intuitively plausible. If K' and K'' are two candidate contractions with respect to $\neg\phi$ and if K' is a subset of K'', then the result of adding ϕ to both of them and closing them up under logical consequences leads to a belief set $Cn(K' \cup \{\phi\})$ that is a subset of the belief set $Cn(K'' \cup \{\phi\})$ (here Cn is the operation of taking the logical consequences of a set of sentences). But is the latter closer to K than the former? There is no justification for saying so. While the latter set contains original beliefs that may have been lost in the former, it also contains novel beliefs that were not held before and can therefore be considered to be gratuitous additions. This fact is basically what Observation 1 is reflecting.

Another advantage of focussing on contractions might be seen in the fact that AGM work with the so-called postulate of Recovery for contractions. This postulate says that when contracting with respect to some sentence ψ, we may withdraw beliefs only to such an extent that adding back ψ immediately after the contraction will make us recover all the original beliefs. This postulate can indeed be regarded as an explicit condition of minimum mutilation. However, as a general defense of minimum mutilation, the argument fails for three reasons. First, Recovery is at best a partial encoding of informational economy, since it does not even disallow amnesic belief change.[30] Second, the information-preserving effects of Recovery evaporate completely when contractions are used as intermediate steps in the construction of belief revisions with the help of the Levi identity. Third, meeting the requirement of Recovery has counterintuitive consequences in many situations. This postulate has accordingly suffered from severe criticism by numerous members of the belief revision community and cannot be considered as belonging to the core of traditional belief revision theory any longer.

We conclude that neither of the two arguments to support the idea of minimal change through a consideration of belief contractions is convincing.

III.B. Reconstructions. The success of the program of Alchourrón, Gärdenfors and Makinson was not in the first place based on their putting together a list of "rationality postulates," but on their showing that the belief change behavior thus axiomatized can be represented by a number of interesting and plausible explicit constructions (such as the above-mentioned revisions based on partial meets and entrenchments). All of these constructions make use of some kind of extra-logical structure of the belief state that guides the selection of most plausible or least plausible worlds or sentences featuring in the process of belief change. Typically this selection is determined by a preference relation that is independent of the particular sentence that is to be accepted or withdrawn. The second line of defense against my attack on the notion of minimal change in belief change is based on the following argument: It is just the upshot of the many representation theorems in the literature that rational belief change can be reconstructed as belief change determined by a minimization condition with respect to some underlying doxastic preference relation.[31] Such preference relations—hidden structures of belief as it were—range alternatively over beliefs, sets of beliefs, models or worlds.[32] And at least in the Grovean possible worlds representation, the minimization procedure has been taken to reflect a form of minimal change: From the models that satisfy the original theory, the agent passes over to the minimally remote ("closest") worlds that satisfy the input sentence.

A first reply to this defense is that contrary to the declarations I cited in Section I, the AGM postulates for belief revision do not place *any* constraints regarding the preservation of beliefs in the case where the incoming information is inconsistent with the current theory—the interesting *belief-contravening* case.[33] So if there is no encoding of minimal change or informational economy in the postulates (and this is what I am claiming), it would be strange if we could conjure up such an idea through mathematical representation theorems.

Secondly, it needs to be supported by extra arguments, I think, that the canonical preference relations constructed in the proofs of AGM-style representation theorems are more than just technical devices and can indeed be given an interpretation that fits the desired economical meaning. Technically, it is of course possible to reverse the underlying preference orderings and select the worlds that are maximally remote ("farthest away") from the prior belief state.

Thirdly, we have seen in Observation 2 that in the case of epistemic entrenchment, the application of the stipulated orderings is not just straightforward minimization.[34]

In other modelings (like partial meet contraction, safe contraction or possible worlds models) minimization is only one step in a complex procedure of constructing revised belief sets, and its effects are at least partially neutralized in subsequent steps of this very procedure. There are tacit principles at work here, according to which a believer should respect ties in her underlying preferences and should treat equally objects that she holds in equal regard.[35] And it is again implicit in the construction of AGM-style belief changes that these principles are given priority over principles of minimal change. As a result, these constructions even license amnesic belief change.

We conclude that the argument to support the idea of minimal change through a rational reconstruction of belief revision in terms of hidden preference relations is not convincing.

III.C. Dispositions. The third line of defense of minimal change concedes that the arguments presented in Section II are correct—provided that they are viewed as an attack on the particular form of early belief change models. But the blame can be laid on an illegitimate identification of belief states with deductively closed sets of sentences, an identification that Alchourrón, Gärdenfors and Makinson seemed to advocate themselves. On a proper understanding, so the defense continues, the formal model of a belief state should already encompass the means necessary for performing belief changes, and should therefore include something like the preference structures we talked about in the previous section. Such a move will allow us not only to provide a smooth mechanism that can easily cope with iterated changes of belief, but also to get a grip on the elusive notion of minimal change in belief-contravening revisions. If we extend AGM theory and conceive more sophisticated structures (e.g., orderings or selection functions over worlds or sentences) as representations of doxastic states, we can find natural ways of defining distances between them. On this construal, the presumption underlying Observation 1, viz., that in order to define minimal changes of belief states one has to compare differences between closed belief sets, is wrong.

Two different but ultimately equivalent ways of implementing this idea are presented in papers by Boutilier and Rott,[36] where the former is based on a possible worlds modeling similar to Grove's, while the latter is based on epistemic entrenchment relations. Both Boutilier and Rott furnish an explicit formal interpretation of minimal change. The basic idea is that in order to effect a change of the belief state represented by a certain ordering (of possible worlds or sentences), one should change in this ordering only the positions of a uniquely identifiable *minimal*[37] set of elements, just as much as is necessary in order to make the change "successful." It has turned out, however, that the possibility to come up with a truly conservative definition of (iterated) belief change has to be dearly purchased. Darwiche and Pearl were the first to make this observation by way of an intuitive counterexample the force of which is acknowledged by Boutilier.[38] A more general twist to the problem is added by the following observation:[39]

Observation 3. If doxastic states encompass revision-guiding structures (like preference orderings or selection functions), then belief-contravening revisions that obey the maxim of minimum mutilation have unacceptable consequences: They violate a requirement of temporal coherence.

We state this theorem without proof. The gist of the argument concerns iterated belief changes and the role of the recency of information. The postulate of success rules that a piece of information that is just coming in (is "most recent") is maximally appreciated and therefore invariably accepted. However, when the next revision takes place, the information just taken up (now the "second most recent" belief) turns out to be very weakly entrenched. As Darwiche and Pearl's red bird example shows, conservative belief change in this sense can be too dismissive: He who revises a *tabula rasa* belief state first by "Fred is a bird," then by "Fred is red," and finally by "Fred is not a bird" will end up in a belief state that does not include "Fred is red"—and that is certainly a counterintuitive consequence.

In another sense, conservative belief change is too tenacious. Revising the belief state that includes "Barney is either French or Flemish" first by "Barney is not French" and then by "Barney is not Flemish" will preserve the initial disjunction (and even yield that Barney is French), which I think is again contrary to intuition. In conservative belief change as defined by Boutillier and Rott, incoming information indeed gets always accepted, but only at the lowest level possible, so in future cases of conflict it has to take all the blame and is the first candidate for removal.[40] In that respect, therefore, old beliefs are treated with more respect than new beliefs. This form of belief change shows no principled stance towards the recency of information and leads to a doxastic behavior that must be called incoherent with respect to time.

In sum, it is true that more encompassing representations of belief change are desirable and even necessary, if we are to deal with the important problem of iterated revisions. However, the most natural suggestion how to reflect the idea of minimal change in such a framework does not lead to a satisfactory solution of the belief revision problem. The argument proposing that the idea of minimal change or informational economy can be supported through an enrichment of the notion of a belief state by structures representing doxastic dispositions is not convincing.

III.D. Truths. Except for the principle of minimum mutilation, little work has been done in the theory of belief revision to account for Quine's criteria that we mentioned in the introduction. It is particularly irritating that the "correspondence" of our beliefs with the real world—Quine's insistence that the empirical phenomena have to be gotten right—is not at all captured by the usual modelings.[41] More than a hundred years ago, William James formulated the main goal of belief fixation and belief change in his famous lecture on "The Will to Believe": "*We must know the truth*; and *we must avoid error*—these are our first and great commandments as would-be knowers."[42] As one of the founding fathers of pragmatism, James is a predecessor of Quine's as well as of Isaac Levi's, a leading philosopher of belief change who did take over James's catchword. The charge now against formal accounts of belief change which can also be turned against my way of capturing minimal change in Principle (1) above, is that one should worry more about truth. Only true beliefs are valuable, and there is no point in preserving false beliefs. A suitably qualified version of (1) that takes into account the basic concern about truth, so the idea of the fourth defense of minimal change, would not run into the difficulties described above.

The concern with truth is to be welcomed, no doubt, but unfortunately it does not offer an escape from the predicament we have come across. First of all, Observation 1 of course remains valid in a version that replaces each belief set by the set of *true* beliefs that it contains.

Corollary 4. Let K' and K'' be two belief-contravening candidate revisions of a consistent and logically closed belief set K by a sentence ϕ. If K' and K'' have different sets of true beliefs, they cannot be set-theoretically compared in terms of the true beliefs on which they differ with K.

But then, looking at symmetric differences does not seem to be what we are interested in, since in general we would not mind, but welcome *many* more true beliefs than we had prior to the revision. The dyadic notion of minimal change (referring to a relation between old and new belief set) is now dominated by the monadic notion of truth (as an aim of the posterior belief set). What we want to make sure is really that we minimize the loss of true beliefs. However, this does not lead to a satisfactory result either, as is borne out by

Observation 5. Let K' be a belief-contravening candidate revision of a logically closed belief set K by a sentence ϕ. Then K' minimizes the set of true beliefs lost from K (amongst all other candidate revisions of K by ϕ) only if K' is *opinionated* in the sense that it contains either ψ or $\neg\psi$, for every sentence ψ.

It is clear that the commitment to opinionated theories is undesirable, especially since the prior theory may well be undecided about countless contingent matters.[43] So minimizing the loss of true beliefs cannot be *the* only criterion relevant for our purposes. It is worth pointing out explicitly that AGM themselves have always been decidedly against the unbridled use of informational economy as it manifests itself in the inflationary behavior of so-called (maxi-)choice contractions and revisions of logically closed belief sets.[44]

I conclude that the argument to support the idea of minimal change through a relativization to true beliefs does not succeed.

As a little digression, we take down the following, perhaps even more surprising observation.

Observation 6. No belief-contravening candidate revision that does not contain every truth strictly enlarges the set of true beliefs. In particular, even if the prior theory has been false while the new piece of information as well as everything else in the posterior theory is true, the set of truths is not strictly increased.

It turns out that except for God (and perhaps various demons who can jump to the whole truth in a single step) there is no one-way road to the truth.[45] For every newly acquired truth we have to pay a price by trading in other truths. We may hope that these truths are less important or fundamental than the ones we acquire, but this is an issue that cannot be settled in an apriori manner.

IV

Getting clear about foundations is not just of theoretical interest, but has a practical effect to it as well: We expect to learn in which direction we are to head if we wish to improve our current theories of belief change, as well as their application in knowledge-based systems. The main point of this paper is that the theories developed by Alchourrón, Gärdenfors and Makinson and their followers are *not* oriented toward the principle of informational economy, and I have found no reason why they should. I have called two versions of this principle dogmas because many researchers (including the present author) seemed to believe in it and recited it time and again, without actually keeping to it when building their theories.

We have seen that contrary to many pronouncements of belief change theorists, it is even difficult to spell out what exactly is meant by the idea of informational economy or minimum utilation, as long as belief change is to satisfy the basic requirements. Theories of belief change in the tradition of Alchourrón, Gärdenfors and Makinson do not align themselves easily with the idea of minimal change, and they are by no means centered on this idea. I have tried to show that two of the most natural ways of fleshing out the idea do not achieve what they are widely supposed to achieve, and that four attempts to defend minimal change against my attack fail.

I should like to stress, however, that I have only been concerned with principles that are supposed to give a *description* of what motivates an important class of existing normative theories about belief change. I have refrained from putting forward any *normative* thesis about belief change myself, neither against nor in favor of conservatism. This is an entirely different undertaking.[46]

Towards the end of the paper, another point has come up. Little work—perhaps no work at all—has been done that reflects Quine's criteria (a) and (b) in the theory of belief revision. In his joint book *The Web of Belief* with J.S. Ullian,[47] Quine added more virtues that good theories should have: modesty, generality, refutability, and precision. Again, belief revision as studied so far has little to offer that would reflect these intuitive desiderata. Except for the issue of conservatism, Quine's list is a list for theory *choice* rather than for theory *change*, because it lists properties that a good posterior theory should have, independently of any properties of the prior theory. It is a strange coincidence that the philosophy of science has focussed on monadic (non-relational) features of theory choice, while philosophical logic has emphasized dyadic (relational) features of theory change. I believe that it is time for researchers in both fields to overcome this separation and work together on a more comprehensive picture.

With the present paper, however, I first of all hope to draw attention to the fact that it is not appropriate to exclusively focus on the idea of informational economy even when talking about nothing but the restricted, AGM-style modelings of belief change. There are various coherence criteria that find expression in formal "rationality postulates": inferential coherence (consistency and closure), dispositional coherence (a kind of semantic representability syntactically encoded in so-called "supplementary" rationality postulates—which are what I would call the hallmark of the AGM theory), as well as temporal coherence (a principled appreciation of the recency of information in iterated belief change). These criteria substantially restrict the dominion of informational economy in formal belief change theories. Having understood this, we may expect that the picture will change again enormously when we set out to live up to the Quinean epistemological virtues. It is time to liberate our minds from the constrictions of informational economy and give conscious admittance to other norms for our ethics of belief.

Appendix: Proofs

Observation 1. 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 beliefs on which they differ with the original belief set.

Proof of Observation 1. Let $\neg\phi$ be in a consistent belief set K, and let K' and K'' be two distinct candidate revisions of K with respect to ϕ. By the postulates of Closure and Success, we know that both K' and K'' are logically closed and contain ϕ.

Using the Δ-notation explained in footnote 21, we want to show that there is a sentence which is in $K\Delta K'$ but not in $K\Delta K''$ and there is also a sentence that is in $K\Delta K''$ but not in $K\Delta K'$.

Since $K' \neq K''$ there is either a sentence in $K' - K''$ or there is a sentence in $K'' - K'$. Without loss of generality, assume the latter, and take some sentence ψ that is contained in K'' but not in K'. Then, by Closure, $\neg\phi \vee \psi$ is in K and in K'' but not in K'. Hence $\neg\phi \vee \psi$ is in $K\Delta K'$ but not in $K\Delta K''$. On the other hand, again by Closure, $\phi \wedge \psi$ is in K'' but neither in K nor in K' (here we also use the consistency of K). Hence $\phi \wedge \psi$ is in $K\Delta K''$ but not in $K\Delta K'$. We conclude that $K\Delta K'$ and $K\Delta K''$ are not related by subset inclusion. *QED*

Observation 2. Suppose we want to revise a belief set by a sentence ϕ and find two elements of the belief set that non-redundantly entail the negation of ϕ. Then it may well be rational, according to the standard belief revision constructions, to restore consistency by removing the *more* entrenched and retain the less entrenched belief. In fact, such a situation can always be identified in an anamnestic revision by a consistent and moderately surprising sentence.

Proof of Observation 2. First we give a simple example of a situation with the characteristics depicted above. Consider $K = Cn(\neg p, q)$ and the following sequence of sentences of decreasing logical strength: $\bot$, $\neg p \wedge q$, $\neg p$, $p \supset q$ and $\top$. Define the ordering $\leq$ by putting $\phi \leq \psi$ iff every sentence in the sequence which implies ϕ also implies ψ.

An illustration of this situation in terms of Grovean systems of spheres (cf. Grove, *op. cit.*) is given in Figure 1 in the main text.

It is easy to check that $\leq$ thus defined is a standard entrenchment relation in the sense of Gärdenfors and Makinson, *op. cit.* Let $<$ be the asymmetric part of $\leq$. By the construction recipe for entrenchment-based revisions, we get $K * p = Cn(\{\phi \in K : \neg p < \phi\} \cup \{p\})$ which equals $Cn(p, q)$. Evidently q and $q \supset \neg p$ form a non-redundant set that entails $\neg p$ ("a minimal culprit set for $\neg p$"), and q is strictly less entrenched than $q \supset \neg p$. And yet q is, but $q \supset \neg p$ is not, contained in $K * p$.

For the general part of Observation 2, let ϕ be a consistent and moderately surprising sentence, and let $K * \phi$ be unequal to $Cn(\phi)$. The former means that there is an α in K such that $\alpha < \neg\phi$, where $<$ is the asymmetric part of the entrenchment relation associated with K. Take such an α, and let β be an element of $K * \phi$ which is not implied by ϕ. We need to find two beliefs ψ and χ which jointly, but not individually imply $\neg\phi$, and are such that $\psi < \chi$ but ψ is maintained while χ is abandoned in $K * \phi$.

Consider

$$\begin{aligned} \psi &:= (\neg\phi \wedge \alpha) \vee (\phi \wedge \beta) \\ \chi &:= \neg\phi \vee \neg\beta \end{aligned}$$

We check that these sentences ψ and χ indeed have the desired properties (for an illustration in terms of Grove spheres, compare Figure 2).

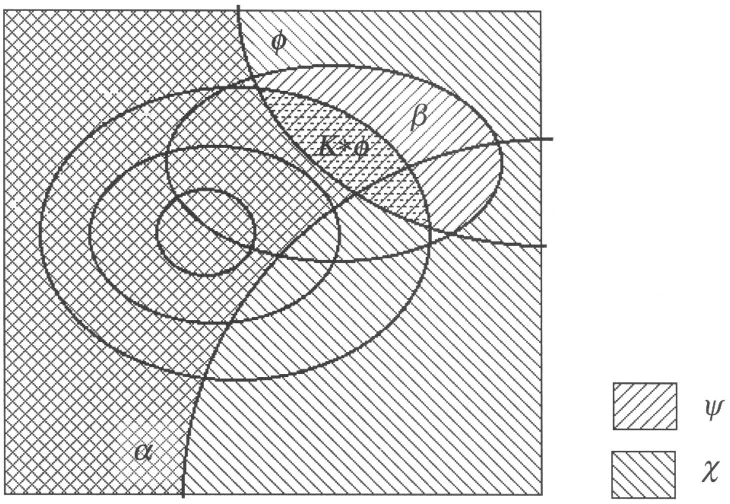

Figure 2: Illustration of the proof of the general part of Obs. 2

First, $\psi \wedge \chi$ clearly implies $\neg\phi$. Second, ψ alone does not imply $\neg\phi$, since the sentences ϕ and β are both elements of the consistent set $K * \phi$ (applying AGM's postulate that consistency can never be occasioned by a consistent input ϕ), and their conjunction $\phi \wedge \beta$ therefore is consistent. Third, χ alone does not imply $\neg\phi$, since ϕ does not imply β and therefore $\neg\beta$ does not imply $\neg\phi$. Fourth, ψ is in $K * \phi$ since both ϕ and β are in $K * \phi$, and $K * \phi$ is closed under logical consequence (using the AGM postulates of Success and Closure). Fifth, χ is not in $K * \phi$, since $\phi \wedge \beta$ is in $K * \phi$ and $K * \phi$ is consistent (again assuming Success and Consistency Preservation). Finally, it remains to show that $\psi < \chi$. (In the following I refer to the convenient axiomatization of entrenchment relations given in Rott, "Preferential Belief Change.") From the irreflexivity of $<$, we get that $\alpha \not< \alpha$. Since $\neg\phi \wedge ((\neg\phi \wedge \alpha) \vee (\phi \wedge \beta))$ implies α, we get by Continuing Up that $\alpha \not< \neg\phi \wedge ((\neg\phi \wedge \alpha) \vee (\phi \wedge \beta))$. From this and $\alpha < \neg\phi$, we get by Conjunction Up that $\alpha \not< (\neg\phi \wedge \alpha) \vee (\phi \wedge \beta)$. Making use of the fact that $\alpha < \neg\phi$ once more, we deduce from this with the help of Virtual Connectivity (a property characteristic of Gärdenfors-Makinson entrenchments) that $(\neg\phi \wedge \alpha) \vee (\phi \wedge \beta) < \neg\phi$. By Continuing Up, we finally get $(\neg\phi \wedge \alpha) \vee (\phi \wedge \beta) < \neg\phi \vee \neg\beta$, that is, $\psi < \chi$, as desired. (Notice that this proof does not depend on a particular construction recipe for entrenchment-based belief changes.) *QED*

Observation 5. Let K' be a belief-contravening candidate revision of a logically closed belief set K by a sentence ϕ. Then K' minimizes the set of true beliefs lost from K (amongst all other candidate revisions of K by ϕ) only if K' is opinionated in the sense that it contains either ψ or $\neg\psi$, for every sentence ψ.

Proof of Observation 5. Let the belief set K be consistent and logically closed, let $\neg\phi$ in K, and let K' be a a candidate revision containing ϕ which is not opinionated. We show that K' does not minimize the set of lost true beliefs from K.

Assume first that ϕ is true. Then the set T of all true sentences is a candidate revision containing ϕ. Take a true sentence ψ that is not contained in K'; there is such a ψ since K' is not opinionated. Then, by logical closure, $\neg\phi \vee \psi$ is a true sentence that was contained in the prior belief set K and is lost in the candidate belief set K'. But $\neg\phi \vee \psi$ is in T, and T loses no true belief from K at all, so T actually loses less true beliefs from

K than K'. Thus K' does not minimize the set of lost true beliefs from K.

Assume secondly that ϕ is false. Then $\neg\phi$ is true. Extend K' to an opinionated set K'', and take some ψ from $K'' - K'$. Then, by logical closure, $\neg\phi \vee \psi$ is a true sentence from K which is lost in K' but not in K''. Since K' is a subset of K'', K'' loses no true sentence from K that K' doesn't lose. So K'' actually loses less true beliefs from K than K'. Thus K' does not minimize the set of lost true beliefs from K. *QED*

Observation 6. No belief-contravening candidate revision that does not contain every truth strictly enlarges the set of true beliefs. In particular, even if the prior theory has been false while the new piece of information as well as everything else in the posterior theory is true, the set of truths is not strictly increased.

Proof of Observation 6. Let $\neg\phi$ be in the belief set K, and let K' be a candidate revision of K with respect to ϕ. By the postulates of Closure and Success, we know that K' is logically closed and contains ϕ.

Now we let K_t and K'_t be the set of true beliefs in K and K' respectively. What we want to show that K'_t is not a strict superset of K_t, i.e., that there is a true sentence which is in K but is lost in K'.

Take some true sentence ψ which is not in K'. Such a sentence exists since K' was assumed not to be omniscient. Now consider $\neg\phi \vee \psi$. This sentence is true, since ψ is true, and it is in K, since $\neg\phi$ is in K and K is logically closed. However, $\neg\phi \vee \psi$ is not in K' since ϕ is in K' and ψ is not in K'. Thus we have found a truth that has been lost.

The second part of Observation 6 follows from the first. *QED*

Notes

[1] Willard Van Orman Quine, "Two dogmas of empiricism," *Philosophical Review*, LX (1951): 20–43. Reprinted in Quine, *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1953), pp. 20-46.

[2] See Carlos Alchourrón, Peter Gärdenfors and David Makinson, "On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions," *Journal of Symbolic Logic*, L (1985): 510–530; Peter Gärdenfors, *Knowledge in Flux. Modeling the Dynamics of Epistemic States* (Cambridge, Mass.: Bradford Books, MIT Press, 1988); Peter Gärdenfors and Hans Rott, "Belief revision," in D. M. Gabbay, C. J. Hogger, and J. A. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming Volume IV: Epistemic and Temporal Reasoning* (Oxford: Oxford University Press, 1995), pp. 35–132; and Sven Ove Hansson, *A Textbook of Belief Dynamics: Theory Change and Database Updating* (Dordrecht: Kluwer Academic Publishers, 1999).

[3] Thanks to Isaac Levi for raising this point.

[4] See in particular Sven Ove Hansson, ed., Special Issue on Non-Prioritized Belief Revision, *Theoria*, LXIII (1999): 1–134.

[5] In AGM theory the success condition has precedence over the consistency condition, since a revision by a contradiction leads to an inconsistent belief set. But this seems to be a contingent rule for a logical limiting case rather than a matter of considered philosophical decision.

[6] David Makinson, "How to Give it Up: A Survey of Some Formal Aspects of the Logic of Theory Change," *Synthese*, LXII (1985): 347–363, here: p. 352.

[7] *Op. cit.*, p. 53.

[8] *Op. cit.*, p. 38.

[9] André Fuhrmann, *An Essay on Contraction* (Stanford: CSLI Publications, 1997), p. 17.

[10] Adnan Darwiche and Judea Pearl, "On the Logic of Iterated Belief Revision," *Artificial Intelligence*, LXXXIX (1997): 1–29, here: p. 2.

[11] *Op. cit.*, p. 87.

[12] Peter Gärdenfors and David Makinson, "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, 1988), pp. 83–95, here: p. 88.

[13] *Op. cit.*, p. 38. Similarly Gärdenfors, "Belief Revision: An Introduction," in Peter Gärdenfors, ed., *Belief Revision* (Cambridge: Cambridge University Press, 1992), pp. 1–28, especially pp. 9 and 17.

[14] *Op. cit.*, p. 24.

[15] Craig Boutilier, "Iterated Revision and Minimal Change of Conditional Beliefs," *Journal of Philosophical Logic*, XXV (1996): 263–305, here: pp. 264–265.

[16] As, by the way, Quine, *op. cit.*, Section 5, claimed the two dogmas of empiricism were.

[17] Introduced by Alchourrón, Gärdenfors and Makinson, *op. cit.*

[18] Introduced by Gärdenfors and Makinson, *op. cit.*

[19] The relevant result is presented in Section 4 of Hans Rott, "Two Methods of Constructing Contractions and Revisions of Knowledge Systems", *Journal of Philosophical Logic*, XX (1991): 149–173. I shall return to this topic in subsection III.B and argue that the interpretation just given is incorrect. Criteria (1) and (2) may even be pulling in opposite directions; compare footnote 35.

[20] The proofs of all observations but one are given in the appendix to this paper. Observation 1 bears some similarity with Miller's and Tichý's celebrated refutations of Popper's early concept of verisimilitude; see David Miller, "Popper's Qualitative Theory of Verisimilitude," *British Journal for the Philosophy of Science*, XXV (1974): 166–177, and Pavel Tichý, "On Popper's Definition of Verisimilitude," *British Journal for the Philosophy of Science*, XXV (1974): 155–160. While Popper, Miller and Tichý are interested in the distances of (false) theories from the theory containing all and only the truths, we are interested here in the distances of (belief-contraveningly revised) candidate theories from a given prior theory.

[21] For any two sets X and Y, the symmetric difference $X \Delta Y$ between X and Y is defined to be $(X \setminus Y) \cup (Y \setminus X)$.

[22] This is a consequence of Observation 2.2 in Carlos Alchourrón and David Makinson, "On the Logic of Theory Change: Contraction Functions and Their Associated Revision Functions," *Theoria*, XLVIII (1982): 14–37.

[23] For official formal definitions, see Gärdenfors and Makinson, *op. cit.*, p. 89, and Rott, "Preferential Belief Change Using Generalized Epistemic Entrenchment," *Journal of Logic, Language and Information*, I (1992): 45–78, here: p. 61.

[24] What *is* excluded by the definition of entrenchment advocated in Gärdenfors and Makinson, *op. cit.*, and generalized in Rott, *op. cit.*, is that the situation of Observation 2 arises when the two "culprits" ψ

and χ, say, *exactly* entail $\neg\phi$, that is, when $\psi \wedge \chi$ is logically equivalent with, but not logically stronger than $\neg\phi$. For then we have, by definition, $\psi < \chi$ if and only if χ is, but ψ is not, in the contraction of the belief set with respect to $\neg\phi$, and therefore χ is, but ψ is not, in the revision of the original belief set with respect to ϕ.

[25] Gärdenfors and Makinson, *Op. cit.*

[26] See Adam Grove, "Two Modellings for Theory Change," *Journal of Philosophical Logic*, XVII (1988): 157–170, and Gärdenfors, *Knowledge in Flux*, pp. 83–86, 94–97.

[27] Here is an objection to this argument from Observation 2: Observation 2 uses an entrenchment relation not in the GM way, but in the way in which Alchourrón and Makinson, "On the Logic of Theory Change: Safe Contraction," *Studia Logica*, XLIV (1985): 405–422, use hierarchies in so-called *safe contractions*. And it is no wonder that such a misapplication should lead to unexpected results.— Rejoinder: Theorem 4(ii) in Rott, "On the Logic of Theory Change: More Maps Between Different Kinds of Contraction Function," in Peter Gärdenfors, ed., *Belief Revision* (Cambridge: Cambridge University Press, 1992), pp. 122–141, shows that entrenchment relations *can* be used for constructing safe contractions and lead to exactly the same results as when applied in the "proper" GM way. But how can this be? In the example just discussed, wouldn't a safe contraction clearly eliminate q and keep $q \supset \neg p$? No. We can read off from Fig. 1 that the belief $q \supset \neg p$ gets lost since it is the minimal element in another entailment set (viz., $\{p \supset q, q \supset \neg p\}$); and q—while eliminated as the minimal element of the set $\{q, q \supset \neg p\}$—is ultimately rederived from the "safe" elements $p \supset q$ and $p \vee q$.

[28] One could try to find a more "global" formulation of the idea. – An anonymous referee pointed out that the following idea of *constrained minimization* might help. Although the example supporting Observation 2 shows that entrenchment minimization *per se* does not serve our purpose, minimization *subject to the constraint that* p *be imported* makes sense, that is, minimization with respect to the relation "q is better entrenched than $q \supset \neg p$ among revisions that import p." Translated to the modeling at hand, this idea might mean two different things. First, reference may be made to the posterior entrenchments of the sentences in questions. This is essentially equivalent to a question about iterated belief change which is something cannot treat here in appropriate generality (but compare Section III.C below). Second, one may relativize the entrenchments by using material conditionals with antecedent p in the comparison themselves. This amounts to a very special method of iterated belief change, so-called "irrevocable belief change" (see Rott, "Two Methods", *op. cit.*, Section 6, and Krister Segerberg,

"Irrevocable Belief Revision in Dynamic Doxastic Logic", to appear in the *Notre Dame Journal of Formal Logic*). According to this proposal, the relevant question is not whether q is more entrenched than $q \supset \neg p$, but rather whether $p \supset q$ is more entrenched than $p \supset (q \supset \neg p)$. It can be verified that this is true in our example, and that indeed the problem mentioned in Observation 2 can never arise if this sort of relativization is employed. But the idea, interesting though it is, departs too much from the wording of (2) to count as a faithful formalization of the principle. It is not posterior entrenchment but prior entrenchment that is supposed to govern changes of belief.

[29] A *candidate contraction* of a belief set K with respect to a sentence ϕ is a subset of K which is logically closed and does not contain ϕ.

[30] In terms of the Grovean possible worlds model, recovery disallows the gratuitous admission of ϕ-worlds in a contraction with respect to ϕ. It says nothing at all about the admission of $\neg\phi$-worlds.— Compare Maurice Pagnucco and Hans Rott, "Severe Withdrawal (and Recovery)," *Journal of Philosophical Logic*, XXVIII (1999): 501–547.

[31] In analogy with the tradition of rational choice theory associated with leading economists like Kenneth Arrow and Amartya Sen, one could advance the slogan "rational change is relational change." In rational choice theory, the corresponding slogan is "rational choice is relational choice." Belief change can accordingly be incorporated as a subtheme into the study of the *homo oeconomicus*; see Rott, *Making Up One's Mind: Foundations, Coherence, Nonmonotonicity* (Habilitationsschrift, Department of Philosophy University of Konstanz, October 1996, revised version to appear with Oxford University Press under the title "Change, Choice and Inference"), and Rott, "Logic and Choice," in Itzhak Gilboa, ed., *Theoretical Aspects of Rationality and Knowledge – TARK 1998*, (San Francisco: Morgan Kaufmann, 1998), pp. 235–248. Gärdenfors-Makinson style entrenchment relations can be conceived as revealed preferences in precisely the sense that is common in rational choice theory.

[32] Compare Sven Ove Hansson, "Hidden Structures of Belief," in André Fuhrmann and Hans Rott, eds., *Logic, Action and Information* (Berlin: de Gruyter, 1995), pp. 79–100. For the classic accounts of the 1980s, see Gärdenfors, *Knowledge in Flux*, Chapters 3 and 4.

[33] The AGM postulates for belief revision include the above-mentioned "basic" requirements of success, logical closure, consistency preservation, as well as a postulate to the effect that it is the content of a new piece of information that matters, not its syntactical formulation. Moreover, AGM have two postulates identifying a revision by a sentence consistent with the prior belief set with the deductive closure of the

union of both. (These two postulates are *the only* AGM postulates for revision that lend themselves to a minimal-change interpretation.) Finally, two "supplementary" postulates of AGM relate revisions by varying inputs. For a more thorough-going interpretation of these postulates, see Rott, "Coherence and Conservatism in the Dynamics of Belief. Part I: Finding the Right Framework," *Erkenntnis*, L (1999): 387–412.

[34] The official Gärdenfors-Makinson recipe for belief contractions based on entrenchments makes use of a disjunction and therefore resists an easy understanding: ψ is in the contraction of K with respect to ϕ if ψ is in K and $\phi \vee \psi$ is more entrenched than ϕ. See Gärdenfors and Makinson, *op. cit.*

[35] See the *principles of preference and indifference* discussed in Pagnucco and Rott, *op. cit.* These principles tell us to go for a compromise if there are multiple optimal solutions, and are thus responsible for the fact that (2) pulls in a different direction than the uncompromising minimal change principle (1). Further relevant arguments are given by Tor Sandqvist, "On Why the Best Should Always Meet," to appear in *Economics and Philosophy*, XVI (October 2000).

[36] See Craig Boutilier, "Revision Sequences and Nested Conditionals," in R. Bajcsy, ed., *IJCAI-93 – Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1993, pp. 519–525, and Boutilier, "Iterated Revision and Minimal Change of Conditional Beliefs," as well as Rott, "Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence," manuscript, ILLC, University of Amsterdam, May 1998.

[37] Modulo the principle of indifference mentioned in footnote 35.

[38] See Adnan Darwiche and Judea Pearl, "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: Morgan Kaufmann, Cal., 1994), pp. 5–23, here: pp. 9 and 19; Darwiche and Pearl, "On the Logic of Iterated Belief Revision," *Artificial Intelligence*, LXXXIX (1997): 1–29, here: p. 10 and 21; and Boutilier, "Iterated Revision and Minimal Change of Conditional Beliefs," *op. cit.*, p. 296.

[39] Taken from Rott, "Coherence and Conservatism in the Dynamics of Belief, Part II."

[40] The input is accepted, but, in the terminology introduced above, its negation is considered to be not even moderately surprising.

[41] Symptomatic of a certain lack of interest in *truth* among belief change theorists is the fact that the most important book on *belief* revision has the title "*Knowledge* in Flux" (Gärdenfors, *op. cit.*) but does not attend to the truth of our alleged "knowledge" (nor to its justification, for that matter).

The same terminological unconcern is present in the commonly used term '*epistemic* entrenchment' which should really be '*doxastic* entrenchment'—unless one does not mind disregarding more than two millennia of philosophical tradition (*vide* Plato). That belief revision today is finally getting closer to the concept of truth can be seen from the recent work of Kevin Kelly and his colleagues. See Kevin Kelly, Oliver Schulte and Vincent Hendricks, "Reliable Belief Revision," in Maria Luisa Dalla Chiara et al., eds., *Logic and Scientific Methods – Proceedings of the 10th International Congress of Logic, Methodology and Philosophy of Science* (Dordrecht: Kluwer, 1997), and Kevin Kelly, "Iterated Belief Revision, Reliability, and Inductive Amnesia," *Erkenntnis*, L (1999): 11–58.

[42] William James, "The Will to Believe," in F.H. Burkhardt, F. Bowers and I.K. Skrupskelis, eds., *The Will to Believe and Other Essays in Popular Philosophy – The Works of William James, Vol. 6* (Cambridge, Mass., and London: Harvard University Press, 1979), pp. 13–33. – Incidentally, one decade later, in 1907, James takes the "observable process ... by which any individual settles into *new opinions*" to show that "in matter of belief we are all extreme conservatives" and that we invariably aim at a "minimum of disturbance" (or "minimum of modification", or "minimum of jolt") of the older stock of beliefs (*Pragmatism – The Works of William James, Vol. 1*, eds. F. Bowers and I.K. Skrupskelis, Cambridge, Mass., and London: Harvard University Press, 1975, pp. 34–35). James says that "loyalty" to older truths is not only the first, but often the only principle for belief change. Although this is formulated as a descriptive statement about empirical believers, it is quite evident that James approved of "extreme" conservatism as a normative doctrine.

[43] Observation 5 stands to Observation 3.2 of Alchourrón and Makinson, "On the Logic of Theory Change: Contraction Functions and Their Associated Revision Functions," essentially as Corollary 4 stands to Observation 1. However, an independent proof of Observation 5 is given in the Appendix for the sake of transparency.

[44] Compare the discussion in Alchourrón and Makinson, *op. cit.*, pp. 20–21, and in Makinson, *op. cit.*, pp. 356–359.

[45] James seems to have felt that such problems may arise: "... it may indeed happen that when we believe the truth A, we escape as an incidental consequence from believing the falsehood B .... We may in escaping B fall into believing other falsehoods, C or D, just as bad as B..." ("The Will to Believe", p. 24)

[46] See especially Gilbert Harman, *Change in View* (Cambridge, Mass.: Bradford Books, MIT Press,

1986) and "Rationality", in E.E. Smith and D.N. Osherson, eds., *Thinking: Invitation to Cognitive Science*, Vol. III (Cambridge, Mass.: MIT Press, 1995), pp. 175–211; reprinted in Gilbert Harman, *Reasoning, Meaning and Mind* (Oxford, Clarendon Press, 1999), pp. 9–45. A good critical discussion of conservatism in contemporary normative epistemology is given by David Christensen, "Conservatism in Epistemology," *Noûs*, XXVIII (1994): 69–89.

[47] Willard Van Orman Quine and Joseph S. Ullian: *The Web of Belief* (second edition, New York: Random House, 1978).

HANS ROTT
Institute for Philosophy
University of Regensburg
93040 Regensburg
Germany
hans.rott@psk.uni-regensburg.de

Additional footnote for page 10:

Consider, for instance, the revision of K by ϕ constructed by means of a contraction of K with respect to $\neg\phi$ and a subsequent expansion with respect to ϕ (the Levi identity). One may choose the contraction to be minimal by taking a maximal subset K' of K that does not imply $\neg\phi$, and then choose to expand K' with minimal change by taking $Cn(K' \cup \{\phi\})$. Unfortunately, the two kinds of minimality do combine into a suitable measure of minimal change for the revision. The result will be just some maximal consistent set of sentences including ϕ, and this set is qualitatively very different from an ordinary belief set because it is opinionated about everything. (Thanks to Eduardo Fermé for drawing my attention to this point.)

Observation relating to footnote 28 on page 27:

Claim. Whenever $\{\psi, \chi\} \vdash \neg\phi$, it holds that

$$\psi \notin K * \phi \text{ and } \chi \in K * \phi \quad \text{iff} \quad \phi \supset \psi < \phi \supset \chi$$

Proof. Let $\{\psi, \chi\} \vdash \neg\phi$. By the AGM conditions for entrenchment-based revision (compare, for instance, Hans Rott, "A Nonmonotonic Conditional Logic for Belief Revision I", in André Fuhrmann und Michael Morreau, eds., *The Logic of Theory Change*, Lecture Notes in Artificial Intelligence **465**, (Berlin: Springer 1991), pp. 135–183, Observation 1), the LHS of the claim holds if and only if

$$(\nvdash \neg\phi \text{ and } \phi \supset \psi \leq \neg\phi) \;\text{ and }\; (\vdash \neg\phi \text{ or } \neg\phi < \phi \supset \chi)$$

By the transitivity of entrenchments, this conditions implies $\phi \supset \psi < \phi \supset \chi$. In order to show that the converse implication is also valid, suppose that $\phi \supset \psi < \phi \supset \chi$. By the Dominance condition for entrenchments, we know that $\neg\phi \leq \phi \supset \psi$. Hence, by transitivity, $\neg\phi < \phi \supset \chi$, and $\neg\phi$ is not a logical truth, by the maximality condition for entrenchment relations. It remains to show that $\phi \supset \psi \leq \neg\phi$. Suppose for reductio that $\neg\phi < \phi \supset \psi$. Then, by the entrenchment properties, we get $\neg\phi < (\phi \supset \psi) \wedge (\phi \supset \chi)$. Since we have the assumption that $\{\psi, \chi\} \vdash \neg\phi$, the sentence $(\phi \supset \psi) \wedge (\phi \supset \chi)$ is logically equivalent with $\neg\phi$, so the extensionality of entrenchments gives us $\neg\phi < \neg\phi$, contradicting the irreflexivity of $<$.