Changing the Theory of Theory Change: Reply to My Critics

Neil Tennant

The British Journal for the Philosophy of Science, Vol. 48, No. 4 (Dec., 1997), 569-586.

Stable URL:
http://links.jstor.org/sici?sici=0007-0882%28199712%2948%3A4%3C569%3ACTTOTC%3E2.0.CO%3B2-W

The British Journal for the Philosophy of Science is currently published by Oxford University Press.

Brit. J. Phil. Sci. **48** (1997), 569–586

DISCUSSION

Changing the Theory of Theory Change: Reply to My Critics

Neil Tennant

ABSTRACT

'Changing the Theory of Theory Change: Towards a Computational Approach' (Tennant [1994]; henceforth *CTTC*) claimed that the *AGM* postulate of recovery is false, and that *AGM* contractions of theories can be more than minimally mutilating. It also described an alternative, computational method for contracting theories, called the Staining Algorithm. Makinson [1995] and Hansson and Rott [1995] criticized *CTTC*'s arguments against *AGM*-theory, and its specific proposals for an alternative, computational approach. This paper replies as comprehensively as space allows.

1 Introduction

In my paper 'Changing the Theory of Theory Change: Towards a Computational Approach' (Tennant [1994]; henceforth *CTTC*), I took *AGM*-theory to task for its reliance on the postulate of 'recovery' for theory contractions, and for its recommendation of what turn out to be more than minimally mutilated results when their contractions are carried out on theories via their finite bases, if they have them. I proposed abandoning the postulate of recovery, and contracting theories instead by means of a method that I called the *staining algorithm*. That algorithm was only sketched in *CTTC*, and a more detailed statement of it was promised for a forthcoming paper. This more detailed statement has now been given in Tennant [1996]. Makinson [1995] defended *AGM*-theory against *CTTC*. Elsewhere (Tennant [1997]) I bring out afresh the deficiencies of *AGM*-theory in the two regards just noted, with particular reference to Makinson's work.

Hansson and Rott [1995], two other figures in the *AGM*-tradition, also ventured a critical response to *CTTC*. They wrote that they welcome 'any chance to open a dialogue with a wider community of researchers on the potentialities and limitations of [the *AGM*-] framework' (p. 362); but, like Makinson, they charge me with misrepresentation and/or misunderstanding of certain important aspects of that theory. This paper airs further my differences of opinion with these three *AGM*-theorists.

2 On the question of normativity

2.1 Finitude and normativity

Hansson and Rott write (p. 373):

> We agree that the logical investigation of belief change should be primarily concerned with the behaviour of idealized, rational agents. This does not, however, mean that it should be concerned only with ideally rational agents who have access to unlimited storing and computational capacity.

Now the issue of ideal rationality in belief change is orthogonal to the issue of finitude (of memory or of computing capacity). Nothing in what I wrote commits me to the view (which I reject) that ideally rational agents either should or even *could* have, in Hansson and Rott's terms, 'unlimited storing and computational capacity'.

I have nothing to say about agents with infinite cognitive resources (space, time, and information). We are *all* finite agents, yet we are able to excogitate various norms and hold each other responsible to them. A computational theory working with finite data structures and invoking only algorithmic procedures and transformations on those data structures can still be a normative

account of what *ought* to happen (in the 'ideal' case) rather than a descriptive account of what *actually* happens in the messy world of performance as opposed to competence.

2.2 The finite predicament

My comments in *CTTC* about the finite *predicament* of cognitive agents presupposed a background of normative expectations *of which one could fall short*. The predicament was that certain logical relations might obtain, or logical properties hold, without our yet being appraised of them. Logical myopia is what puts us in the finite predicament, not innocence of whatever logical norms have to be respected. The myopia is of the kind that prevents us from seeing that the norms ordain a certain result; it should not be regarded as evidence that there cannot be any such pertinent norms in the first place.

Hansson and Rott attribute to me (p. 373) the two requirements of 'uncompromising idealization and computational tractability' and say that these 'do not seem to be compatible in an obvious way'. They fail, however, to provide any textual evidence that I am an *uncompromising* idealizer. Indeed, the following textual evidence (*CTTC* at p. 890) is to the contrary:

> When a model of godly competence appeals to infinitary syntactic objects (such as the logical closure of a set of sentences, each with its potentially infinite set of justifications from the axioms), then a parallel model of our finitary handling of perforce finite fragments of such infinitary objects will have to put up with some occasional rough in order to be generally smooth.

Hansson and Rott also go astray (at p. 369, fn. 13 of their paper) about what I called (at p. 889) the 'working identity' of a theory. They suggest that I meant a base for the theory. But the 'working identity' of a theory was defined on p. 889 of *CTTC* as being given, at any stage of its development, 'by the finitely many justifications that [had] thus far been provided for claims recognized to be part of the theory ... At any working stage, only finitely many different pedigrees would have been discovered for any given claim.' The notion of working identity is crucial to an understanding of the new perspective I was urging on the problem of theory change.

3 Invariance under choice of underlying logic

Hansson and Rott rely on classical logic, and so are not sensitive to the point that one's method of contraction ought to be invariant across choice of underlying logic. For example, they say that because one would in due course discover a proof of a from its consequences $a \vee b$ and $a \vee \neg b$, one would be

forced, upon contracting with respect to a, to decide which of $a \vee b$ and $a \vee \neg b$ had also to go. But I disagree on two points. First, the underlying logic of the system might be a sub-classical logic in which the rule of constructive dilemma does not hold. Secondly, one of the main points in *CTTC* concerned 'totally dependent progeny' of a sentence a that is to be excised. If the system development renders both $a \vee b$ and $a \vee \neg b$ totally dependent on a, then *both* $a \vee b$ and $a \vee \neg b$ will be eliminated at the downward step immediately following the staining of a. If, however, either of the sentences $a \vee b$ or $a \vee \neg b$ had at least one justification, already explicitly registered in the system, from a set of premisses Δ other than $\{a\}$, then that sentence might survive the contraction, provided that at least one of those other supporting premiss-sets Δ was intact as well.

4 The technical definition of system development

Hansson and Rott regret that *CTTC* lacked the kind of technical precision (in the definition of a system development) that would be called for in a mathematical journal, and ask a series of questions about what I might have meant by a 'system development'. But *CTTC* only aimed to give the broadest features of the staining algorithm pending its detailed formulation elsewhere. To that end I sketched clearly the main ideas behind a system development, and was content to leave the absolutely precise (and necessarily more prolix) details to the reader. It was clear that I intended a system development to be the formal counterpart of a logically articulated structure of finitely many beliefs in the system of beliefs of a cognitive agent, some of which beliefs might be derived from yet others, and some of which beliefs would otherwise count as 'self-justifying' within the system, and be able to serve as starting points for justifications. I called the justificatory pedigrees 'proofs', and pointed out that we could restrict our attention, within any proof, to just its set of premisses and its conclusion. I made the point that any sentence (serving as a belief in the system) might have more than one set of premisses on which it depended, within that system. It follows as a matter of course that any premiss thus involved would itself have to *be in the system*, hence, in turn, either be justified by yet other sentences in the system, or be self-justifying. Hansson and Rott's suggestion that my technical definition (in so far as I was offering any such thing) left it open how the system was to be 'closed' should be assessed in this light.

5 Hansson and Rott's concessions

Hansson and Rott make two major concessions in response to two of my main critical points. These critical points were:

1. The *AGM*-postulate of recovery can be shown to be false; and
2. The *AGM*-method of safe contraction, even on finite bases, is more than minimally mutilating.

Hansson and Rott still, however, fail to appreciate another one of my main critical points, namely that

3. even if the *AGM*-method of finite base contraction could do as well as the staining algorithm in the finite and well-founded case, it cannot match the staining algorithm in the coherentist (non-well-founded) case.

On the first issue, they make two further claims that are in some tension. First, they suggest (1) is not original, having been anticipated by other writers in the *AGM* 'tradition'; and, second, they try to downplay the importance of that very point as devastating for the 'classic core' of the *AGM*-tradition. But if (1) is really not a very important point, even if true, one would not expect any fuss about originality. And if other writers had indeed already made point (1) before me, the fact remains that these writers did not follow up where their counterexemplary intuitions led them, and point out the devastating ramifications of the falsity of the postulate of recovery for the classic core of the *AGM*-tradition—that is, for the method of partial meet contraction and the method of safe contraction. Strip away those two methods, and abandon the postulate of recovery, and there is hardly any of the original *AGM*-theory left. So-called *AGM*-theorists would be practising so-called *AGM*-theory in name alone.

Despite their concessions, Hansson and Rott still maintain that recovery is not the, or even a, 'main foundation stone' of *AGM*-theory. This does not accord with my reading of the historical record of the extant literature (for details of which, see Tennant [1997]). It also involves an appeal to Makinson's dispensability-for-revision thesis (Makinson [1985]), which, as I shall now argue, carries no conviction once one appreciates that it is really a dispensability-for-*revishun* thesis—where *revishun* is the AGM version of what genuine revision would be.

6 The status of recovery

6.1 Contraction or contrakshun?

Recovery, I claimed—and still do claim—is demonstrably false, even for *theories* and even for *irredundant bases*. I gave more than one example in *CTTC* to show why. Unfortunately, Makinson does not address my examples. Instead, he takes exception to my calling recovery 'the main foundation stone' of *AGM*-theory. He writes

> [Tennant] repeatedly asserts that the 'main foundation stone' of the AGM tradition is the postulate of recovery. This is a serious misapprehension.

Makinson's claim here is curious, given that he himself said (in Makinson [1987], pp. 383–94):

> 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. 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 is 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 does not have the value he imputes to it. The uniquely determined contraction operation *will not be demanding enough.*[1] 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 problem here 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.

My major 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

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

connections between them. For Makinson's 'Observation' (*loc. cit.*, p. 389) should be read as concerning only the *AGM* notions of *contrakshun* and *revishun*. The real question is not how the *AGM* notions are interrelated; 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.

6.2 On the 'emergence' of recovery

6.2.1 Hansson and Rott on recovery

Hansson and Rott claim (p. 366) that

> it is ... fairly difficult to construct a plausible operation of contraction that operates on belief sets and does not satisfy recovery. Hence, recovery has a strong standing as an emerging property, rather than as a fundamental postulate, of belief set contraction.

Now we must bear in mind that by 'belief set' Hansson and Rott mean a *theory*, that is, a logically closed set of sentences. Even with this qualification, what they say can be refuted. The staining algorithm, applied to finite system developments, provides a completely adequate method of contraction. It is demonstrably minimally mutilating, and recovery fails for its results—which is exactly how matters ought to be. (See Tennant [1996] for details.) Moreover, in response to the anticipated objection that by confining oneself to finite system developments the advocate of the staining algorithm is sacrificing generality, I can present the following result, due to Harvey Friedman:

> *Upward Finitizability Theorem*
> For every axiomatizable theory T there is a decidable base B and some natural number n such that every theorem p of T follows from at most n distinct minimal p-implying subsets of B.

A subset C of B is minimal p-implying just in case C implies p, and any p-implying subset of B all of whose members are implied by C implies all members of C. (It follows from compactness that every minimal p-implying subset of B is finite.) The number n is called the *minimal implying index* of the base B. In the general case, n can be chosen as 1. There are also Upward Finitizability results for more 'natural' axiomatizations B of the given theory T, where n is slightly higher, and implications are *modulo* some distinguished finite proper subset B^* of B. (See Friedman and Tennant [1997] for further details.)

What these Upward Finitizability results ensure is that confining oneself to finite system developments and the staining algorithm entails *absolutely no loss of generality vis-à-vis* axiomatizable theories. This is to be contrasted with the state of affairs afflicting that version of *AGM*-theory, due to Hansson [1993a], in

which one confines contractions to finite bases, but seeks a surrogate for recovery. Hansson's surrogate condition entails that every contraction of a finitely axiomatized theory has to be finitely axiomatized. Another result due to Friedman shows that *this* finite-base contraction approach is doomed to irremediable loss of generality:

> There is a finitely axiomatized theory none of whose partial meet contractions that are not full meet contractions is finitely axiomatizable.

6.2.2 Makinson on recovery

Makinson, like Hansson and Rott, claims that recovery 'emerges' for contraction of *theories*. In his review he distinguishes partial meet [contrakshun], safe [contrakshun] and [contrakshun] via entrenchment. Then he writes:

> recovery is a property that emerges from the above three approaches when they are applied directly to belief sets already closed under logical consequence.

The implication, at that point in his review, was that I had somehow missed this fact of 'emergence', or failed to make some necessary discrimination between belief sets already closed under logical consequence, and belief sets that were not. But by his very own words, all three of the approaches to theory contraction that he distinguished stand indicted—*for they admit recovery*. Moreover, the reader is being asked to believe that the 'emergence' in question is somehow *in the (normative) phenomena themselves*, and that it therefore has to be respected in our theoretical modelling. The 'emergence' of recovery, however, is entirely an artefact of an inadequately motivated set of mathematical operations that aim to provide explications of the pre-theoretical notion of contraction. *Pace* Makinson, recovery 'emerges' only for *contrakshun*, not *contraction*.

The claimed 'emergence' of recovery for theories holds only for the peculiar kinds of mathematical functions that *AGM*-theory has allowed itself to consider. It is only too narrow a notion of contraction (that is, *contrakshun*) that allows recovery in; a suitably broadened notion of contraction allows us to exclude this unwelcome intruder. I claim that a suitably broadened notion is available in *CTTC*.

6.3 Recovery was a legitimate target

CTTC gave sound reasons for mistrusting the *AGM* notions as surrogates for the pre-theoretical explicanda. Nothing that Makinson, or Hansson and Rott, say in their responses to *CTTC* engages its central challenge to the *AGM*-theory.

It should be obvious that recovery is the 'main foundation stone' of the

AGM-theory. After all, the remaining postulates governing contrakshun (even those embodying extensionality requirements) are toothless, in so far as they are satisfied by cutting all the way back to just logic. But according to Makinson, as already quoted, to think that recovery is the 'main foundation stone' would be to labour under 'a serious misapprehension'. Rather, he tells his reader:

> If one is to search for 'foundation stones' for the three approaches in the AGM tradition, one can say that the central idea of the *partial meet approach* is the intersection of suitably selected maximal non-implying sets, whilst that of *safe contraction* is the elimination of minimally secure elements of minimal non-implying sets. For *contraction via entrenchment* the key idea is to use a relation of 'entrenchment' between statements, relative to the theory under contraction, constrained by the requirement that it is not only connected but also well-behaved with respect to conjunction, *to permit explicit definition of the result of contraction* [My emphases—NT].

Two negative observations recorded in Tennant [1997] are relevant to the matter in hand. I show there, with fully detailed formal proofs, that the *barest essentials* of the partial meet approach and of safe contraction commit one to recovery, *regardless* of the method (called 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.

There is no way that *AGM*-theorists can possibly downplay the importance, for *AGM*-theory, of recovery. In trying to make out that the *real* achievement of *AGM*-theory is to be located in these other, supposedly distinct, 'central ideas' of partial meet contraction, safe contraction and contraction *via* entrenchment, Makinson is trying to deflect properly targeted criticism. *All three approaches 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).

7 Entrenchment and entrenshmunt

7.1 Makinson on entrenchment

Entrenchment can be thought of in two different ways: as an ordering of

sentences, or as an ordering of *sets* of sentences. The latter sort of ordering is what is used for so-called 'relational partial meet' contrakshun.

That I understood the difference should have been apparent from pp. 866–7 of *CTTC*, where I wrote that *AGM*-theory's

> aim is to establish representation theorems of the form 'Contraction satisfies such-and-such global postulates if and only if it is given by a function defined thus-and-so by appeal to the entrenchment relation [among (sets of) sentences] with such-and-such features'.

Yet Makinson complains that I 'misuse ... well-defined technical terms', and says that I 'constantly [conflate] and indeed at times [confuse]' the three *AGM* approaches to contraction. 'In this connection,' he writes,

> it should be emphasized that throughout the literature on belief-change, 'entrenchment' is a precisely defined and severely constrained technical term, employed in only one of the three AGM constructions. Readers will be confused by [Tennant's] use of the term, without warning, to refer to almost any kind of preference ordering employed to assist a selection process.

Despite Makinson's claim, however, there is no one precise definition of entrenchment 'throughout the literature on belief-change'. There is already the vacillation, which I have noted and taken account of, between entrenchment conceived as a relation among sentences, and entrenchment conceived as a relation among *sets* of sentences. Moreover, I introduced an entrenchment relation in *CTTC for my own purposes*, to assist with the staining algorithm; and when I did so I gave a perfectly precise definition of what I meant by entrenchment. None of my criticisms of *AGM*-theory's commitment to recovery depended on having any particular (and different) 'precisely defined and severely constrained' notion of entrenchment—for, as stressed above (and rigorously proved in Tennant [1997]), recovery is the ineluctable byproduct of a certain definitional strategy for contrakshun functions, and is *independent* of the precise structural features of entrenchment.

Moreover, Makinson's claim that two of the three *AGM* approaches do *not* appeal in any way to entrenchment is true only by appeal to the letter of particular definitions. There is no denying that with the method of safe contraction, the ordering with respect to which safety of beliefs is determined is, intuitively, very much like an entrenchment relation among sentences, as that relation might be pre-theoretically understood.

That observation leaves only the partial meet approach as innocent of any kind of entrenchment notion—hence, so much the worse for partial meet contraction. Not only does it yield recovery, but it also *ignores* a major source of constraints on the process of contraction (namely, one's possible—even if partial—'sacrificial preferences' among various beliefs), a

source that is incontrovertibly 'in the phenomena' for which a theory of theory change has to account.

Indeed, it was because of my concern to represent the *intuitive* notion of entrenchment—the one that reveals itself *in the phenomena*—that I defined entrenchment slightly differently, as a relation among sentences, than Gärdenfors had done. Unlike Gärdenfors, I do not require the formal notion of entrenchment to be *complete* (or *connected*). Completeness (or connectedness) is a matter of every pair of sentences being comparable in at least one direction—that is, requiring that at least one of them be at least as well entrenched as the other. This simply fails to be the case with the intuitive notion of entrenchment, even when the would-be contractor of a belief set is an expert logician seeking to abide by all reasonable logical norms. Gärdenfors's notion, then, should be called *entrenshmunt*, to mark how unrealistic it is as a modelling match for what is supposed to be in the phenomena being modelled.

7.2 Hanson and Rott on entrenchment

Hansson and Rott's complaint about my use of the term 'entrenchment' echoes Makinson's similar complaint in his review. But no small community of logical specialists working on the topic of theory dynamics should be allowed to lay claim to a monopoly on the use of that very versatile word of ordinary English.

The word readily suggests itself for a variety of related uses, more or less literal, more or less metaphorical, in epistemology at large. The term 'entrenched', with or without a comparative modifier 'relative(ly)', is on the tip of one's tongue in any post-Positivist discussion in epistemology. Gärdenfors himself adopted the term in 1984 because it was already in wide use in epistemological writings. Indeed, he even pointed out that relative entrenchment can be pragmatically informed, and is not simply determined by strength of evidential support, or likelihood of truth. If a sentence played an extremely important organizational and integrative role within a theory, then it could acquire a higher degree of entrenchment (even if one were inclined to regard such a high-level sentence as having a more 'speculative' or 'precarious' status in the light of further evidence).

Moreover, *AGM*-theorists have not given us a *sensible* and obviously *applicable* definition of the term, a definition that would have made it abundantly clear that the actual phenomena in the normative enterprise of theory contraction and revision lent themselves to non-question-begging representation by means of the defined term. The strict technical term of the *AGM*-theorists (which, as indicated above, I shall call *entrenshmunt*) is forced to obey certain definitional constraints which are highly implausible if thought of

as characterizing the intuitive, pre-formal phenomena involved in holding various beliefs in a logically organized way on the basis of certain evidence.

Entrenchment as intuitively understood is hardly a complete or connected relation across sentences in our systems of belief. There could well be sentences x and y such that it is neither the case that x is more entrenched than y, nor the case that y is more entrenched than x. Nor need this be owing to the fact that x and y are *equally* entrenched, that is, 'tied' under considerations of entrenchment. Rather, it could be owing to the fact that they are simply incomparable, and that the believer would not be in a position even to begin to marshall such considerations as would give rise to such a tie, or break it in favour of one or the other of the two sentences. Yet consider for a moment the *AGM*-condition (E3) on entrenshmunt, as given by Hansson and Rott:

$$\text{(E3) } \phi \leq \phi \wedge \psi \text{ or } \psi \leq \phi \wedge \psi$$

In conjunction with the unobjectionable conditions

$$\text{(E1) if } \phi \leq \psi \text{ and } \psi \leq \xi \text{ then } \phi \leq \xi$$

and

$$\text{(E2) if } \psi \in Cn(\phi) \text{ then } \phi \leq \psi$$

this condition (E3) has the immediate consequence that entrenshmunt is connected. Hansson and Rott note this fact (at p. 363), but fail to draw the obvious conclusion that the very disconnectedness of the intuitive entrenchment relation, in general, makes their chosen condition (E3) extremely implausible as part of a formal characterization of the intuitive notion.

If Hansson and Rott were to be held to their own exacting standards of respect for the evidence coming from mistaken reasoners, they would never insist on a condition such as (E3) on entrenshmunt. I for one do not allow that any two arbitrarily chosen beliefs of mine admit of a clear relative entrenchment decision. Certainly, some of my beliefs are better entrenched than others, and some of them are equally entrenched; and, I hope, the relation of entrenchment among my beliefs satisfies the conditions of transitivity (E1) and dominance (E2), for these are rationality constraints to which I willingly submit. But apart from that, I maintain that there really is no further information that can be warrantedly extracted from my asymmetric willingness (whenever it obtains) to sacrifice one belief rather than another when I am called upon to cease believing both of them. I demur, that is, at being 'represented' or 'formalized', *qua* cognitive agent, the way that *AGM*-theorists would have it, when they are availing themselves of the notion of entrenshmunt (among *my* beliefs) for their theoretical purposes. If this strenuous denial of the applicability of the *AGM*-machinery of entrenshmunt in my own case is not to be heeded, then Hansson and Rott owe us a convincing case for why this is so.

The intuitive and pre-formal notion of entrenchment, properly understood, is there to be appealed to as affording more definitive reasons, in certain situations, for opting for one theory contraction $T -_1 p$ over another, $T -_2 p$. There can be alternative contractions of the same theory T with respect to the same sentence p. All that entrenchment considerations (properly conceived and characterized) are supposed to do is help one narrow down the range of possible choices for the eventual contracted outcome. This help comes in the form of local indications of which of two sentences x and y to give up if, say, they are both in some minimal implying set for some other sentence z that we have already decided to give up. If it happens to be the case that y is more entrenched than x (for the believer in question), then we (the contractors of his system of beliefs) will give up x on his behalf and hold on to y. That is all the operational import of considerations of entrenchment. No epistemologist or reasonably reflective ordinary believer would ever venture to suggest that there should always be, in the background, some amazingly fertile body of considerations that would afford, for any two chosen sentences x and y, *exactly one* of the following possibilities:

1. x is more entrenched than y;
2. y is more entrenched than x;
3. x and y are equally entrenched.

Suppose, though, that it turns out that the contractional upshot of equal entrenchment of x and y is exactly the same as the incomparability of x and y—incomparability being the fourth possibility that I am saying is mistakenly ignored by *AGM*-theory when it deals with entrenshmunt. It still would not follow that (E3) was a correct characterization of entrenchment thus extended.

The only unobjectionable conditions on (strict) entrenshmunt are those of transitivity (E1 above) and dominance (E2 above), from which it follows that

> if x logically implies, but is not logically implied by y, then y is more entrenched than x ($y < x$).

Assume now that a and b are logically independent—that is, neither one of them logically implies the other. Now note that $a \wedge b$ logically implies, but is not logically implied by a. Hence $a \wedge b$ is less entrenched than a.

Likewise, $a \wedge b$ logically implies, but is not logically implied by b. Hence $a \wedge b$ is less entrenched than b. Yet (E3) tells us that $a \wedge b$ is at least as well *entrenshed* as one or other of a or b—whence we may conclude that entrenshmunt and entrenchment are two very different matters.

Hansson and Rott object (p. 365) that (E2), which I grant, is

> inconsistent with the idea [Tennant] later entertains that the elements of a theory's base may be more entrenched than 'their distant consequences' (p. 875).

Here they misconstrue me. One can perfectly well have each of the axioms $a_1, \ldots, a_n$ more entrenched than a distant consequence b of $\{a_1, \ldots, a_n\}$, while yet b is more entrenched than the *conjunction* $a_1 \wedge \ldots \wedge a_n$. Hansson and Rott fail to note that (E2) applies to *single sentences*, whereas in the passage they criticize I was talking about distant consequences of *sets of* sentences.

I was not myself maintaining that, given the proper notion of entrenchment,

> the elements of a theory's base may be more entrenched than 'their distant consequences' ...

What I was maintaining, and still maintain, is that

> for every sentence a in the base B, a will be more entrenched than any distant consequence of B

rather than

> for every sentence a in the base B, a will be more entrenched than any distant consequence of a

which is the view that Hansson and Rott mistakenly attribute to me. The latter view would indeed be in direct contradiction with (E2); but it is no claim of mine.

8 On finite base contraction

Makinson downplays the importance of what I called the 'downward' step of the staining algorithm in *CTTC* as follows:

> it was already pointed out by Fuhrmann [*J. Philos. Logic* 20 (1991), no. 2, pp. 602–625] that, for base contraction, [Tennant]'s 'downward' step *takes care of itself*—it is effected automatically by taking the logical closure of the contracted base only. Statements in the closure of the original base that are left unsupported by the contracted base are thus eliminated (my emphasis—NT).

I had, however, anticipated precisely this point in footnote 23 on p. 880, which I shall here quote in full:

> Note that for Fuhrmann it would only be the Upward pass that came into play. He makes no provision for the Downward pass required on the new perspective. And if in reply he were to claim that the effects of Downward passing would anyway be secured somehow 'in the wash' by operating on *bases* of theories rather than on the whole theories themselves, one could point out that he is thereby limiting himself to the foundationalist case. The Staining Algorithm, by contrast, applies to both the foundationalist and the coherentist case.

It is relevant here that the subtitle to *CTTC* is 'towards a computational approach'. I was concerned to develop the rudiments of a theory of theory

change that would be *implementable*. For that to be the case, theories cannot be over-idealized as logically closed, hence consisting of infinitely many sentences. Rather, logical closure has always to remain as an operation that *could* be carried out, but perforce actually *has* been only to a very limited and necessarily finite extent. No computer can hold an infinite data structure—not even if it is supposed to be an infinite theory. It might well hold a *finite* set of principles 'as' that theory, provided that their logical closure is indeed the theory in question. (Such a finite set of principles, with their pedigrees of justification, forms what I call a *system development*). The computer might even hold finitely many axiom *schemata*, each of which has infinitely many instances; but it could have *instantiated* those schemata only finitely many times by any given stage. All this is simply part of what I discussed as the 'finite predicament', at pp. 889–92 of *CTTC*.

Given that we are in the finite predicament, I was interested in a computational theory of theory change that would go to work on the necessarily finite representations or data structures with which a proper analysis would furnish us, as input for whatever contraction and revision algorithms we could develop. Even when dealing with a finitely axiomatized theory, it would be a profligate waste of past computational effort—effort expended in developing the theory to the point thus far reached—to effect a contraction with respect to a present statement p of the theory by cutting all the way back to some favoured proper subset of the (finite) set of axioms, and then having to *re-derive* whatever would survive the contraction—that is, whatever had already been derived from axioms unaffected by the expurgation in question.

Yet Makinson assumes that all the theorist has to point out is that the surviving statements would be rederiv*able* after only an Upward pass of my staining algorithm; hence, the Downward pass would be otiose. What he ignores is that from the computationalist perspective the really otiose move is to undertake all that extra deductive effort now entailed by the need to *re-generate* the surviving consequences of the old theory.

The Upward and Downward pass of the staining algorithm were designed so that computational logic's labours would not be lost. Suppose that you have the (finitely developed) theory containing p, whose member sentences have been laboriously derived—if not assumed as axioms—by means of *past computed proofs*. On my account one would have preserved as part of the data structure of the theory thus far unfolded the *premiss pedigrees* of the statements derived. When faced with the task of contracting with respect to p, the algorithm swings into play with both Upward *and Downward* sweeps in order respectively to 'stain' the worst culprits responsible for the erstwhile presence of p, as well as those past statements of the theory that would now go begging for justification. Computationally, it would simply be a carefully controlled shake-out. Once the stained sentences have been thrown out,

what is left of the data structure—which in general would be *much more* than the mere 'contracted finite base' of axioms—is ready to be built on by future computational effort. Only now we know that future computational effort is not being spent re-doing (a possibly very large part of) what had already been done. We shall not have to re-derive the surviving theorems all over again.

9 The deficiencies of the '*en bloc*' approach

Safe contraction is a paradigm example of how contrakshun can be doubly off-beam. Consider once again the base $\{a, a \rightarrow b\}$. Denote by $<$ the ordering relation with respect to which 'safeness' of beliefs is to be judged. Assume that neither $a < a \rightarrow b$ nor $a \rightarrow b < a$ holds. Then the safe contrakshun of $[a, a \rightarrow b]$ with respect to b will fail to imply a and fail to imply $a \rightarrow b$; but it will imply $b \rightarrow a$. 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. Nowhere in his review does Makinson confront my central charge that such *en bloc* elimination sins against the requirement of minimal mutilation; nor does he address my central claim that the virtue of the staining algorithm is that it *avoids* this defect common to both his method of safe contraction and Fuhrmann's method of finite base contraction.

10 Further clarifications

Makinson claims that my 'pedigrees of justification' are *sequences of statements*, and adds that these are 'presumably from the same belief state, although [Tennant] is not explicit'. Nothing in what I wrote justifies such a construal. I made it clear that the various pedigrees II were *proofs*. But the antiquated Hilbert notion of a proof as a sequence of statements is one that I decline to have visited upon me. Those proofs are best thought of as *natural deductions*, in the style of Gentzen and Prawitz. As such, those proofs will often contain sentences (labelling nodes of the proof tree) that do *not* themselves belong to the belief set for whose members the proofs provide justification. An obvious example would be the antecedent of a conditional, when it occurs as an assumption for conditional proof; or, even more to the point, the sentence x when it is assumed for *reductio ad absurdum*, in order to justify the belief $\neg x$.

Makinson also misreads me in connection with my method of choosing sentences for excision via the staining algorithm. This algorithm, he says,

> intends to eliminate at least one minimally secure element from each minimal implying subset ([Tennant] announces exactly one, but as the sets need not be disjoint, there is no guarantee that this is always possible).

Nowhere do I 'announce exactly one'. The context (p. 878) was:

> *It would be nice if* ... exactly one ... (my new emphasis—NT).

In this comment I was *presupposing* that unique choices for excision from minimal implying sets are not possible *in general*, and making the point that there could, nevertheless, be many practical cases in which such unique choices *could* be made. *Whether* this were indeed so in any particular case would depend on the structure of the theory (or system development) being contracted, the pattern of entrenchment of statements within it, the available pedigrees present at the time of the contraction, and the sentence with respect to which one was contracting. But a moment's reflection on the problem

$$[a, b, c] - (a \wedge b) \vee (b \wedge c) \vee (a \wedge c)$$

where one is indifferent among a, b, and c, shows that at least one of the minimal implying sets $\{a, b\}$, $\{b, c\}$, $\{a, c\}$ would have to take a double whammy upon the indicated contraction.

11 An apology in closing

Makinson accuses me of 'disregard of relevant results in the literature cited', and of 'apparent unawareness of other highly relevant contributions'. He does not state any result that I had allegedly disregarded, nor adduce any evidence of culpable ignorance. He cites only two papers by Hansson [1993a, b] and one by Nayak [1994].

It would be an unnecessary diversion to explain why the papers by Hansson, in the light of the preceding discussion, provide no countervailing considerations in favour of the postulate of recovery. The remaining 'highly relevant contribution', that of Nayak, was published after *CTTC*. I shall end this reply, therefore, with an apology for my reprehensible lack of clairvoyant powers in my coverage of the relevant literature.

Acknowledgements

I would like to thank David Papineau for editorial suggestions that helped conserve original *Sinn* while tinting the *Färbung*. Thanks also to those who participated in the interdisciplinary Mechanization of Inference Seminar at the Ohio State University in Autumn Term 1996—especially Harvey Friedman and Victor Marek.

Department of Philosophy
Ohio State University

References

Alchourrón, C., Gärdenfors, P. and Makinson, D. [1985]: 'On the Logic of Theory

Change: Partial Meet Contraction and Revision Functions', *Journal of Symbolic Logic*, **50**, pp. 510–30.

Friedman, H. and Tennant, N. [1997]: 'Minimal Impliers and the Theory of Theory Contraction', in preparation.

Fuhrmann, A. [1991]: 'Theory Contraction through Base Contraction', *Journal of Philosophical Logic*, **20**, pp. 175–203.

Hansson, S. O. [1993a]: 'Theory Contraction and Base Contraction Unified', *Journal of Symbolic Logic*, **58**, pp. 602–25.

Hansson, S. O. [1993b]: 'Reversing the Levi Identity', *Journal of Philosophical Logic*, **22**, pp. 637–99.

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

Makinson, D. [1987]: 'On the Status of the Postulate of Recovery in the Logic of Theory Change', *Journal of Philosophical Logic*, **16**, pp. 383–94.

Makinson, D. [1995]: Review of Tennant [1994] in *Mathematical Reviews*, no. 95i:03065.

Nayak, A. [1994]: 'Foundational Belief Change', *Journal of Philosophical Logic*, **23**, pp. 495–533.

Tennant, N. [1994]: (*CTTC*): 'Changing the Theory of Theory Change: Towards a Computational Approach', *British Journal for the Philosophy of Science*, **45**, pp. 865–97.

Tennant, N. [1996]: 'The Staining Algorithm for Theory Contraction', available as a downloadable **.dvi** file from http://www.cohums.ohio-state.edu/philo/tennant_pubs.html.

Tennant, N. [1997]: 'On Having Bad Contractions: or, No Room for Recovery', *Journal of Applied Non-Classical Logics*, **7**, pp. 241–66.