On Having Bad Contractions, or: No Room for Recovery*

Neil Tennant

The Ohio State University

July 10, 2004

Abstract

The well-known *AGM*-theory of theory-contraction and theory-revision, due to Alchourrón, Gärdenfors and Makinson, relies heavily on the so-called postulate of *recovery*. This postulate is *supposed* to capture the requirement of 'minimum mutilation'; but it does not. Recovery can be satisfied even when there is more mutilation than is necessary. Recovery also ensures that very often *too little* is given up in a contraction. In this paper I bring out clearly the deficiencies of the *AGM*-theory in these two regards, showing how it is doubly off-beam. I show that some of the most serious inadequacies of the *AGM*-theory derive from early claims in some of its founding contributions, claims that have not been seriously questioned within the tradition since. The upshot of these investigations is that recovery cannot, and should not, be recovered. Theory contraction is hysteretic. Whether the *AGM*-theory can now recover is a good question.

1 Knowledge- and belief-sets versus logicians' 'theories'

Ramanujan, the great mathematician, claimed that an angel visited him in his sleep at nights, and wrote upon a tablet various theorems that Ramanujan ought to prove. Ramanujan would dutifully prove them; and his

*This paper appeared in *Journal of Applied Non-classical Logics*, 7, 1997, pp. 241-266. This form corrects a bad typo—a deleted backslash in the LaTeX source code of the published version—which deformed one of the proofs. Otherwise, it is exactly as published.

contribution to the expansion of our knowledge in mathematical analysis was considerable.

Expanding a set of known propositions involves providing justification for believing some new proposition to be true. The proposition is 'new' in the sense that it has not heretofore been proved. The first principles from which such proofs might proceed will already be consciously and explicitly known. Thus there is an everyday sense in which we are willing to talk of 'expanding' our knowledge even when the 'expansion' involves proving something that follows logically from propositions already known to be true.

This is a clue to the inadequacy of the view that a knowledge- or belief-set (that is, a set of propositions known or believed) has to be thought of as *deductively closed*. Typically, such sets are *not* deductively closed; and that is why we are willing to talk of 'expansion' (conceived of as *proper* expansion) when we prove a new result, albeit from principles already 'known' (or 'believed') and thus already in the set before the expansion in question.

Our knowledge- and belief-sets are inadequately modeled as deductively closed sets of sentences — that is, as what logicians standardly refer to as 'theories'. But even taking into account their lack of logical closure, our knowledge- and belief-*sets* are not good representations of our knowledge- and belief-*systems*. For our knowledge- and belief-*systems* consist, typically, of *finitely many* propositions, or proposition-schemata, *along with 'justifications'* that are articulated to varying degrees. The belief-*set* could be called the *core* of the belief-*system*; it represents what propositions are 'in' the system, and therefore believed. The belief-*system* goes *beyond* this core — it represents in addition *how* each proposition came to be believed; that is, what the believer's *reasons* are for believing it. In order for me to maintain a form of conversational continuity with other writers' contributions on the topic of theory contraction, I have to be careful to talk about belief-*sets* when contrasting certain ideas of mine with ideas of theirs. On other occasions, when I venture into new territory with considerations relating exclusively to *systems*, I can stress the difference by using the appropriate terminology.

The first time that a 'new' theorem of a mathematical theory is proved, the sole justification for it will be the new proof offered. Mathematicians usually succeed in providing yet further, different proofs of a theorem already proved. They might seek to constrain further the first principles appealed to in the proof, eliminating, if possible, 'unnecessary' ones. They might also seek different deductive routes to the same theorem, ones which (when codified as proofs) reveal more clearly why the result holds, or which lend themselves to generalization in the solution of whole classes of related conjectures. The upshot, then, is that a typical 'item' in a knowledge- or belief-*system* is a (known or believed) proposition furnished with possibly more than one — but still only *finitely many* — *justifications*, or *epistemic pedigrees*, or *proofs*.

2 Bases, developments and theories

Let us return for the time being just to sets of sentences, and ignore the need to take into account the justificatory pedigrees available for each sentence in a knowledge- or belief-set. We still need to distinguish three things: *bases*; what I shall call *developments*; and *theories*. The phrase 'belief-set' is ambiguous among all three of these.

A *theory* is a logically closed set of sentences: that is, a set of sentences that contains every logical consequence any of its subsets. Obviously any theory is infinite, and will contain logic, which can be thought of as consisting of the logical consequences of the empty set. I shall assume throughout that completeness holds, that is, that the semantic notion of logical consequence and the proof-based notion of deducibility coincide in extension; and I shall use '$\vdash$' to stand for this relation, whether it be conceived model-theoretically or proof-theoretically. Thus $X \vdash x$ will mean that the sentence x is deducible from (is a logical consequence of) the set X of sentences. When X is empty, we shall simply write $\vdash x$.

A *base* is a set of axioms (finite or infinite) for a theory, and is (almost always) not itself logically closed. A theory is the logical closure of any base for it. When A is the logical closure of a set B of sentences I shall write $A = [B]$. In particular, A is a theory if and only if $A = [A]$. The set of logical truths is just $[\emptyset]$.

A base B is called *irredundant* just in case removing any one of its members results in the generation of a different theory: that is, if and only if

$$\forall b \in B\ [B \setminus b] \neq [B].$$

One can have different bases for one and the same theory; indeed, one can have different *irredundant* bases for one and the same theory. For take the two irredundant bases $\{a,\ b\}$ and $\{a,\ a \to b\}$. From the first of these bases one can deduce $a \to b$ by means of one step of $\to$I; and from the second one can deduce b by means of one step of $\to$E. In each case that yields the same *development* $\{a,\ b,\ a \rightarrow b\}$; and from there on everything will coincide, and the same unique theory will be generated.

This example serves also to illustrate what I shall mean by a *development*. In the case where the base is finite, a development will be any set midway between that base and its full logical closure. To be realistic, a development moreover is *always* a *finite* set:

$$\text{(finite) base } \{a,\ b\} \subset \text{ (finite) development } \{a,\ b,\ a \rightarrow b\}$$
$$\subset \text{ full (infinite) theory } [a,\ b]$$

The first containment is to be understood as a proper one by virtue of the presence, in the development, of at least one falsifiable non-member of the base in question. Merely extending a base by adding only logical truths does not produce a true development. In our example, what we added in order to produce the development was the sentence $a \rightarrow b$, which is both falsifiable and not itself in the base $\{a,\ b\}$.

When the base is infinite, however, the first containment above will not hold; for a development, we have stipulated, is always to be finite. In the case where the base is infinite, therefore, the development will be some partial logical closure of some finite subset of that base, by means of at least one falsifiable non-member thereof — that is, it will consist of finitely many members of the base plus finitely many sentences that they jointly imply, among which will be at least one falsifiable non-member of the base. A development in this case will not itself be a subset of the (infinite) base in question, but its logical closure will be a subset of the theory determined by that base.

Note that any development must, according to our definition, contain some redundancy; that is, no development can be an irredundant base. When a development is fleshed out with proofs to justify each sentence in the development, it becomes a *system development*. Moreover, a sensible view is that all system developments are *finite*, even if the base of available axioms is an infinite set. For at any stage of our history we shall only ever have produced finitely many proofs. And since the proofs themselves are finite, we shall have invoked only finitely many of these axioms.

3 Contractions

Back to Ramanujan — not the historical one, but rather a fictional one for expository purposes. Suppose Ramanujan is interested in theories and belief sets quite generally, and not just mathematical ones. (What I shall be saying, however, will apply to mathematical theories as well as to non-mathematical ones.) Ramanujan's angel now comes to him in his sleep holding a different kind of tablet. No longer does the angel write down what he wants Ramanujan to believe, and to find a reason for believing it. Rather, the angel writes down what he wants Ramanujan to *cease believing.* Indeed, when the angel's practice changed, he had an initial conversation with the dreaming Ramanujan about the new tack. The angel spake thus:

> In the past I have told you what propositions are worth believing. But from now on I shall tell you what propositions you should *no longer believe.* Do not, however, leap to any conclusions about their contradictories. I'm not telling you to *disbelieve* them; I just want you not to *believe* any proposition I may happen to write on the tablet. But, as far as each such proposition will be concerned, I want you to *keep an open mind.* For who knows — one day you might have to recruit back the beliefs I've asked you to let go. Or, you may have to take on some of their contradictories. But, for the time being, I can't say which. So do please keep an open mind. Obey my injunctions by giving up as *little* as possible. And if ever I instruct you to cease believing a logical truth, you can disregard that instruction. Take seriously only my proscriptions of propositions that are not logical truths!

Ramanujan has been assigned the general task of *contracting* a set of beliefs with respect to any given (contingent[1]) proposition implied by it.

Ramanujan falls to reflecting on how best to carry out the angel's commands. He knows the shape of the problem in general: at any stage he will have a set A of current beliefs. There will be only finitely many of them; but, *potentially*, at least, there is an infinite set of beliefs derivable from A. This is because the believer is rationally committed to any proposition that can be shown to be in the logical closure of A. (Among these, for example, will be all the logical truths in the language of A.) So any changes Ramanujan proposes to make to a belief-set will have ramifying consequences.

[1]A contingent proposition is one which is neither logically true nor logically false. Contracting with respect to a logical truth involves doing nothing. And the reason why contraction with respect to a logical falsehood is not explicitly considered is that one assumes an abiding commitment to consistency of belief anyway, in the form, perhaps, of the angelic proscription 'Do not believe $\bot$!'

Exactly which propositions are to be sacrificed from a theory in a rational process of contraction with respect to one of its consequences will in general depend on how the theory is articulated (including, among other considerations, what might be taken as its base) and what the 'sacrificial preferences' are of the theorist whose theory it is. Ramanujan, for example, may have no preference for a over b, and *vice versa*. But he may also have a very strong preference for the implication $a \rightarrow b$ over a. Take now our two earlier irredundant bases $\{a,\ b\}$ and $\{a,\ a \rightarrow b\}$, which, as we noted, generate the same theory (call it Θ).

If Ramanujan holds this theory Θ on the basis of $\{a,\ b\}$ and is told by the angel to give up the conjunction $a \wedge b$, he might do any one of the following:

1. give up both a and b and therefore 'go back to scratch', as it were, contracting all the way to the theory of the empty base (i.e. logic);

2. opt to give up a, even though he has no particular reason for preferring b to a — in which case the contracted theory will be $[b]$; or

3. opt to give up b, even though he has no particular reason for preferring a to b — in which case the contracted theory will be $[a]$.[2]

If, on the other hand, Ramanujan holds the theory Θ on the basis of the *other* base $\{a,\ a \rightarrow b\}$, then he should contract it with respect to $a \wedge b$ as follows:

> First, note that one has to give up one or other (possibly both) of a, b, in order to get rid of their conjunction $a \wedge b$. If, on the one hand, a is given up first, then there will no longer be any justification for b, since b had been derived by $\rightarrow$E from the two members a, $a \rightarrow b$ of the base; so b will be given up also. If, on the other hand, b is given up first, then, since it is jointly implied by the two members of the base $\{a,\ a \rightarrow b\}$ in question, one will have to give up one of them too. But, *ex hypothesi*, one's preference is for $a \rightarrow b$ over a; whence one will give up a. Thus one gives up both a and b regardless of which of these two one gives up first. The residue will therefore be $[a \rightarrow b]$.

[2] A more sophisticated 'look ahead' method might disincline one to plump for this option, since it involves the loss of $a \rightarrow b$, which we are supposing to be preferred to a. But here I am invoking considerations that do not fall within the ambit of any mathematical provisions by AGM-theory. Indeed, AGM-theory allows *only* option (1) under the hypotheses of the example (on their method of finite base contraction).

This makes contraction look rather intensional as a process, depending on more than just the sentence-membership of the theory (or belief-system) being contracted; and indeed it is. Each 'belief' α within the *system* A consists in the proposition a believed *along with* such justifications, or derivational pedigrees, as a might have acquired by that stage. The proposition a is the member of the core; while the item α 'dresses up' the proposition a further, with proofs, so that it can be an entry in the *system*. The general form of a justified item in the belief *system* is therefore

$$\alpha =< a, \{\Pi_1, \ldots, \Pi_n\} >$$

where the Π_i are the various justifications so far provided for a. A belief-*system* is more than a mere set of beliefs. The various justifications provided for each belief are partially constitutive of the system's identity. A system will have the set of beliefs as its *core*; but it will not be fully determined until all of the believer's *justifications* for beliefs in that core have been provided. If the system (or system-user) has bad memory, some of the fine internal structure of these Π_i might have been lost. In that case, one could retreat to a coarser kind of representation for system items:

$$\alpha =< a, \{\Delta_1, \ldots, \Delta_n\} >$$

where each Δ_i is the set of premisses (undischarged assumptions) of the erstwhile proof Π_i. Note that there is no requirement to the effect that every member of such a set Δ_i should itself be an *axiom*, or *first principle*, of the 'theory' of which A is a (necessarily finite and) partial development. Rather, all that is required is that any member of such Δ_i be itself justified within the system A. We do, characteristically, build upon our past successes — a process whereby a derived result can become the starting point for fresh inquiry. In logicians' parlance, a member of a set Δ_i within an entry of the form just displayed above could be a 'cut formula' (or lemma) for the ultimate proof, from first principles, of the corresponding result a.

It is an interesting question, and one worth exploring — but one which is unable to be posed on the AGM-approach — whether making do with the most recent lemmata (as members of Δ_i) adduced for the proof(s) of a, rather than tracing logical dependencies all the way back to those members of the base in question on which a ultimately rests, substantially affects the outcome(s) of the contraction process.[3] If it did, then this would be yet another manifestation of the intensionality of that process.

[3] I have in mind here the contraction process outlined in Tennant [1994].

Ramanujan knows that the angel could write on his tablet any proposition p, regardless of whether it features as an already justified item in the currently developed system A. If there is as yet no justification, within A, of the prohibited proposition p, then Ramanujan will have to live in constant fear and trembling of the prospect that his system might unexpectedly turn up such a would-be justification for p, contrary to the angel's wishes. Ramanujan will just have to be very vigilant to avert such a prospect; but, apart from wondering about what proofs of p there *might*, unknown to him (at present) be, there is precious little that he can *do*, for the time being, to carry out the angel's wish that he not believe p. If there *is* already a proof Π for p in A, however, so that A has an entry of the form

$$\pi = < p, \{\ldots, \Pi, \ldots\} >$$

then matters are different. Ramanujan in this case has his work cut out for him; for he has to excise p and disable every one of those justifications like Π. Fortunately the cruder form of representation

$$\pi = < p, \{\ldots, \Delta, \ldots\} >$$

(where Δ is the set of premisses of the justification Π) suffices for this doxastic surgery. For all that Ramanujan has to do is remove some member, or members, from each such set Δ, so as to make what is left over *insufficient* for the justification of the proscribed proposition p.

Exactly how Ramanujan will contrive to do this is a matter we can leave, at this stage of our tale, to his own sharp wits. I offered some quite detailed suggestions as to the appropriate method in Tennant [1994], and have just illustrated it above with the toy examples of contracting $[a,\ b]$ and $[a,\ a \rightarrow b]$ with respect to $a \wedge b$. Let me remark, however, on one strategem that Ramanujan considers and quickly rejects.

4 What do we require of a contraction?

4.1 Taking the intuitionistic case seriously: the condition for successful contraction

Ramanujan recalls that the angel has never expressed to him any preference for classical as opposed to intuitionistic logic. Indeed, for all Ramanujan knows, the angel might well be an intuitionist. For (so this story goes) all the proofs that Ramanujan has found in the past on the angel's promptings have been intuitionistic ones. 'Aha!' schemes Ramanujan, 'I'll just respond to every angelic proscription of belief p (at least where p is not itself a negation) by believing $\neg\neg p$ instead, and insisting on using only intuitionistic logic. Then no-one can force me to admit that I'm still committed to p, and I will have obeyed the angel's injunction.'

The angel, reading Ramanujan's thoughts, raises an eyebrow. Ramanujan immediately has second thoughts, recalling the angel's insistence that he *keep an open mind* concerning the proscribed propositions. How can one keep an open mind with respect to p if one believes $\neg\neg p$ even if not, admittedly, believing p itself? Surely an open mind means that one can proceed either to adopt p or to adopt $\neg p$ as a 'new' belief *without thereby incurring any inconsistency*? But if Ramanujan were to pursue the 'cheeky intuitionist' path, he would be unable subsequently to embrace $\neg p$ without thereby engendering conflict with his belief $\neg\neg p$, even in the absence of p.

Realizing this, Ramanujan sees that one has to be careful when formulating the conditions for success in the project of contracting a set A of beliefs with respect to any of its consequences p. The contraction $A - p$ (that is, the core of the contracted *system*) must not only fail to imply p ; it must also be consistent with $\neg p$ (that is: $A - p, \neg p \nvdash \bot$). Against the background of classical logic, these two requirements are equivalent; against the background of intuitionistic logic, however, they are not. In the intuitionistic case the way to express the condition for success is to insist that the contraction $A-p$ be consistent with $\neg p$. This will guarantee that $A-p$ fails to imply p. Intuitionistically, merely insisting that $A - p$ fail to imply p is not enough to guarantee that $A - p$ will also be consistent with $\neg p$.

Let us now be mindful once again of the distinction drawn earlier between knowledge- or belief-*sets*, and knowledge- or belief-*systems*.

The task of contracting a belief-system A with respect to one of its consequences p is this: make the (core of the) result consistent with $\neg p$; and make it as extensive a sub-system of A as you can. The first requirement is that of SUCCESS; the second requirement is that of MINIMUM MUTILATION. Moreover, note that in a *system*, as opposed to a *set*, mutilation is all the greater the more *proofs* or *justifications* one loses. One would like to be able to conserve as much as possible of past computational effort, in the form of previously discovered proofs *that ought still to be available* after the contraction has been completed.

4.2 Withdrawal operations

In order to connect once again with previous writers' treatments, I now revert to talking of belief-sets rather than belief-systems. Consider the following platitudes about contraction of consistent sets A. Any operation '$-$' satisfying these postulates is called a *withdrawal* operation:[4]

1. $A - p$ is still a belief set; indeed,

2. $A - p$ is a subset of A;

3. to contract A with respect to a proposition p that is not even implied by A is easy — simply do nothing! — ; and

4. if p is not a logical truth, then $A - p, \neg p \nvdash \bot$.

5. if p and q are logically equivalent, then $A - p$ and $A - q$ have the same consequences.

The fourth of these is our SUCCESS condition above. There is a little more to it than meets the eye. If p is a logical falsehood, then $\neg p$ is a logical truth; whence if $A - p, \neg p \nvdash \bot$, it will follow that $A - p \nvdash \bot$. So any contraction with respect to a logical falsehood is required to be consistent. This is an exceptionally strong requirement on contraction, and arguably one which should not be imposed if one takes paraconsistent logic seriously. For the paraconsistentist, there could be 'localized' contradictions lurking within a theory, and a contraction might be required (for other reasons than *this* particular inconsistency) 'elsewhere' in the theory. In such a case there is no *prima facie* reason why a contraction to remove a particular logical falsehood $q \wedge \neg q$ should at the same time fix whatever it is that leads to some different inconsistency $r \wedge \neg r$ elsewhere. Thus the paraconsistentist might wish to reject condition (4) and replace it with something else. We shall not pursue this line of inquiry on behalf of the paraconsistentist here; but it is worth noting the considerations involved.

The fifth postulate above is in effect a requirement that the process of contraction be *extensional*. In an ideal world of logically omniscient rational

[4] The terminology is from Makinson [1987], p.388. The first four of these postulates are the so-called Gärdenfors postulates for contraction, other than the controversial (and false) postulate of *recovery (vide infra)*. Note that my form of (4) is equivalent in the classical case to the form in which it is given by Gärdenfors (that is, $A - p \nvdash p$). His form, however, unlike my form, would be inappropriate for the non-classical case, as we have just seen above.

agents it may be; but in a world of finite rational agents, striving to respect logical norms but conditioned by bounded resources for computation, it will not be. A (subset of a) base might have been developed in various directions, yielding a system development which happens to include the logically equivalent propositions p and q, but not as yet any proof of their equivalence. One could perfectly well imagine a situation where p and q had sufficiently different pedigrees by a given stage of system development for it to make a difference which one of them was chosen as the proposition with respect to which to contract. The extensionality requirement will be honoured in the breach, especially whenever the underlying logic of the theory is undecidable. And, even when decidable, the decision problem may be so intractable that it would be unreasonable to suppose that we could, computationally, conjure up a proof of the equivalence of p and q in order to avert the prospect of contracting differently with respect to each of them.

Collectively, the five postulates given above, though true of the normative enterprise of contraction (insofar as it concerns cores), are still *utterly impotent*. For they are true of a contraction process that involves always cutting (the core) back from A to mere logic![5]

4.3 Minimum mutilation

Ramanujan's angel, however, would do more than raise an eyebrow if Ramanujan were simply to retreat all the way back to logic whenever faced with the proscription of a given proposition p. For, remember, the angel had told him to *give up as little as possible*! Admittedly this was a vague injunction, but still it carries considerable weight. Cutting all the way back to logic would clearly be a breach of this injunction in general. We need a further, sixth postulate on contraction: one that will give expression to the requirement of MINIMUM MUTILATION.

It has been suggested that the following captures this requirement (where once again we speak of sets rather than systems):

> 6. RECOVERY: If p is implied by A, then whatever is implied by A is implied by $A - p, p$.

The idea can be attributed to Gärdenfors [1982]:

> ... it is reasonable to require that we get all of the beliefs in [our

[5] I am not the first writer to have made this obvious point; *cf.* Hansson [1991], p.253.

> theory] back again after first contracting and then expanding with respect to the same belief. Formally, this idea is captured by the ... axiom [of recovery]. (pp.93-94)

This dogmatic assertion hardly counts as an argument. In Gärdenfors [1988] the argument for the axiom had still not been improved:

> The criterion of informational economy requires that [the contraction of a theory] be a 'large' subset of [the theory]. Formally this can be expressed as [the axiom of recovery]. (p.67)

Note that our statement of recovery has it that contraction with respect to any *consequence* p of the set A should admit recovery of all consequences upon restoration of p to $A - p$. It would not be strong enough, to capture the intended spirit of minimum mutilation, to require only that contraction with respect to any belief *in* A should admit recovery of all consequences of A upon restoration of that member to the contracted set. For A in general is not logically closed, and indeed is only finitely developed.

Not much interest attaches, therefore, to the claim of Alchourrón and Makinson [1982], p.25, that recovery holds when applied to irredundant sets of propositions; for their form of recovery is stated with the antecedent 'If p is *already in* A ...'. (For Alchourrón and Makinson, an irredundant set of propositions is one none of whose members follows from the rest. This is clearly equivalent to the definition given above.) They are talking about recovery for contraction with respect to *members*; I am talking about recovery for contraction with respect to *consequences*. In the case of *theories*, there is no difference between these two ways of talking, for all consequences are members; but in the case of belief sets *not* closed under logical consequence, there is a great difference, for then there will be consequences that are not members. In the latter case the really interesting form of recovery is recovery for contraction with respect to consequences, that is, when the proscribed belief newly restored *followed from* A, without necessarily being itself a *member of* A.

Recovery is a seductive idea: it is the rash supposition that we shall be able give up so little upon giving up p that, were we to be given just p back again, we would be able to recover *everything* (by way of consequence) that we had secured before. But its attraction is only apparent. Recovery is false; and it fails utterly to give expression to the requirement of minimum mutilation.

5 Why recovery is false

It is a strange feature of intellectual progress that general principles can acquire many adherents even when simple counterexamples to them are at hand. Such is the case with recovery, as we shall now show. We postpone to the next section our examination of the history of its largely uncritical acceptance by adherents of the AGM-tradition of theory change.

We use once again one of our earlier, very simple examples. Take two atomic propositions a and b, and a conditional connective $\rightarrow$ that at least admits of *modus ponens*.

First counterexample to recovery: Take the single belief a, and let it be one's sole starting point in some inquiry. a implies $b \rightarrow a$. Thus $[a] - (b \rightarrow a)$ must be $[\emptyset]$. Now restore $b \rightarrow a$. What one gets is $[b \rightarrow a]$, which does not imply a.

If one has relevantist qualms over the meaning of $\rightarrow$ in this example, one may use disjunction instead: $[a] - (a \vee b)$, $a \vee b \not\vdash a$.

Second counterexample to recovery: One can well imagine a theory (or set of beliefs) based on the 'first principles' a and $a \rightarrow b$. Such principles would not need proof; thus their 'entries' (if we were to consider the belief-system in question) would be $<a, \{\overline{a}\}>$ and $<a \rightarrow b, \{\overline{a \rightarrow b}\}>$ respectively.

One can furthermore imagine the believer having a strong preference for $a \rightarrow b$ over a, were he ever to be faced with the prospect of having to give up at least one, but not necessarily both of them. We shall express such a preference by means of the 'inequality' $a < a \rightarrow b$.

Suppose now that Ramanujan's angel tells him not to believe the proposition b. What is he to do? He has the entry

$$\beta = < b, \ \frac{a \quad a \rightarrow b}{b} >$$

and he is told that b is unwanted. Thus he must disable the little proof

$$\frac{a \quad a \rightarrow b}{b}$$

The obvious way to do this is to *give up* a, since, *ex hypothesi*, we have $a < a \rightarrow b$. The contracted belief-set that results is (the logical closure of) $a \rightarrow b$. Remember that we contracted *with respect to* b. Now, when we restore b, we obtain (the logical closure of) $\{a \rightarrow b, b\}$. In order to obtain from *this* the member a of the uncontracted theory, one would need to commit the fallacy of affirming the consequent.

Note how devastating this simple counterexample is. It has the same force whether or not we regard the set of beliefs as logically closed. For contraction with respect to *consequences*, one cannot guarantee recovery even of *members* of the belief-set contracted! And this holds even for *irredundant* bases! (For $\{a,\ a \rightarrow b\}$ is irredundant.)

It is no rejoinder to these counterexamples to recovery to insist that recovery holds for *theories*, where by 'theory' one means no more than a bare logically closed set of sentences, without any extra information about what might be taken as its base, or what the structure of justifications might be for its various theorems. I am saying that recovery *fails* for *theories*, precisely because the 'bare logical closure' conception is *too thin* for our theoretical needs in the theory of theory change. Our intuitions answer to an implicitly richer, more structured and articulated notion of a theory than is given by bare set-membership of a logical closure. For our purposes, a fully characterized theory would be an even richer object than the logical closure: it would articulate all the logical dependencies by means of which such closure is generated, and it would indicate what basis there might be for such closure.

The falsity of recovery drives home a basic conclusion: *theory contraction is hysteretic.*

6 Why recovery fails to give expression to the requirement of minimum mutilation

Suppose $A \vdash p$ and A is finitely axiomatizable. Then $A = [t]$, for some sentence t. Consider the theory $[p \rightarrow t]$. It is an awful mutilation of $[t]$. Yet it is a successful 'contraction' of $[t]$ with respect to p *that satisfies recovery.*

More generally, suppose A is any theory implying p and $A = [B]$, where the base B may be infinite. Consider the theory $[p \rightarrow b | b \in B]$. It is an awful mutilation of $[B]$. Yet it is a successful 'contraction' of $[B]$ with respect to p *that satisfies recovery.* (In fact, $[p \rightarrow b | b \in B]$ is the so-called 'full meet' contraction of $[B]$ with respect to p. This observation, for the case where B is a singleton or A, I owe to Harvey Friedman (personal communication). Another, equivalent characterization of the full meet contraction of A with respect to p, due to Alchourrón and Makinson, is $[A \cap [\neg p]]$.)

As a special case, take $B = A$. Consider the theory $[p \to a | a \in A]$. It is an awful mutilation of A, obtained as it is by replacing every non-logical truth a in A by the conditional assertion $p \to a$. Yet it is a successful 'contraction' of A with respect to p *that satisfies recovery.*

7 The uncritical acceptance of recovery

I have already cited Gärdenfors's inadequate justifications offered for the axiom of recovery when he first proposed it and when he consolidated the work of the AGM-tradition in his book in 1988. I propose now to take a cursory look at what other writers in this tradition have said about recovery.

Alchourrón and Makinson [1982] offered a 'proof' of recovery for *theories* which went as follows. In the following quotation, the only change I make is to use my own stroke symbol '$-$' for the contraction operation, rather than their symbol, which has a dot above the stroke. Note that in their paper the empty set is represented by the symbol ϕ rather than the usual $\emptyset$. Also, since A is a theory, $p \in A$ is equivalent to $A \vdash p$. Alchourrón and Makinson reasoned as follows:

> ... one can easily derive [recovery] under the assumption that A is a theory. For suppose $x \notin Cn(\phi)$ and $x \in A$ and $y \in A$: we need to show $A - x \cup \{x\} \vdash y$. Since Cn includes tautological implication, it will suffice to show that $A - x \vdash \neg x \vee y$. But since $y \in A$ and A is a theory, $(\neg x \vee y) \in A$, so if $A - x \nvdash (\neg x \vee y)$ then $(\neg x \vee y) \notin A - x$,
>
> so $(A - x) \cup \{\neg x \vee y\} \vdash x$ [†],
>
> so $(A - x) \cup \{\neg x\} \vdash x$, so $(A - x) \vdash x$, contradicting $A - x \in A \perp x$.

I have labelled [†], and displayed, the step in their reasoning to which I want to draw attention. This 'proof', when analyzed carefully as a natural deduction, turns out not to need to appeal to any strictly classical properties of the consequence relation. One can be a little more economical with one's assumptions than Alchourrón and Makinson were, without changing the leading idea of their attempted proof. Cleaned up as much as possible, it can be formalized as follows. I write $A \vdash p$ instead of their $p \in Cn(A)$; I write A, p instead of $A \cup \{p\}$; and I use $x \to y$ instead of their $\neg x \vee y$. Their reasoning, then, has the following formal structure:[6]

[6] The formal proof given here in effect formalizes Alchourrón and Makinson's argument as it proceeds from the point immediately following the first occurrence of the word 'so'

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{A-x\nvdash x\to y}(1)}{x\to y\notin A-x}}{A-x, x\to y\vdash x}(\dagger) \qquad \dfrac{}{\neg x\vdash x\to y}(\text{by logic})}{A-x,\neg x\vdash x}(\text{CUT})}{A-x\vdash x}(\vdash\text{ classical})}{\bot}(\text{Defn. }-)}{A-x\vdash x\to y}(\text{classical reductio})\,(1)}{A-x,x\vdash y}(\text{by }\to\text{E})
$$

Now this is a rather peculiar sort of proof to proffer for recovery. It attempts to show that contracting a theory will not make it lose too many teeth; but the attempt is completely broken-backed. The full statement of recovery is a (universally quantified) conditional:

if $A\vdash y$ then $A-x, x\vdash y$.

Yet no use is made, in this 'proof' by Alchourrón and Makinson, of the assumption (for conditional proof) that $A\vdash y$. Nor is any use made either of the assumption $\nvdash x$ $(x\notin Cn(\emptyset))$ or of the assumption $A\vdash x$ $(x\in A)$, both of which were enunciated at the outset. Instead, what the 'proof' would actually have established (were it correct) is the rather startling general claim $A-x, x\vdash y$ — which would mean that, upon restoring x to the contracted theory $A-x$, one could infer *any* proposition, whether or not it was in the theory A !

Clearly there is something wrong here. The 'proof' above cannot be formally correct. Indeed, the step marked (†) is a primitive one, which Alchourrón and Makinson, as we saw from the quotation above, take without any justification. And it is fallacious. Their 'proof' therefore carries no conviction. Here is an obvious counterexample to the step (†): Let A be $[a,\ a\to b]$ where $a < a\to b$. Then $A-b$ will be $[a\to b]$. Note that $b\to a\notin A-b$. Moreover, $[A-b,\ b\to a]$ is just $[a\to b,\ b\to a]$, which does not contain b. Setting $b=x$ and $a=y$, we have invalidated step (†).

No writer within the tradition, so far as I know, has balked at this egregious fallacy; but some have ventured to offer problematic examples for the axiom of recovery. Among these were Niederée [1991], Fuhrmann [1991] and Hansson [1991]. What would baffle the logician newly introduced to the theory of theory change is why no-one had the good sense simply to disavow

in the quotation given.

recovery *even for theories.* Instead, we find in the abstract of Hansson [1991] the claim

> ...core-retainment together with four of the other Gärdenfors postulates implies recovery for logically closed belief sets. Reasonable contraction operators without recovery do not seem to be possible for such sets.

and in the body of the text (p.255)

> Since it does not seem sensible for a contraction operator to violate core-retainment or any of [conditions (1)-(5) above], a reasonable contraction operator without the recovery property does not seem to be possible.

The quote from Hansson's abstract continued:

> Instead, however, [contraction operators without recovery] can be obtained for non-closed belief bases.

Fuhrmann [1991] shares this view. He turned his attention to contractions of *finite bases* of belief-sets (whether these sets are logically closed or not) and tried to account for how to contract these more modest logical objects, the finite bases.

The prevailing view within the AGM-tradition, then, even among would-be critics of contraction, seems to be that (a) recovery must hold for contraction of *theories*; and (b) recovery need not hold for contraction of *finite bases.* I contend that (a) is false; and that the way AGM-theorists have tried to establish (b) shows how little they appreciate the force of the requirement of MINIMUM MUTILATION.

8 The challenge posed to AGM

Against this orthodox view, Tennant [1994] proposed a general method of contracting belief-*systems* (and hence the sets that are their cores) with respect to any of their consequences. The method would apply in principle even to logically closed systems (i.e. theories). It respects, and exploits, the fact that believed propositions sometimes do, and generally ought to, come with their justificatory pedigrees attached to them. It heeds such information when it is available. The method also respects, and exploits, the fact that there is sometimes information available about the relative vulnerability of beliefs (taken pairwise). That is, one can have two beliefs a and b and prefer one to the other (say, a to b — represented as $b < a$). The operational meaning of such a preference is that, if one is asked to give up one or other, or possibly both, of these two beliefs, and one can make do by giving up just one of them, then one will give up b and retain a. Information about such a preference ordering can be very partial; but, when it is available, it *can* and *ought* to be used in the ideal contraction process.

The method of contraction I have proposed is devoid of any general commitment to recovery, and, *much more so than any of its AGM-related rivals*, it seeks to achieve *minimum mutilation* of belief-systems (hence also of belief-sets) in the contraction process. Its key departure from *AGM* theory consists in this methodological maxim:

> *Do not make your first priority that of formulating a condition on the* **goal** *or* **outcome** *of applying one's method of contraction as a way of giving expression to the requirement of minimum mutilation. Rather,* **make minimum mutilation manifest in the method** *of contraction itself.*

The method proposed in Tennant [1994] could realistically be *implemented*, and makes manifest how minimum mutilation is achieved. All one needs to do is code belief-systems as suitable data structures, structures embodying our conception of them as finitely developed systems of consequences whose entries contain propositions with various justificatory pedigrees; and then work out an intelligent way of exploiting whatever information one has about the relative entrenchment of the beliefs involved (that is, the 'sacrificial preferences' of the believer).[7]

[7]It is an interesting question whether one has to make do with a 'greedy' algorithm when effecting local excisions on the basis of rankings afforded by the logical facts and the entrenchment relation; or whether there is some globally optimizing method of contraction that can exploit both those sources of constraints on the outcome. In further work reported in Tennant [1996a], a provably minimally-mutilating version of the staining algorithm is developed in full and explicit detail. Marek and Truszczynski are also currently investigating the possibility of extending their treatment of 'revision programming', which affords some nice global optimization results, to take care of information about relative entrenchment by means of *default* rules, which were not treated in Marek and Truszczynski [forthcoming]. The idea is that the entrenchment fact $a < b$ could be expressed by the default rule

$$\frac{\mathrm{out(a)} \vee \mathrm{out(b)} : \mathtt{M}(\mathrm{out(a)})}{\mathrm{out(a)}}$$

Finally, my own method of contraction (suitably idealized for non-finite belief-systems) was indifferent to whether the belief-system was logically closed or not; indifferent to whether the underlying consequence relation was that of classical or non-classical logic; indifferent to whether the consequence relation was compact or not; indifferent to whether it satisfied full, unrestricted cut or not; and indifferent to whether one's justificatory pedigrees were well-founded or coherentist.

To summarize: the method of contraction laid out in Tennant [1994] is:

- applicable to both finite and infinite belief-systems
- applicable to both logically closed systems and ones that are not logically closed
- sensitive to justificatory pedigrees of beliefs within a belief-system
- sensitive to the believer's possible preferences among various beliefs
- non-committal with regard to recovery
- minimally mutilating
- implementable
- invariant across choice of logic and 'frill' structural features of the consequence or deducibility relation[8]
- invariant across different 'epistemological orderings' (well-founded or coherentist)

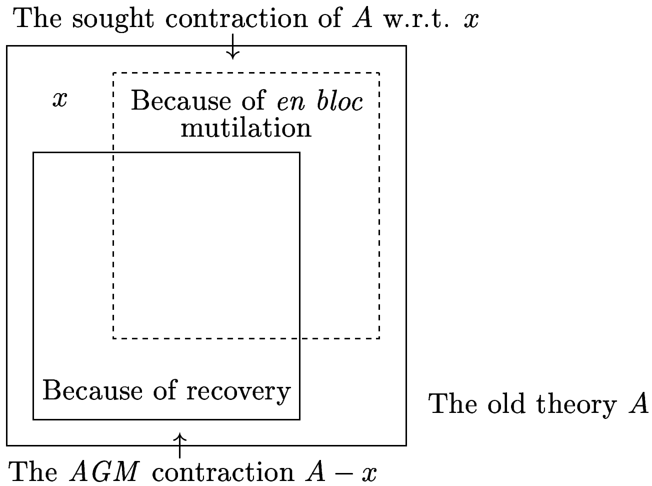
Tennant [1994] contained two main contentions that would have been embarrassing to AGM-theory.[9] The first was that recovery was false, even for logically closed belief-sets. The second was that even those AGM-theorists who had retreated to work on finitely based sets were offering the reader

[8]What I do *not* regard as 'frills' for $\vdash$ are the rules of *Cut for* $\bot$ and *Cut for Gain*. *Cut for* $\bot$ says that if $X \vdash A$ and $Y, A \vdash \bot$ then (some subset of) $X \cup Y$ is inconsistent. *Cut for Gain* says that if $X \vdash A$ and $Y, A \vdash B$ then for some subset Z of X, Y, either $Z \vdash B$ or $Z \vdash \bot$. These two cut rules are, together, weaker than the rule of unrestricted cut that says that if $X \vdash A$ and $Y, A \vdash B$ then $X, Y \vdash B$; yet they suffice for all our cumulative deductive needs in both mathematics and empirical science.

[9]The measure of such embarrassment may be judged from the harshness of the review by Makinson in *Mathematical Reviews*, no. 95i03065. Hansson and Rott [1995] also joined the fray on behalf of AGM-theory. I reply to both these critiques in Tennant [1996b].

badly mutilated contractions. I identified what I called the *en bloc* approach to getting rid of unwanted propositions, and criticized it for maximizing, not minimizing, the mutilation inflicted by their method of contraction. As it happens, *both* kinds of deficiency are evident in Alchourrón and Makinson's method of so-called 'safe' contraction.

A diagram will help to clarify this criticism of AGM-theory:



The dashed box represents the correct contraction of A with respect to x. The inner solid box represents the demonstrable shortcomings of an AGM-contraction $A - x$ in the style of so-called 'safe' contraction. The diagram shows that safe contraction is doubly off-beam. Because of recovery, this kind of AGM-contraction in general contains consequences that it should not contain; and because of its *en bloc* method of excising members of minimal implying sets, it fails to contain consequences that it ought to contain.

Every kind of contraction so far proposed within the AGM literature suffers from one or other, and sometimes both, of these two defects. That they are indeed defects can be established by consulting clear intuitions on carefully chosen examples.

We clearly need a method of contraction which, unlike virtually all those methods in the AGM tradition, does *not* allow recovery to hold in general; and which, unlike many a method in the AGM tradition, does *not* inflict the sometimes *maximizing* mutilations involved when prosecuting the *en bloc* method. That is why the method of 'staining' unwanted sentences was proposed in Tennant [1994].

Why did an implementable, minimally mutilating theory of theory contraction take so long to be proposed? And why did the AGM-theorists not abandon recovery *even for logically closed sets*? The answer to both of these questions is, I believe, to be found in the near-hypnotic hold on their minds of the *search for representation theorems* for suitably defined 'contraction operations'. The excitement and potential elegance of the pure mathematics got in the way of thinking seriously about the accuracy of the modeling provided by the theory for the pre-theoretic but normative phenomena of theory revision. Even when abandoning recovery, these theorists seek surrogates for it so that they can prove some form of representation theorem for their newly defined 'contraction operations'. They have discovered, ruefully, that there are intrinsic limitations on their methods for defining their contraction 'functions' for *theories* so as to avoid commitment to recovery. But they have not learned from this the main lesson: which is that *they shouldn't be trying to define 'nice' contraction functions for which they can then prove their representation theorems.*

AGM-theory suffers from what one might call the *functionality fetish.* It has completely lost sight of the fact that contraction need not be construed as a many-one relation between pairs of the form $\lhd$theory, proposition$\rhd$ and theories. Why *should* we expect 'the' contraction $A - p$ to be single-valued? Why not consider rather a relation $\mathcal{C}$ of contraction whose typical (set-theoretical) member would have the form $\ll A, p >, B >$? For any given theory A and proposition p, there could well be a whole spectrum of possibilities counting as 'reasonable contractions' of A with respect to p :[10]

$$\begin{array}{l} \mathcal{C}(< A, p >, B_1) \\ \mathcal{C}(< A, p >, B_2) \\ \vdots \\ \mathcal{C}(< A, p >, B_n) \\ \vdots \end{array}$$

Once again an extremely simple example serves to make the point. Take the theory $[a,\ b]$, where neither $a < b$ nor $b < a$ holds — that is, where one has no preference for a over b or *vice versa*, should one ever be forced to give up one

[10]It has been gratifying to learn recently, since this paper was written, that awareness of such multiple alternatives is to be found in a mathematically elegant account of data base revision: see Marek and Truszczynski [forthcoming]. The only problem with the latter account is that it does not at present provide any obvious way of handling *preferences* among items as registered by an entrenchment relation. But see the previous footnote.

or other (or possibly both) of them. What is 'the' contraction of this theory with respect to $a \wedge b$? The standard answer from AGM-theorists in the 'base contraction' camp is that the only proper thing to do in the circumstances described is to cut back to logic, jettisoning both a and b in one's 'even-handed' approach to avoiding commitment to their conjunction. (This is an example of what I have called the *en bloc* approach, at work to wreak *maximum* mutilation!) No-one has ever given a satisfactory *philosophical* rationale for such a policy of contraction in the case of 'ties' of this kind. The intuition seems to be that unless one has a sufficient reason for getting rid of just one of a, b but not the other, then one should simply get rid of both — a plague on both their houses, as it were! This is like banging together the heads of two squabbling children when one does not know who the provoking culprit was.

On the view that I wish to oppose to AGM-theory, there are (as we have already seen above) *three* possibilities open to us when contracting $[a,\ b]$ with respect to $a \wedge b$. One possibility is indeed to cut back to the empty theory (that is, to logic). But two equally inviting alternatives are the contracted theories $[a]$ and $[b]$. Against the Continental insistence on a principle of sufficient reason we oppose the pragmatic British dictum that half a loaf is better than no bread. Rather than be left with no theory whatever on which, say, to base survival-relevant decisions, one might as well take a gamble on the culprit conjunct, since it may be that only one of a, b is in error. With no theory, one will die; with a non-negligible chance of having a correct theory, one has a non-negligible chance of surviving. One can imagine a rather convincing evolutionary epistemological argument being made out here for contraction in our simple example having the spectrum of three possibilities rather than just one; and, indeed, for the less severe contractions being *preferable* to the more severe one!

We see, then, that AGM-style contraction is both too *severe* (because of its *en bloc* pruning in the example of $[a,\ b] - (a \wedge b)$ where neither $a < b$ nor $b < a$) *and* too *lax* (because it permits recovery when it should not, as with the example $[a,\ a \rightarrow b]$ where $a < b$). AGM-style contraction is *doubly off-beam.*

9 The inevitability of recovery on the functional fetishist approach

In this section I shall show precisely why the unwanted principle of recovery is forced upon one if one tries to define a contraction function in any of the ways that AGM-theorists have considered. The proofs that I shall provide are essential improvements on those of Alchourrón and Makinson, since they eschew some of the unnecessary assumptions that the latter make about the nature of the relation $\vdash$.

9.1 Meet contraction

Suppose $A \vdash x$ and $\nvdash x$. We do not assume $\vdash$ to be classical. F will be a placeholder for predicates; t will be a placeholder for singular terms.

B is a maximal subset of A such that $\neg F(B) =_{df}$
$B \subseteq A \wedge \neg F(B) \wedge \forall C((B \subset C \wedge C \subseteq A) \rightarrow F(C))$.

Abbreviate this to $B \rhd A : \neg F(B)$.

This abbreviation will be able to feature as the major premiss of the following derived rule of inference:

$$\dfrac{B \rhd A : \neg F(B) \qquad \begin{array}{c} \underbrace{\dfrac{}{B \subseteq A}{\scriptstyle (i)} \quad \dfrac{B \subset t \quad t \subseteq A}{F(t)}{\scriptstyle (i)} \quad \dfrac{F(B)}{\bot}{\scriptstyle (i)}} \\ \vdots \\ G \end{array}}{G}{\scriptstyle [i]}$$

The step marked $[i]$ (with brackets) discharges the assumptions or applications of rules of inference marked (i) (with parentheses).

$A \perp x =_{df} \{B | B \rhd A : B, \neg x \nvdash \bot\}$. (Here $\neg F(B)$ is "$B, \neg x \nvdash \bot$" — the property that we have seen is needed in order to fend off the cheeky intuitionist.) Thus $A \perp x$ is the set of subsets of A that are maximal with respect to the property of being consistent with $\neg x$.

Since $A \vdash x$ and $\nvdash x$ we have that $A \perp x$ is non-empty.

Now let $\gamma(A \perp x)$ be *any* non-empty subset of $A \perp x$. $\bigcap \gamma(A \perp x)$ will be called a *(partial) meet contraction of* A *with respect to its consequence* x. If $\gamma(A \perp x)$ is a singleton, then $\bigcap \gamma(A \perp x)$ is called a *maxichoice* contraction. If $\gamma(A \perp x)$ is $A \perp x$ then $\bigcap \gamma(A \perp x)$ is called a *full meet* contraction.

Consider now the following proof

$$\begin{array}{c} \underbrace{\underline{u \in A\ ,\ \neg x \vdash u}} \\ \Theta(u) \\ \bigcap \gamma(A \perp x) \vdash u \end{array}$$

where the notation $\Theta(u)$ emphasizes that u is a term, occurring in sentences within the proof Θ, for which other terms could be substituted without loss of proofhood:

$$\dfrac{\dfrac{\dfrac{{\scriptstyle(3)}\overline{}}{\bigcap\gamma(A \perp x) \nvdash u}}{\exists B \in \gamma(A \perp x)\ B \nvdash u} \qquad \dfrac{\dfrac{\dfrac{{\scriptstyle(2)}\overline{}}{B \in \gamma(A \perp x)}}{B \rhd A : B, \neg x \nvdash \bot} \quad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{{\scriptstyle(1)}\overline{}}{B \subseteq A} \quad \dfrac{{\scriptstyle(2)}\overline{}}{B \nvdash u}}{B \subset B, u} \quad \dfrac{\dfrac{{\scriptstyle(1)}\overline{}}{B \subseteq A} \quad u \in A}{B, u \subseteq A}}{B, u, \neg x \vdash \bot}{\scriptstyle(1)} \qquad \neg x \vdash u}{B, \neg x \vdash \bot}{\scriptstyle\text{CUT}}}{\bot}{\scriptstyle(1)}}{\bot}{\scriptstyle[1]}}{\dfrac{\bot}{\bigcap \gamma(A \perp x) \vdash u}{\scriptstyle(3)}}{\scriptstyle(2)}$$

Note the application [1] of the rule stated above. In this application the term t in our general statement of the rule is chosen to be B, u (that is, $B \cup \{u\}$). Note also that the foregoing reasoning holds for sets A that need not be theories (i.e. closed under deducibility — $A = [A]$); and for any consequence or deducibility relation $\vdash$, whether or not it is *classical*, whether or not it is *compact*, and whether or not it satisfies *unrestricted* CUT.

Now put $x \to a$ in place of u and extend the proof $\Theta(x \to a)$ as follows:

$$\begin{array}{c}\underbrace{\dfrac{A=[A]\quad \dfrac{A\vdash a}{A\vdash x\to a}\text{by }{\to}\mathrm{I}}{x\to a\in A}\qquad\qquad \dfrac{}{\neg x\vdash x\to a}}_{\displaystyle\Theta(x\to a)}\\ \bigcap\gamma(A\perp x)\vdash x\to a\\ \hline \bigcap\gamma(A\perp x),x\vdash a\end{array}$$

This yields our first major negative observation:

> Every possible kind of *meet contraction* on *theories* A (even those with deducibility relations that are non-classical and/or non-compact and/or not unrestrictedly transitive) with respect to *consequences* satisfies recovery of all consequences:
>
> if $A\vdash a$ then $\bigcap\gamma(A\perp x),x\vdash a$.

9.2 Safe contraction

B is a *minimal* subset of A such that $F(B)$
$=_{df}\ B\subseteq A\wedge F(B)\wedge\forall C(C\subset B\to\neg F(C))$

Abbreviate this to $B\lhd A:F(B)$.

Note that we have the derived rule

$$\dfrac{B\lhd A:F(B)\qquad\qquad \begin{array}{c}\underbrace{\dfrac{}{F(B)}{\scriptstyle(i)}\ ,\ \dfrac{t\subset B\quad F(t)}{\perp}{\scriptstyle(i)}}\\ \vdots\\ G\end{array}}{G}{\scriptstyle[i]}$$

a is a $<$-minimal element of $B=_{df}\ a\in B\wedge\neg\exists b\in B\ b<a$

Abbreviate this to $a\triangleleft B$.

$A/x\ =_{def}\ \{a\in A|\neg\exists B(a\triangleleft B\wedge B\lhd A:B,\neg x\vdash\perp)\}$

$A\cap[A/x]$ is called the ($<$-) safe contraction of A with respect to x.
(If $A=[A]$ then $A\cap[A/x]=[A/x]$.)

Note that we have the following rules of inference:

$$\frac{t \in B}{(B \setminus t, t) = B} \qquad \text{(by the logic of sets)}$$

$$\frac{t \in B}{B \setminus t \subset B} \qquad \text{(by the logic of sets)}$$

$$\frac{t \in A \quad t \notin \{x \in A | \neg\psi(x)\}}{\psi(t)} \qquad \text{(by the logic of sets)}$$

$$\frac{\Delta, \phi \vdash \bot \quad \psi \vdash \phi}{\Delta, \psi \vdash \bot} \qquad \text{(CUT for } \bot\text{)}$$

$$\dfrac{t \lhd B \qquad \begin{matrix} \overline{t \in B}^{(i)} \\ \vdots \\ G \end{matrix}}{G}(i) \qquad \text{(by definition of } t \lhd B\text{)}$$

Consider the following proof $\begin{matrix} \underbrace{u \in A\ ,\ \neg x \vdash u} \\ \Xi(u) \\ u \in A/x \end{matrix}$ constructed by means of the rules just given:

$$\dfrac{\dfrac{\dfrac{u \in A \qquad \overline{u \notin A/x}^{(5)}}{\exists B(u \lhd B \wedge B \lhd A : B, \neg x \vdash \bot)} \qquad \dfrac{\overline{u \lhd B \wedge B \lhd A : B, \neg x \vdash \bot}^{(4)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overline{u \in B}^{(2)}}{(B \setminus u, u) = B} \quad \overline{B, \neg x \vdash \bot}^{(1)}}{B \setminus u, u, \neg x \vdash \bot} \quad \neg x \vdash u}{B \setminus u, \neg x \vdash \bot} \quad \dfrac{\overline{u \in B}^{(2)}}{B \setminus u \subset B}}{\bot}(1) \qquad \overline{B \lhd A : B, \neg x \vdash \bot}^{(3)}}{\bot}[1] \qquad \overline{u \lhd B}^{(3)}}{\bot}(2)}{\bot}(3)\ \wedge\mathrm{E}}{\bot}(4)}{u \in A/x}(5)$$

Note that this proof does not assume that A is a theory; and does not assume any properties of $<$. (This is an important point, to which we shall return when rebutting a criticism of Makinson's below.) The proof $\Xi(u)$ establishes a perfectly general result about 'safe' contractions A/x: to the effect that any member of A that follows from $\neg x$ will follow from A/x; and this holds whatever the nature of A and whatever the nature of the ordering $<$ by means of which 'minimal' elements of A with respect to x may be defined.

Now put $x \to a$ in place of u and extend the proof $\Xi(x \to a)$ as follows:

$$\begin{array}{c}
\underbrace{\dfrac{A=[A]\quad \dfrac{A\vdash a}{A\vdash x\to a}\,\text{by}\to\mathrm{I}}{x\to a\in A}\qquad\qquad \dfrac{}{\neg x\vdash x\to a}}\\
\Xi(x\to a)\\
\dfrac{x\to a\in A/x}{A/x,x\vdash a}
\end{array}$$

This yields our second major negative observation:

> Every possible kind of *safe contraction* of *theories* A (even those with deducibility relations that are non-classical and/or non-compact and/or not unrestrictedly transitive) with respect to *consequences* satisfies recovery of all consequences:
>
> if $A \vdash a$ then $A/x, x \vdash a$.

We have already seen above, however, that recovery of consequences does not even hold for *theories*. Hence:

> *No kind of meet contraction and no kind of safe contraction could possibly be the correct explication of contraction rationally and intuitively understood.*

The AGM-theory's stock of possible contraction operations has been garnered with only optimistic generalities and analogies in mind. We saw this when we discussed the provenance of recovery in Gärdenfors's writings. The *philosophical* or *epistemological* arguments given for recovery were exceedingly thin. The AGM-theory appears not to have been constructed by following the more sensible and cautious method of studying particular examples (including — and perhaps especially — 'small' ones) in order to test one's preliminary intuitions and thereby gain some guidance as to what general conjectures about these operations can and should be borne out once one provides some theoretical operations to do the job of contraction.

10 Makinson on the status of recovery

Recovery is demonstrably false, even for *theories* and even for *irredundant bases*. I have given more than one example to show why. Makinson has observed[11]

> Recovery ... plays a central and apparently essential role in each of the two representation theorems of [AGM 1985] ...

Of course we must pay close attention to Makinson's qualification 'apparently' in this quote. The point of the paper from which this quote is taken was to show that there was a sense in which recovery 'is innocuous, facilitating proofs without generating new properties' (*loc.cit.*, p.383).

But the result that 'shows' this shows no such thing. It is philosophically worthless. The result in question is Makinson's 'Observation' on p.389 of the paper cited, to the effect that any 'withdrawal' operation (i.e. a contraction operation *not* assumed to satisfy recovery) uniquely determines a contraction operation (satisfying recovery) that is what Makinson calls 'revision equivalent' to that withdrawal operation. Moreover, the contraction operation in question is the *laxest* one [my terminology — NT] of all the withdrawal operations that are revision-equivalent to the given withdrawal operation.

The value of this result hinges entirely on what revision equivalence amounts to. Two withdrawal operations are revision-equivalent just in case they give rise to the same revision operation via the so-called Levi identity. This identity states that the revision of a theory with respect to a statement p is obtained by adding p to that theory's contraction with respect to $\neg p$ (and then closing under logical consequence).

We now see why Makinson's result just stated is of no value. The uniquely determined contraction operation *will not be demanding enough*![12] And this is because the five postulates of AGM-theory for revision in its own right are, *themselves*, not demanding enough! So it is *easy* to secure the 'revision equivalence' claimed in Makinson's result.

The mistake arises from thinking of 'revision' and 'contraction', as these notions occur in AGM-theory, as identical (even if only extensionally) to the actual normative operations that are the target of theoretical explication.

[11] Makinson [1987], pp.383-394.

[12] — at least in certain respects. Ironically the same AGM contraction operation can be *too* demanding in other respects!

My whole contention is that AGM-theory's notions of 'contraction' and 'revision' are off-beam; they are *not* the correct explications that we seek of those operations as they are intuitively, rationally and pre-theoretically understood. I propose therefore to mark the contentious status of the AGM-notions by calling them *contrakshun* and *revishun*.

Once we see that the AGM notions are *both* off the mark as explications, we need no longer be impressed by results internal to AGM-theory that establish connections between them. For Makinson's 'Observation' (*loc. cit.*, p.389) can now be re-stated as follows:

> Any 'withdrawal' operation uniquely determines a *contrakshun* operation (satisfying recovery) that is *revishun*-equivalent to that withdrawal operation. Moreover, the *contrakshun* operation in question is that *laxest* one of all the withdrawal operations that are *revishun*-equivalent to the given withdrawal operation.
>
> Two withdrawal operations are *revishun*-equivalent just in case they give rise to the same *revishun* operation via the so-called Levi identity. This identity states that the *revishun* of a theory with respect to a statement p is obtained by adding p to that theory's *contrakshun* with respect to $\neg p$ (and then closing under logical consequence).

The real question is not how the AGM notions are inter-related; but, rather, whether each of these notions itself measures up to the demands of material adequacy to what one might call the pre-theoretical but normative phenomena.

Note that in Makinson [1987] it is conceded that recovery fails when contraction and expansion are performed on a finite (hence not logically closed) set of sentences; but that recovery holds for theories. What I have shown is that recovery should fail to hold in general even for theories, provided only that they are conceived as involving not only logical closure but also information about justificatory pedigrees of each theorem.

The barest essentials of the partial meet approach and of safe contraction, as we have seen, commit one to recovery, *regardless* of the method γ of 'selection' of 'suitable' maximal non-implying sets on the partial meet approach, and *regardless also* of the nature of the ordering with respect to which 'safeness' is to be construed for the method of safe contraction. As far as contraction *via* entrenchment is concerned, the commitment to recovery is just as immediate and inevitable. This is because all that the appeal to available facts about entrenchment yields is a particular selection γ with respect to which one then proceeds in accordance with the method of partial meet contraction.

All three AGM approaches — partial meet contraction, safe contraction, and contraction *via* entrenchment — directly and immediately entail recovery, *without the aid of any particular assumptions concerning the auxiliary notions respectively involved* (such as how one selects the sets whose intersection one wishes to take; or what the relation is like with respect to which one judges 'safeness'; or what particular properties are enjoyed by the entrenchment relation).

Of the three *AGM* approaches, that of safe contraction is a paradigm example of how contrakshun can be doubly off-beam. Consider once again the base $\{a,\ a \to b\}$. Denote by $<$ the ordering relation with respect to which 'safeness' of beliefs is to be judged. Assume that neither $a < a \to b$ nor $a \to b < a$ holds. Then the safe contrakshun of $[a,\ a \to b]$ with respect to b will fail to imply a and fail to imply $a \to b$; *but* it will imply $b \to a$! Thus it both implies too little and implies too much.

Implying too much is the defect of recovery. Implying too little is the defect of what I called the *en bloc* approach to eliminating minimally secure elements of minimally implying sets. (The reader may wish to consult once again the diagram given above.) It is time for *AGM* theorists to contract with respect to recovery, even if precious little of *AGM* theory can be preserved in the result.

References

1. Alchourrón and Makinson [1982]: 'On the logic of theory change: Contraction functions and their associated revision functions', *Theoria*, 48, pp.14-37.

2. Alchourrón and Makinson [1985]: 'On the Logic of Theory Change: Safe Contraction', *Studia Logica*, 44, pp.405-422.

3. Alchourrón, Gärdenfors and Makinson [1985]: 'On the Logic of Theory Change: Partial Meet Contraction and Revision Functions', *Journal of Symbolic Logic*, 50, pp.510-530.

4. Fuhrmann [1991]: 'Theory contraction through base contraction', *Journal of Philosophical Logic*, 20, pp.175-203.

5. Gärdenfors [1982]: 'Rules for Rational Changes of Belief', pp.88-101 in *Philosophical Essays Dedicated to Lennart Aqvist on his Fiftieth Birthday*, Philosophical Studies published by the Philosophy Society and the Department of Philosophy, Uppsala University.

6. Gärdenfors [1988]: *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, MIT Press.

7. Hansson [1991]: 'Belief Contraction Without Recovery', *Studia Logica*, 50, pp.251-260.

8. Hansson and Rott [1995]: 'How Not to Change the Theory of Theory Change: A Reply to Tennant', *British Journal for Philosophy of Science*, 46, pp.361-80.

9. Makinson [1987]: 'On the status of the postulate of recovery in the logic of theory change', *Journal of Philosophical Logic*, 16, pp.383-394.

10. Marek and Truszczynski [forthcoming]: 'Revision Programming', *Theoretical Computer Science.*

11. Tennant [1994]: 'Changing the Theory of Theory Change: Towards a Computational Approach', *British Journal for Philosophy of Science*, 45, pp.865-897.

12. Tennant [1996a]: 'The Staining Algorithm for Theory Contraction', forthcoming.

13. Tennant [1996b]: ‘Changing the Theory of Theory Change: Reply to my Critics’, forthcoming.