The Rational Mind

Scott Sturgeon

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2020

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019950787

ISBN 978–0–19–884579–9

DOI: 10.1093/oso/9780198845799.001.0001

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

Contents

5

The Belief Model (AGM)

5.1 Preview

In this chapter we lay out one of the best-known formal models of coarse-grained epistemic attitudes. The model was invented by Carlos Alchourrón, Peter Gärdenfors, and David Makinson in a series of letters they sent to one another in the 1970s. Eventually they realized that their independently developed thoughts about rationality were converging. The resulting formalism is the Belief Model (also known as the AGM Model).

Like other models of rationality, the Belief Model has three moving parts: a theory of states, a transition theory, and a linking theory. The model uses a particular formalism to represent rational attitudes at a time, builds on that formalism to represent rational shifts between attitudes across time, and locates a kind of epistemic commitment which forges an internal link between the two stories. First we'll look at its theory of states, then we'll unpack the transition theory, and then we'll clarify how the model sees the two things fitting together.

5.2 The Belief Model's Theory of States

The Belief Model recognizes no matter of degree when it comes to epistemic attitudes. Basically the thought is that one either believes something or one does not. On this kind of view, epistemic attitudes can be usefully indexed in a couple of different ways. On the one hand, we might index a configuration of rational beliefs, say, by appeal to *what* is believed in the configuration; and were we to do so it would be natural to index a configuration of beliefs with a set of propositions (i.e. a set of things believed). On the other hand, we might index a configuration of beliefs by appeal to sentences used to express what is believed, in which case a configuration of beliefs would be naturally indexed by the collection of sentences. The Belief Model takes the latter approach.

Think of it this way. Suppose you are asked to say something which you believe, and in reply you sincerely utter sentence S. Next you are asked to say something else you believe, so you sincerely reply with sentence S*. Let the process continue until you have nothing left to say. At that point gather up all the sentences you've used and put them into a set $\underline{B}$. Then everything you believe is expressed by a member of $\underline{B}$, and every member of $\underline{B}$ expresses something you believe. Intuitively, for any proposition $\boldsymbol{p}$: you believe $\boldsymbol{p}$ exactly when there is a sentence S such that (i) S expresses $\boldsymbol{p}$, and

The Rational Mind. Scott Sturgeon, Oxford University Press (2020). © Scott Sturgeon.
DOI: 10.1093/oso/9780198845799.001.0001

(ii) S is a member of B. In a good sense, then, B captures your configuration of beliefs at the time in question.

Yet B isn't a belief set in the technical sense of the Belief Model, for it is insufficiently ideal. Like the Bayesian model of chapter 2, the Belief Model of this chapter makes idealizing assumptions when choosing its formalism. This amounts to ignoring certain aspects of real-world cognition (about which we'll have more to say in Chapter 13). In the present case, the Belief Model idealizes away from two aspects of real-world cognition. It refuses to allow a belief set to contain logical conflict between its members—even tacit conflict—and it refuses to allow a belief set to be logically incomplete in a certain way. Consider these idealizations in turn.

First, the Belief Model idealizes away from logical conflicts which plague ordinary thought. When such problems infect a thinker's beliefs, those beliefs cannot all be true: logic ensures at least one of them is mistaken. Such conflict may be deep and/or difficult to see but it is there as a matter of logic (i.e. logic alone is sufficient to ensure the conflict is present); and when that happens, a given thinker cannot have correct beliefs through and through, for she is bound (by logic) to have made a mistake. The Belief Model idealizes away from such bother, presuming that rational sets of belief logically hang together. The model requires that such sets be logically consistent:

(Con) Belief sets are logically consistent.

Second, the Belief Model precludes a certain kind of logical incompleteness in its belief sets. The model idealizes away from a certain sort of open-endedness of ordinary thought. To see the point clearly recall your belief set B. As it happens you do not accept everything logically entailed by things you believe, for like everyone else you have not had time (or inclination) to work everything out. This means that your belief set B is in a certain sense not fully logical—or perhaps better put: it is not fully informed by logic. Your belief set does not contain all that is logically entailed by its members. From a logical point of view that set would be improved were it to contain everything logically entailed by things you believe. After all, your beliefs would then be logically complete.

The Belief Model ignores such incompleteness and presumes that rational agents have chased down all logical implications of things they believe. The model presumes that their belief sets are logically complete:

(Entail) For any belief set **B** and sentence S: if the members of **B** logically entail S, then S belongs to **B**.

This principle guarantees that belief sets are fully logical in the following sense: there is no way to break out of them by chasing down the logical consequences of their members. Everything entailed by a given belief set is already in that set. The technical way to express this is to say that belief sets are 'closed' by logical consequence. But all that means is that belief sets satisfy the requirement (Entail). The model insists that everything entailed by members of a given belief set is also a member of that set.

Principles (Con) and (Entail) are idealizing constraints in the Belief Model. They directly echo powerful ideas emphasized over a century ago by William James. In particular, James emphasized that at least two goals seem to be deeply connected with our cognitive practice. In some sense we aim in thought to

(T) Believe truth.

But we also aim in thought to

(F) Avoid error.[1]

James pointed out that these are different goals. The acquisition-of-truth goal (T) is maximally satisfied by believing everything, after all—and so everything true—while the avoidance-of-error goal (F) is maximally satisfied by suspending judgement in everything, i.e. by never believing or rejecting anything at all. These two cognitive goals, then—(T) and (F)—pull us in opposite directions. They tug against one another in our practice.

Intuitively: the value of cottoning onto truth encourages epistemic risk, while the disvalue of stumbling into error counsels epistemic caution. Our cognitive practice pulls us in opposite directions, then, in the sense that one of its aims motivates risk-taking while another of its aims motivates intellectual caution. Epistemic rationality seeks a good mix of risk and caution, one which well reflects the value of truth and the disvalue of error.

The demands of logical consistency and closure—(Con) and (Entail) above—fall out of these twin goals. One way to take (F) to a logical extreme, after all, is to hear it as saying that we should not accept a claim if it is guaranteed by logic to be false. But an even stronger reading hears it as saying that we should not accept a collection of claims if logic ensures that they jointly contain an error. The first reading counsels one never to accept a logical contradiction, the second never to accept a collection of claims which harbours logical conflict in its members. This latter thought is exactly the content of (Con), so the consistency requirement can be seen as the demand to avoid error taken to a logical limit.

Similarly, one way to take the truth goal (T) is to hear it as saying that we should accept anything which is guaranteed by logic to be true. But an even stronger reading would hear it as saying that we should accept anything logically implied by things we accept. The first reading counsels one to accept logical truths, so to say; the second reading counsels one to accept logical consequences of things accepted. This latter thought is exactly the content of (Entail). Hence the model's closure requirement can be seen as the demand to believe truth taken to a certain kind of logical limit.

Yet one thing is clear: neither the consistency requirement (Con) nor the closure requirement (Entail) is satisfied by any real person. Indeed neither of these requirements *can* be satisfied by any real person. Our cognitive limits preclude our doing so: any being with a rich and changing set of beliefs like ours, who also satisfies the consistency requirement (Con), or the closure requirement (Entail), thereby fails to be recognizably human in her cognition. Any such being is recognizably a super-human thinker, someone whose cognition goes far beyond the reach of ordinary folk.

This prompts an obvious question: how could it be a necessary condition on human rationality—even full human rationality—that our epistemic attitudes satisfy the consistency requirement (Con) or the closure requirement (Entail)? After all, it is

[1] For an argument that these goals underpin the accuracy-based epistemology of credence, see my 'Epistemology, Pettigrew Style' (*Mind,* 2018).

not humanly possible to satisfy either of these requirements. The attendant thought here—once the question has been posed explicitly—is that it is *not* a necessary condition on human rationality—even full human rationality—that our epistemic attitudes satisfy requirements like (Con) or (Entail). The thought is that it is not possible for humans to satisfy such conditions, and, for this reason, doing so cannot be a requirement on rationality, even full rationality.

This line of thought is grounded in a Kantian idea:

(Kant) One ought rationally to satisfy condition C *only if* one can satisfy condition C.

Approaches to rationality which aim to respect this Kantian principle insist that the bounds of human rationality—even full human rationality—do not outstrip the bounds of human possibility. On this way of thinking about things, a configuration of epistemic attitudes is rational only if it is a humanly-possible configuration of attitudes, i.e. only if it is possible for humans to manifest that configuration of attitudes.

The Belief Model resolutely rejects this Kantian perspective. It insists that full rationality places requirements on attitudes which human cannot satisfy. Hence the theory is meant to detail various constraints on epistemic attitudes which are not within our power to satisfy. Those constraints are constitutive of what we might think of as super-human rationality, a kind of rationality manifested only by super-human agents, a kind of rationality to which real-world agents can aspire, perhaps, but never properly manifest. Just as our actions in a lifetime can only aspire to be morally perfect, the thought would be—with no real human being capable of leading a morally perfect life—so it is with our epistemic lives: no human can adopt a configuration of attitudes in line with the Belief Model.[2]

With such preliminaries in place, then, it is easy to state the model's constraints on the configuration of epistemic states. At any given moment, the model insists that an agent's attitudes shouldbe consistent and closed by logic. Put another way: at any given time, a rational configuration of states should be modelled by a belief set in the technical sense.

5.3 The Belief Model's Transition Theory

The Belief Model's transition theory is not nearly so simple; but like its theory of states it's best to sneak up on the model's transition theory by appeal to the underlying psychology it presupposes. So recall that the model kicks off, like a great deal of everyday practice, by seeing coarse-grained epistemic attitudes as a three-part affair: either you believe something, disbelieve it, or suspend judgement. The Belief Model recognizes three attitudes you might take to a proposition P. In shifting your take on P, therefore, the model (and much of commonsense) requires that there be mental movement *from* one of these three attitudes *to* one of the others.

[2] Chapter 13 discusses a weakening of (Kant) which proves central to rationality.

The underlying psychology recognizes no other shift in view. Hence there are exactly six ways to adjust your take on P:

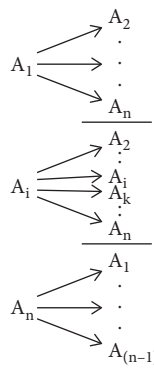
1. Shift from belief to disbelief
2. Shift from belief to suspended judgement
3. Shift from suspended judgement to belief
4. Shift from suspended judgement to disbelief
5. Shift from disbelief to belief
6. Shift from disbelief to suspended judgement.

Within the Belief Model framework a shift in one's take on P occurs only if one of these six changes takes place. But note they are incompatible: no two of them can happen at once. There can be a model-recognized shift in one's take on P, therefore, only if *exactly* one of the six changes takes place.[3]

This suggests the Belief Model transition theory will have six rules for changing one's mind: one for when belief shifts to disbelief, one for when belief shifts to suspended judgement, and so forth. But that is not how it goes. In fact the Belief Model endorses just *one* kind of non-trivial change: the model boils everything down to logical changes plus a single non-logical rule; and it does this by making a series of alignment assumptions. These are claims to the effect that the presence of one attitude lent to a content perforce aligns with a potentially different attitude lent to a potentially different content.

The model uses five such alignment assumptions. Three of them look initially plausible—at least as idealizations—and they permit the reduction of changes (a) thru (f) down to three types of shift in view. A pair of further alignment assumptions are then brought into play in order to line up the three remaining types of shift with one *uber*-rule. But neither of these further alignment assumptions looks at all obvious or commonsensical, even as idealizations. They do make for a highly simple

[3] More generally: suppose you can take a finite number of belief-like attitudes A1,..., An to a proposition P. Then you have these possible shifts in view concerning P:



For each belief-like attitude Ai that you might take to a proposition P, there are (n–1) other belief-like attitudes you might shift Ai into. When we assume there are n belief-like attitudes at your disposal, therefore, there are n(n–1) possible shifts in view.

theoretical foundation, though, a view on which all rational shift of opinion is defined by logic plus a single non-logical rule.

This is good work if you can get it. In the next chapter it will argued that the work is not to be had, but in this chapter we merely lay out the manner in which the Belief Model bases its transition theory on a single non-logical rule.

Alignment Assumption 1

The Belief Model treats disbelief as if it is belief in negation, marking no difference between *dis*belief in P and belief in ¬P. In effect the model assumes that disbelief is definable by appeal to belief and negation, so its first alignment assumption is

(AA_D) Disbelief in P always aligns with belief in ¬P.

For instance: disbelief in the claim that grass is green is treated by the model as if it is belief in the claim that grass is not green; disbelief in the claim that Oswald acted alone is treated by the model as if it is belief in the claim that Oswald did not act alone; and so forth. The Belief Model insists that disbelief and belief in negation stand or fall together.

This might be so for at least two reasons. Deep down it might turn out that disbelief is really nothing more than belief in negation. Or it might turn out that the notionally-distinct attitudes are psychologically distinct, deep down, but that rational agents disbelieve exactly when they believe a related negation. On this latter view, disbelief is a psychological attitude over and above belief in negation; but rationality requires that an agent invests disbelief exactly when she believes the relevant negation. We shall take no stand here on such a non-reductive endorsement of (AA_D). The next chapter resists the reduction of disbelief to belief in negation.

Alignment Assumption 2

The Belief Model treats belief in a doubly negated content as if it is belief in that which is doubly negated, marking no difference between belief in ¬¬P and belief in P. In essence the model assumes that belief in a double negation is definable by appeal to belief itself. Its second alignment assumption is

($AA_{\neg\neg}$) Belief in ¬¬P always aligns with belief in P.

For instance: belief in the claim that it is not the case that it is not the case that grass is green is treated by the model as if it is belief in the claim that grass is green; belief in the claim that it is not the case that it is not the case that Oswald acted alone is treated by the model as if it is belief in the claim that Oswald acted alone; and so forth. The Belief Model insists that belief in a double negation and belief in what is so negated stand or fall together.

Again this might be so for at least two reasons. Deep down it might turn out that belief in a double negation is really nothing more than belief in what is so negated. Or it might turn out that the notionally distinct states are psychologically distinct, deep down, but that rational agents believe a double negation exactly when they believe what is so negated. We shall take no stand here on either of these routes to ($AA_{\neg\neg}$). The next chapter presents prima facie worries for both lines of thought.

Alignment Assumption 3

The Belief Model treats suspended judgement as if it is the absence of belief and disbelief, marking no difference between suspending judgement in P and failing both to believe and to disbelieve P. In essence the model assumes that suspended judgement is definable by appeal to belief and negation. Its third alignment assumption is

(AA_S) Suspended judgement in P always aligns with failure of both belief in P and disbelief in P.

For instance: suspended judgement in the claim that grass is green is treated by the model as if it is avoidance of belief and disbelief in the claim that grass is green; suspended judgement in the claim that Oswald acted alone is treated by the model as if it is avoidance of belief and disbelief in the claim that Oswald acted alone; and so forth. The Belief Model insists that suspended judgement and the absence of belief/disbelief stand or fall together.

This too might be so for at least two reasons. Deep down it might turn out that suspended judgement is really nothing more than the joint absence of belief and disbelief. Or it might turn out that suspended judgement is something more that such an absence, but, that rational agents suspend judgement exactly when they neither believe nor disbelieve. We shall take no stand here on such a non-reductive endorsement of (AA_S). The next chapter resists a reductive approach to suspended judgement.

Now, recall the six types of shift in view mentioned earlier:

1. Shift from belief to disbelief
2. Shift from belief to suspended judgement
3. Shift from suspended judgement to belief
4. Shift from suspended judgement to disbelief
5. Shift from disbelief to belief
6. Shift from disbelief to suspended judgement.

Alignment assumptions concerned with disbelief and suspended judgement permit (a) thru (f) to be rewritten solely in terms of belief and negation. As follows:

applying (AA_D) to (a) yields

(a)* Shift belief in P to belief in ¬P;

applying (AA_S) to (b) and then (AA_D) to the result yields

(b)* Shift belief in P to neither belief in P nor belief in ¬P;

applying (AA_S) to (c) and then (AA_D) to the result yields

(c)* Shift neither believing P nor believing ¬P to belief in P;

applying (AA_S) to (d) and then (AA_D) twice to the result yields

(d)* Shift neither believing P nor believing ¬P to belief in ¬P;

applying (AA_D) to (e) yields

(e)* Shift belief in ¬P to belief in P;

and applying (R_S) to (f) and then (AA_D) twice to the result yields

(f)* Shift belief in ¬P to neither belief in P nor belief in ¬P.

In a nutshell: the alignment assumptions (AA_D) and (AA_S) reduce disbelief and suspended judgement to belief plus negation. This means (a)-(f) can be rewritten as (a)*-(f)*. Shift in disbelief and suspended judgement boils down to shift in belief plus negation. That is ensured by the model's alignment assumptions for disbelief and suspended judgement.

Further still: the placeholder in the starred rules above—namely P—can go for any proposition whatsoever. So the alignment rule ($AA_{\neg\neg}$) can be used to reduce the starred principles to *three* kinds of change. Specifically:

- ($AA_{\neg\neg}$) implies (a)* and (e)* are instances of a single kind of change: shift from belief to opposing belief. The Belief Model calls that ***revision***.
- ($AA_{\neg\neg}$) implies (b)* and (f)* are instances of a single kind of change: shift from belief to suspended judgement. The model calls that **contraction**.
- ($AA_{\neg\neg}$) implies (c)* and (d)* are instances of a single kind of change: shift from suspended judgement to belief. The model calls that **expansion**.

In a nutshell: the Belief Model begins with the common-sense thought that one believes, disbelieves, or suspends judgement. This creates six ways to shift one's opinion. The model also accepts alignment assumptions for disbelief, double negation and suspended judgement. Those assumptions boil down the six types of shift-in-view to three: revision, contraction, and expansion. All of these transitions begin with a set of beliefs which is consistent and closed by logic. Revision is then shifts from belief to opposing belief, contraction is then shift from belief to suspended judgement, and expansion is then shift from such judgement to belief.

Let's formalize all this in the style of the model.

Suppose you are in an epistemic state characterized by belief set **B**. Let ¬P be a member of **B** (i.e. you believe ¬P). Then *the revision of **B** by P* is the shift from **B** to a new belief set containing P. The model names this sort of transition '(**B***P)', so we have

(**B***P) = the result of revising **B** by P.

Intuitively, revision of belief occurs when you are rationally turned entirely around on a claim: first you endorse it, then you come to endorse its logical opposite (and then you settle everything else accordingly).

But suppose you start out believing P rather than believing its negation. Then *the contraction of **B** by P* is the shift from **B** to a new belief set not containing P. The model names this sort of transition '(**B**-P)', so we have

(**B**-P) = the result of contracting **B** by P.

Intuitively, contraction of belief occurs when you rationally give up a commitment to a claim: first you endorse it, then you fail to endorse or to reject (and then you settle everything else accordingly).

Finally, suppose neither P nor ¬P starts out in **B**. By the model's lights you begin with suspended judgement in P, neither believing nor disbelieving. Then *the expansion of **B** by P* is the shift from **B** to a set containing P. The Belief Model names this sort of transition '(**B**+P)', so we have

(**B**+P) = the expansion of **B** by P.

Intuitively, expansion of belief occurs when you rational commit to a claim: first you have no such commitment, then you endorse the claim in question (and then you settle everything else accordingly).

The Belief Model takes revision, contraction and expansion as explanatorily fundamental. Once we get clear on the alignment claims (AA_D), ($AA_{\neg\neg}$) and (AA_S)—once we get clear, that is to say, on the alignment of disbelief and suspended judgement to belief plus negation—it turns out there are three notionally basic shifts in view: revision, contraction, and expansion. It comes as a shock, then, to see that the Belief Model boils down these three types of transition to shifts defined by logic plus one non-logical rule. This further bit of the model is not the slightest bit obvious, so we'll need a good argument for it if we're to endorse the overall view. Our next task is to explain the rest of the model. In the next chapter we'll examine it critically.

5.4 Further Claims About Rational Transitions

At the end of the day the Belief Model endorses a single kind of non-logical belief change. The key to its simple transition theory comes via two further claims: one concerns expansion, the other concerns revision. We'll take them in order and then spend a good deal of time explaining the non-logical rule endorsed by the model.

The root idea behind the model's transition theory is *epistemic conservatism*. This principle has it, roughly, that a rational shift in view should be minimal, that it should be the least drastic change in belief called for by the required epistemic transition. Once such a principle is accepted—and it seems entirely natural to accept it, of course—it is also natural to say three further things: one about rational expansion, one about rational contraction, and one about rational revision. Namely:

- *Expansion* should be the minimal shift in belief brought on by adding something about which one had suspended judgement.
- *Contraction* should be the minimal shift in belief brought on by suspending judgement in something previously believed.
- *Revision* should be the minimal shift brought on by accepting something reviously rejected.

These principles capture the philosophical spirit of the Belief Model's transition theory, and they look entirely sensible on their face. As we saw in the previous chapter, moreover, a principle very like the three above captures the spirit of formal work on the rational shift of fine-grained epistemic attitudes. Epistemic conservatism underwrites most work on rational shift in view.

Now, in this conservative spirit the model puts forward the idea that expansion is defined by logic. The expansion of **B** by P is set equal to the set got by adding P to **B** and then simply tossing in all the new logical consequences. We have the following rule for expansion:

$$(+) \quad (\mathbf{B} + \mathrm{P}) = \{\Phi : \{\mathbf{B} \cup \{\mathrm{P}\}\} \Rightarrow \Phi\}.$$

More in English: the expansion of **B** by P is the set got by adding P to **B** and then closing the result by logic. This is the set of claims Φ such that Φ is entailed by **B**-members together with P.

This is a very simple idea. Recall that expansion is meant to be a minimal shift from suspended judgement to belief. Recall too that states of belief are assumed to be consistent and closed by logical implication. If such an idealized initial state involves suspended judgement in P, then, adding P will generate no conflict in the result. After all, the initial state is consistent and closed by implication, but contains neither P nor its negation. So the initial state does not contain claims which logically entail P or its negation. Adding P to that initial state thus yields no logical difficulty. Principle (+) then insists that the expansion of **B** by P should result in the set you get when you add P to **B** and then toss in all the resulting logical consequences.

The idea is motivated by epistemic conservatism: expansion is meant to be the minimal shift in view brought on by adding something about which one had previously suspended judgement. Since the Belief Model construes all rational configurations of belief as closed by logical implication—i.e. they all obey the closure rule (Entail)—the logical consequences of P plus one's initial beliefs must result from coming to accept P. But rock-ribbed epistemic conservatism then counsels adding no more than one logically must after that. When expanding by P, then, the model's rule (+) says that one should accept P along with everything one previously accepted, plus anything new which logically follows in light of coming to belief P, but no more than this. If that is right, however, then rational expansion may be defined by logic. It is simply the addition of new content plus whatever follows from the resulting enriched configuration of belief. This is the model's first claim about rational state transition which does not seem initially obvious. In the next chapter we'll see why that is so.

The second such claim concerns rational revision. Recall that this sort of transition involves the move from rejection to acceptance. Since the model's belief sets are logically consistent, revision will oblige *taking things out* of your belief set as well as putting things in. The model then appeals to epistemic conservatism to motivate a natural thought, namely, that the revision of belief set **B** by claim P is itself a two-step affair: first one removes the minimal from **B** so as to preclude a commitment to ¬P, then one inserts the minimal to the resulting impoverished set of commitment to ensure a commitment to P. But recall that the contraction of **B** by P is itself meant to be the first of these steps, while expansion of the result by P is itself meant to be the second of these steps. This leads the model to endorse what is known as the *Levi identity*:

$$(*) \quad (\mathbf{B}*\mathrm{P}) = ((\mathbf{B} - \neg \mathrm{P}) + \mathrm{P}).^{4}$$

This equation sets the revision of **B** by P equal to the upshot of a two-step process: first the contraction of **B** by ¬P, then the expansion of the result by P.

Now, recall that rational revision is meant to be the minimal shift from belief to opposing belief. Rational revision by P is thus meant to be the minimal shift from disbelief in P to belief in P. Rule (*) says that such a minimal shift can be got by contracting and then by expanding. On this view the shortest cognitive route *from* a belief set committed to ¬P *to* a belief set committed to P involves subtracting ¬P and then adding P. If this is right, revision boils down to contraction plus expansion.

[4] See (Levi, 1977), or (Gärdenfors, 1988).

That is the Belief Model's second claim about rational shift in view which does not seem initially obvious. In the next chapter we'll see why that is so.

In a nutshell, then, the model makes two unobvious claims about rational state transition. Rule (+) says that rational expansion boils down to logic. Rule (*) says that rational revision boils down to contraction and then expansion. Together these rules imply that rational revision boils down to contraction plus logic, which means that *every type of rational shift in view boils down, ultimately, to contraction plus logic*. In this way the Belief Model endorses an expansion function defined by logic, a non-trivial contraction function yet to be glossed, and their iteration in reverse order. In a good sense, then, the model's transition theory boils down to one non-logical bit of theory: contraction. That is our next topic.

5.5 Contraction

We have noted that there is really a single driving force behind the Belief Model's approach to rational shift in view: epistemic conservatism. Gärdenfors calls this 'the principle (or criterion) of informational economy'; and he makes a number of non-trivial claims about it. In particular, he puts forward the follow thoughts about epistemic conservatism:

> The key idea is that, when we change our beliefs, we want to retain as much as possible of our old beliefs—information is in general not gratuitous, and unnecessary losses of information are therefore to be avoided. (Gärdenfors 1988: 49)

> The criterion of informational economy demands that as few beliefs as possible be given up so that the change is in some sense a *minimal* change of **B** to accommodate P . . . there are in general several ways of fulfilling the minimality requirement. (ibid.: 53)

> When forming the contraction (**B**-P) of **B** with respect to proposition P, the principle of informational economy requires that (**B**-P) contain as much as possible from **B** without entailing P. (ibid.:75)

We can capture the vibe here with the principle of information economy:

(Info-Econ) Shift in view should avoid needless loss of belief.

Given the idealizing assumptions of the Belief Model, this principle can look quite compelling. After all, it is a presupposition of the model that one starts in a rational belief state. As Gärdenfors puts it: belief sets are immune to 'forces of internal criticism'; so whenever an external force induces a rational shift in view, it seems plausible that that shift should itself echo its cause but no more than its cause. In other words, the rational shifts in view should be exactly commensurate with their causes or prompts. They should not be over-reactions. They should involve only the minimal perturbance of an equilibrium position.

Who could argue with that?

Well, no one really. But it is natural to ask for clarification. In particular, it is natural to ask for clarification concerning the notion of minimality at work in the foregoing line of thought. Since the principle of informational economy is given a non-trivial role in the Belief Model, we need a clear specification of how minimal change is to be understood.

That specification is found in Gärdenfors's final quote above. Read literally, the idea is that the contraction of **B** by P should contain all of **B** it can while remaining a belief set and avoiding P. Or in other words: (**B**-P) should retain as much of **B** as possible while being consistent, closed by logic, yet failing to retain P. On this view, the contraction of **B** by P is the minimal perturbance of **B**-members brought on by the need to remove P yet preserve consistency and closure.

To formalize this it helps to use notions from set theory. So forget belief sets for a moment and think of sets more generally. We say that set **S** is a *subset* of **S***—written $\mathbf{S} \subseteq \mathbf{S}^*$–when everything in **S** is also in **S***. Formally put

$$(\subseteq) \quad \mathbf{S} \subseteq \mathbf{S}^* \text{ iff } (\forall x)[(x \in \mathbf{S}) \supset (x \in \mathbf{S}^*)].$$

And we say **S** is a *proper subset* of **S***—written $\mathbf{S} \subset \mathbf{S}^*$—when everything in **S** is also in **S*** but **S*** contains members not in **S**. In other words, **S** is a proper subset of **S*** when two things happen: **S** is a subset of **S***; and **S*** is not a subset of **S**. The second condition obliges something to be in **S*** but not in **S**. Formally put

$$(\subset) \quad \mathbf{S} \subset \mathbf{S}^* \text{ iff } \{(\mathbf{S} \subseteq \mathbf{S}^*) \,\&\, (\exists x)[(x \in \mathbf{S}^*) \,\&\, (x \notin \mathbf{S})]\}.$$

Of course belief sets are special kinds of sets: they have logically consistent members and they are closed by logical implication: belief sets harbor no conflict and anything implied by their members is itself a member.

This means the contraction of **B** by P must do more than remove P. It must also ensure that no **B**-claims which remain in the contraction jointly imply P. For belief sets are closed by logic: anything implied by their members is in them. So the Belief Model faces a general question about contraction: when contracting a belief set **B** by a claim P, what more than P should be removed from **B**?

This is one place where the principle of informational economy plays a crucial role. Think back to final quote from Gärdenfors:

> When forming the contraction (**B**-P) of **B** with respect to proposition P, the principle of informational economy requires that (**B**-P) contain as much as possible from **B** without entailing P.

The idea is that the principle of informational economy should guide retraction. When contracting **B** by P, for instance, and deciding what must be taken out of **B** in addition to P, one should be maximally conservative when tossing things out. One should remove only what is obliged by the demands of consistency, closure, and the removal of a commitment to P. One should remove only what is demanded by logic in light of these needs. So the idea is basically this:

(**B**-P) = The belief set **B*** which is a subset of **B**, does not contain P, and is no proper subset of anything which also does both of these things.

More formally put, Gärdenfors's guiding thought is

(G) (**B**-P) = The belief set **B*** such that
(1) $\mathbf{B}* \subseteq \mathbf{B}$
(2) $P \notin \mathbf{B}*$
(3) $\neg(\exists \mathbf{B}\wedge)[$(a) $\mathbf{B}* \subset \mathbf{B}\wedge$
(b) $\mathbf{B}\wedge \subseteq \mathbf{B}$
(c) $P \notin \mathbf{B}\wedge]$.

It is important to realize that (G) is built from a certain take on what minimal shift in view amounts to, i.e. a certain understanding of informational economy. Two ideas ground clauses in principle (G):

1. Informational economy entails that contraction leads to a subset of what is contracted;
2. **X** is less radically shifted from **Z** than is **Y** iff **X** and **Y** are subsets of **Z**; but **Y** is a *proper* subset of **X**.

Clause (1) of principle (G) itself grows from 1 while clause (3) of that principle itself grows from 1 and 2 together. The net result is a view on which the rational contraction of **B** by P is the belief set which is a subset of **B**, removes P from **B**, yet is no proper subset of anything which manages also to do both those things.

To get a feel for what is going on here, we should test the guiding thought in the neighbourhood with some tinker-toy examples. To that end, let **Con**(**S**) be the set of logical consequences of **S**:

$$\underline{\text{Con}}(\mathbf{S}) \;=\; \{\text{logical consequences of } \mathbf{S}\}.$$

Assume all the sets we are now dealing with are consistent; and let **B** be the belief set consisting of P, Q and their logical consequences:

$$\mathbf{B} = \underline{\text{Con}}(\text{P,Q}).$$

B is a belief set in the technical sense of the Belief Model. It captures the configuration of belief which endorses P, Q, and all that follows logically from them.

Now consider the contraction of **B** by P. In our notation this is

$$[\underline{\text{Con}}(\text{P,Q})] - \text{P}.$$

It seems pre-theoretically plausible that this contraction should be simply the consequences of Q by itself (i.e. **Con**({Q}). But it also looks as if the set of those consequences itself plays the **B***-role in the Gärdenfors principle (G), which in turn looks to confirm the guiding thought behind (G). More formally put, when **B** is **Con**({P,Q}) we have:

Con({Q}) = The belief set **B*** such that
(1) **B**∗ ⊆ **B**
(2) P ∉ **B**∗
(3) ¬(∃**B**^)[(a) **B**∗ ⊂ **B**^
(b) **B**^ ⊆ **B**
(c) P ∉ **B**^].

And this suggests that principle (G) captures the idea of minimal perturbance in a belief set brought on by the need to remove a given claim while retaining consistency and closure.

Having said that, take a second look at Gärdenfors's guiding thought about contraction:

(G) (**B**-P) = The belief set **B*** such that
(1) **B**∗ ⊆ **B**
(2) P ∉ **B**∗
(3) ¬(∃**B**^)[(a) **B**∗ ⊂ **B**^
(b) **B**^ ⊆ **B**
(c) P ∉ **B**^].

This principle assumes that there exists a unique belief set which plays the $\mathbf{B}^*$-role. That cannot be taken for granted. Although it is true in some cases, as we've just seen, it is not true in others. Indeed (G)'s existence and uniqueness assumptions can each fail: there may be no belief set to play the $\mathbf{B}^*$-role in (G), or there may be more than one set to play it. Everything depends on the details of the case.

1. For instance, take any belief set $\mathbf{B}$ and any truth of logic L. Consider the contraction of $\mathbf{B}$ by L. There is simply no way to remove L from $\mathbf{B}$ yet keep the result closed by logic, for L is a consequence of logic. This means that L is a member of every belief set and thus no such set satisfies (G2): no belief set fails to contain L. In this sort of case the existence assumption of (G) fails: when L is a truth of logic, there is no belief set to play the $\mathbf{B}^*$-role in (G).

2. Other cases involve a failure of (G)'s uniqueness assumption. Let $\mathbf{B}$ equal the consequences of X and Y:

$$\mathbf{B} = \underline{\mathbf{Con}}(\{\mathrm{X},\mathrm{Y}\}).$$

Let P be the conjunction of X and Y:

$$\mathrm{P} = \mathrm{X}\&\mathrm{Y}.$$

Then P is a member of $\mathbf{B}$, of course, since it is a trivial consequence of X and Y together. But consider the contraction of $\mathbf{B}$ by P. Here there are *two* sets which play the $\mathbf{B}^*$-role in (G): one is $\underline{\mathbf{Con}}\{\mathrm{X}\}$, the other is $\underline{\mathbf{Con}}\{\mathrm{Y}\}$. We have

(1) $\underline{\mathbf{Con}}\{\mathrm{X}\} \subseteq \mathbf{B}$
(2) $\mathrm{P} \notin \underline{\mathbf{Con}}\{\mathrm{X}\}$
(3) $\neg(\exists \mathbf{B}\wedge)[$(a) $\underline{\mathbf{Con}}\{\mathrm{X}\} \subset \mathbf{B}\wedge$
(b) $\mathbf{B}\wedge \subseteq \mathbf{B}$
(c) $\mathrm{P} \notin \mathbf{B}\wedge]$.

And we also have

(1) $\underline{\mathbf{Con}}\{\mathrm{Y}\} \subseteq \mathbf{B}$
(2) $\mathrm{P} \notin \underline{\mathbf{Con}}\{\mathrm{Y}\}$
(3) $\neg(\exists \mathbf{B}\wedge)[$(a) $\underline{\mathbf{Con}}\{\mathrm{Y}\} \subset \mathbf{B}\wedge$
(b) $\mathbf{B}\wedge \subseteq \mathbf{B}$
(c) $\mathrm{P} \notin \mathbf{B}\wedge]$.

Nevertheless: $\underline{\mathbf{Con}}\{\mathrm{X}\}$ and $\underline{\mathbf{Con}}\{\mathrm{Y}\}$ are distinct sets. So in this case (G)'s uniqueness assumption fails: more than one set plays the $\mathbf{B}^*$-role in (G).

To sum up: for arbitrary $\mathbf{B}$ and P, the Belief Model wants the contraction of $\mathbf{B}$ by P to be the minimal perturbance of $\mathbf{B}$ brought on by the need to remove P while preserving consistency and closure. When smallness of change is measured set-theoretically—as in 1 and 2 above—a problem crops up: sometimes no such change exists (much less a minimal one), while other times multiple changes satisfy the criteria.

What to do?

Well, think back to the model's guiding thought:

(G) (**B**-P) = The belief set $\mathbf{B}^*$ such that
(1) $\mathbf{B}* \subseteq \mathbf{B}$
(2) $\mathrm{P} \notin \mathbf{B}*$
(3) $\neg(\exists \mathbf{B}^\wedge)[$(a) $\mathbf{B}* \subset \mathbf{B}^\wedge$
(b) $\mathbf{B}^\wedge \subseteq \mathbf{B}$
(c) $\mathrm{P} \notin \mathbf{B}^\wedge]$.

When a set of claims plays the $\mathbf{B}^*$-role in (G), it is a belief set by hypothesis. So any set which plays the $\mathbf{B}^*$-role in (G) is itself consistent and closed by logic. But any such set is also a subset of **B**, since it satisfies condition (1) of principle (G); and any such set is a maximal subset of **B** too, since it satisfies conditional (3) of principle (G). So any set which plays the $\mathbf{B}^*$-role in (G) is a *maximal non-P belief subset of **B***.

This is a mouthful we'd do well to shorten. When a set plays the $\mathbf{B}^*$-role in (G), then, let us say that it is a 'maximal subset of **B**', but always keep in mind that the set in question is a maximal non-P subset of **B**. OK, put all the maximal subsets of **B** together into a set:

$$(\mathbf{B}\perp\mathrm{P}) = \{\text{maximal subsets of } \mathbf{B}\}.$$

This set is composed of other sets, of course, since its members are sets which play the $\mathbf{B}^*$-role in (G). And we have seen that there are exactly three cases to consider when it comes to members of ($\mathbf{B}\perp$P): the case in which no sets play the $\mathbf{B}^*$-role in (G), the case in which precisely one set does so, and the case in which multiple sets play the $\mathbf{B}^*$-role in (G):

(A) ($\mathbf{B}\perp$P) = the empty set Ø (i.e. the set with no members at all).
(B) ($\mathbf{B}\perp$P) = a set with just one member.
(C) ($\mathbf{B}\perp$P) = a set with more than one member.

Eventually we'll see that the model treats the first two cases here as special instances of the treatment it gives to the third case; but superficially, at least, the model defines contraction rather differently in each of these cases. So it is helpful to consider them in turn.

Case (A). Here the set of maximal subsets of **B** is empty: there are no sets which are consistent, closed by logic, yet fail to entail P. But consider

$$\mathbf{L} = \{\text{logical truths}\}.$$

L is consistent and closed by logic, we shall suppose; so **L** is the ideal pure logician's belief set, capturing exactly the dictates of logic. Now ask yourself this: is P a member of **L**?

There are two sub-cases to consider: when P is a member of **L**, and when it is not. Suppose we are in the first sub-case: P is a member of **L**. Then it is clear why there are

no maximal subsets of **B**; for P is guaranteed by logic to be true, so no belief set can avoid a commitment to P. On the other hand, suppose we are in the second sub-case: P is not a member of **L**. Then intuitively we can add **B**-claims to **L** until logic demands that P be added as well. Just before reaching that point, however, our construction will be a maximal subset of **B**. But that contradicts the defining feature of Case (A), namely, that (**B**⊥P) is empty. So we may conclude:

$$[(\mathbf{B}\perp \mathrm{P}) = \emptyset] \text{ iff } \mathrm{P} \in \mathbf{L}.$$

In English: the set of maximal subsets of **B** is empty just when P is ensured by logic.

Suppose, then, that (**B**⊥P) is empty: there are no maximal sub-sets of **B**, i.e. no set plays the **B***-role in (G). When that happens the Belief Model says that contraction comes to nothing, as it were, for the model sets the contraction of **B** by P equal to the original set **B**. In effect the model says that in this particular Case—when there are no maximal subsets of **B**—the process of contraction puts one back into the very state that one started out in.

This aspect of the model could be viewed as its expression of the opinion that one cannot rationally retract something which is guaranteed by logic to be true. Or perhaps it is no more than a technical convenience. At this stage we needn't decide why the Belief Model treats retraction in this peculiar way; we need only note that when there are no maximal subsets of **B**, the model insists that the contraction of **B** by P is identical to **B** itself:

$$(\mathbf{B} - \mathrm{P}) = \mathbf{B}.$$

Case (B). Next suppose that the set of maximal subsets of **B** has exactly one member. There is a unique set **B*** such that

$$\begin{array}{ll} (1) & \mathbf{B}* \subseteq \mathbf{B} \\ (2) & \mathrm{P} \notin \mathbf{B}* \\ (3) & \neg(\exists \mathbf{B}\wedge)[(a)\ \mathbf{B}* \subset \mathbf{B}\wedge \\ & \qquad (b)\ \mathbf{B}\wedge \subseteq \mathbf{B} \\ & \qquad (c)\ \mathrm{P} \notin \mathbf{B}\wedge]. \end{array}$$

In this case

$$(\mathbf{B}\perp \mathrm{P}) = \{\mathbf{B}^*\},$$

so it's no surprise that the Belief Model sets the contraction of **B** by P equal to **B***:

$$(\mathbf{B}\text{-}\mathrm{P}) = (\mathbf{B}\perp \mathrm{P})\text{'s one member} = \mathbf{B}^*.$$

This reflects the model's commitment to the principal of informational economy, with the idea being that minimal information should be lost in contraction. But recall that the Belief Model measures relative amounts of information across sets with the subset relation. Hence it sees minimal damage to information after P-removal as some kind of maximal non-P subset. When the set of maximal subsets of **B** has exactly one member, then, it is unsurprising that the model sets the contraction of **B** by P equal to **B** itself.

Case (C). Finally suppose that the set of maximal subsets of **B** has more than one member. There are multiple sets $\mathbf{B^{1,2,3\ldots n}}$ which play the $\mathbf{B^*}$-role in (G). This means that for any such set $\mathbf{B^i}$:

(1) $\mathbf{Bi} \subseteq \mathbf{B}$
(2) $P \notin \mathbf{Bi}$
(3) $\neg(\exists \mathbf{B}\wedge)$[(a) $\mathbf{B^i} \subset \mathbf{B}\wedge$
(b) $\mathbf{B}\wedge \subseteq \mathbf{B}$
(c) $P \notin \mathbf{B}\wedge$].

Recall that the model measures loss of information with the subset relation. In this case there are several maximal subsets. Hence the model cannot use its standard take on informational economy to pin down how contraction should go. When there are several maximal subsets of **B**, and information is measured by the subset relation, the principle of informational economy does not pin down a unique solution to the contraction problem. It does not pin down a unique set to be the contraction of **B** by P.

What to do?

At this point the Belief Model makes appeal to a new resource. In effect it admits that not all belief sets are equal, that some are more *epistemically entrenched* than others. Gärdenfors has this to say about such entrenchment:

> This is a notion that [applies] to a single sentence . . . It is the epistemic entrenchment of a sentence in an epistemic state that determines the sentence's fate when the state is contracted . . . When a belief set **B** is contracted, the sentences in **B** that are given up are those with the *lowest* epistemic entrenchment . . . The fundamental criterion for determining the epistemic entrenchment of a sentence is how useful it is in inquiry and deliberation. Certain pieces of [information] are more important than others when planning future actions, conducting scientific investigations, or reasoning in general . . . The epistemic entrenchment of a sentence is tied to its explanatory power and its overall informational value within the belief set. (Gärdenfors 1988: 86–7)

The idea, then, is to use entrenchment to break ties between maximal subsets of **B**. When multiple sets satisfy the model's guiding thought (G)—that is to say, when multiple sets $\mathbf{B^{1,2,3\ldots n}}$ are such that for any of them $\mathbf{B^i}$:

(1) $\mathbf{B^i} \subseteq \mathbf{B}$
(2) $P \notin \mathbf{B^i}$
(3) $\neg(\exists \mathbf{B}\wedge)$[(a) $\mathbf{B^i} \subset \mathbf{B}\wedge$
(b) $\mathbf{B}\wedge \subseteq \mathbf{B}$
(c) $P \notin \mathbf{B}\wedge$],

entrenchment is used to break the tie. Epistemic entrenchment is used to build the contraction of **B** by P from the maximal subsets of **B**.

Here's how it goes. First one compares entrenchment of maximal subsets of **B**, with there being two sub-cases to consider: either one finds a maximal subset more entrenched than all others, or one does not. If there is a winner on the entrenchment front—i.e. a best-entrenched maximal subset of **B**—then, the Belief Model sets the contraction of **B** by P equal to that best-entrenched set. In this sub-case we have

(**B**-P) = The belief set $\mathbf{B}^{e}$ which is more entrenched than any other maximal subset of **B**.

If no maximal subset of **B** is uniquely best entrenched, however, there will be entrenchment ties amongst the maximal subsets of **B**. In this scenario there are multiple maximal subsets of **B** which are all equally well entrenched, and which are all more entrenched than other maximal subsets of **B**. In that case the Belief Model says the following about contraction:

(**B**-P) = {Φ: Φ belongs to *each* best-entrenched maximal subset of **B**}.

In other words: when there are multiple best-entrenched maximal subsets of **B**, the contraction of **B** by P is built from what they have in common.

So we can chase down the Belief Model's take on contraction with three questions. For any **B** and P

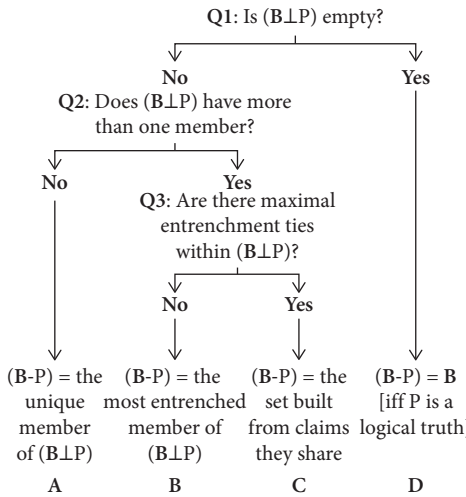


Figure 5.1

But notice that (A) is a special case of (B) and (B) is a special case of (C). This means that the model's treatment of contraction boils down to this thought:

(**B**-P) = **B**, when P is logical necessary
{what is common to the most entrenched members of (**B**⊥P)}, otherwise.

Once more set theory helps make this precise.

When **S** and **S*** are sets their *intersection* is the set containing members of both. This is written $\mathbf{S} \cap \mathbf{S}^{*}$:

$$(\cap) \quad \mathbf{S} \cap \mathbf{S}^{*} = \{x : x \text{ is a member of } \mathbf{S} \text{ and } \mathbf{S}^{*}\}.$$

When set **S** is composed of other sets $\mathbf{S}^{1,2,3 \cdots n}$, the *generalized intersection* of **S**—written ∩**S**—is the set of things common to all sets in **S**, i.e. common to all the $\mathbf{S}^{i}$s:

$$(\cap\mathbf{S}) \quad \cap\mathbf{S} = \{x : x \text{ is a member of } \mathbf{S^1}, \text{ and } \mathbf{S^2}, \text{ and ... and } \mathbf{S^n}\}.$$

Now think back to the set of maximal belief subsets of **B**, (**B**⊥P). We know that for any **B** and P

(A) (**B**⊥P) = the empty set Ø

or

(B) (**B**⊥P) = a set with exactly one member

or

(C) (**B**⊥P) = a set with more than one member.

And we have seen that in Case (A) the Belief Model sets the contraction of **B** by P equal to **B** itself, while in Case (B) the model sets the contraction of **B** by P equal to (**B**⊥P)'s sole member; but in Case (C), when more than one set satisfies the model's guiding thought (G), the model calls on the idea of epistemic entrenchment, collecting together the most-entrenched members of (**B**⊥P). The model then sets the contraction of **B** by P equal to the set of things those most-entrenched members of (**B**⊥P) have in common.

Let **max-E**[(**B**⊥P)] be the set of most entrenched maximal subsets of **B**. Note when it has just one member **B***:

$$\cap\mathbf{max\text{-}E}[(\mathbf{B}\perp P)] = \mathbf{B}^*.$$

This is why the Belief Model can treat Cases (B) and (C) in a single way, with the following contraction rule

(*) CONTRACTION

For any **B** and P: (**B**-P) = **B**, if (**B**⊥P) = Ø;
$\cap\mathbf{max\text{-}E}[(\mathbf{B}\perp P)]$, if (**B**⊥P) ≠ Ø.

In effect the contraction rule lays down a function from pairs of things—a belief set and one of its members—to other belief sets. And there is a technical name for this sort of function—a *partial meet contraction function*—but don't worry about what that means. Gärdenfors canvasses the idea nicely:

> The intuitive idea is that the [entrenchment function **max-E**] picks out the maximal subsets of (**B**⊥P) that are epistemologically most entrenched. Thus a proposition P is in (**B**-P) iff it is an element of all the epistemologically most entrenched maximal subsets of **B**. (Gärdenfors 1988: 80)

And now the guts of the model's transition theory are easy to state. It says that shift between epistemic states should boil down to expansion by logic, contraction, and their iteration in reverse order. That is all there is to it.

5.6 Postulates

The Belief Model's transition theory is built from three basic ideas:

1. When shifting view, don't give up something unless forced to by logic;
2. When shifting view, don't take on something unless forced to by logic;

3. Where ever possible measure the 'size' of what is given up or taken on by inclusion relations between belief sets.

These ideas drive the construction of the model in two different ways. They motivate the Levi identity, as we'll see, and they motivate various *Postulates* for revision, expansion, and contraction. These Postulates are rules which specify how a rational shift in view should proceed. In his landmark book *Knowledge in Flux* Gärdenfors proves that the Postulates and the Levi identity 'hang together' in a mutually supporting way (which we'll explain in a moment).

In constructing his Postulates, however, it is important to emphasize that Gärdenfors appeals time and again to informational economy. In turn this leads him to a system of rules for expansion, revision, and contraction. Here is the system:

Expansion Postulates

(+1) (**B**+P) is consistent and fully logical.
(+2) P is in (**B**+P)
(+3) Everything in **B** is also in (**B**+P)
(+4) If P is in **B**, then (**B**+P) = **B**
(+5) If everything in **B** is also in **B***, then everything in (**B**+P) is also in (**B***+P)
(+6) (**B**+P) is the smallest set satisfying (+1)—(+5).

Revision Postulates

(*1) (**B***P) is consistent and fully logical
(*2) P is in (**B***P)
(*3) Everything in (**B***P) is also in (**B**+P)
(*4) If ¬P is not in **B**, then everything in (**B**+P) is also in (**B***P)
(*5) (**B***P) is logically inconsistent if but only if P is logically inconsistent
(*6) If P and Q are logically equivalent, then (**B***P) = (**B***Q)
(*7) Everything in [**B***(P&Q)] is also in [(**B***P)+Q]
(*8) If ¬Q is not in (**B***P), then everything in [(**B***P)+Q] is also in [**B***(P&Q)]

Contraction Postulates

(−1) (**B**-P) is consistent and fully logical
(−2) Everything in (**B**-P) is also in **B**
(−3) If P isn't in **B**, then (**B**-P) = **B**
(−4) If P is not a logical truth, then it is not in (**B**-P)
(−5) If P is in **B**, then everything in **B** is also in [(**B**-P)+P]
(−6) If P and Q are logically equivalent, then (**B**-P) = (**B**-Q)
(−7) Everything in both (**B**-P) and (**B**-Q) is also in [**B**-(P&Q)]
(−8) If P isn't in [**B**-(P&Q)], then everything in [**B**-(P&Q)] is also in (**B**-P)

These Postulates are justified by loose intuition and the principle of informational economy. The idea is put forward repeatedly that information is precious, that one should not lose information in changing one's mind unless forced to do so, and that logic is the source of such force. These thoughts are quickly used to codify the Postulates above and then Gärdenfors sets about proving various non-trivial technical things about the Postulates. Often those things are quite interesting. We shall focus on two results.

First: Expansion Postulates fit together with the idea, used in §1.4, that expansion works like set-theoretic union plus deduction:

Expansion-as-Deduction Theorem
A function '+' obeys (+1)−(+6) iff $(\mathbf{B} + \mathrm{P}) = \{\emptyset : \{\mathbf{B} \cup \{\mathrm{P}\}\} \Rightarrow \emptyset\}$.
Recall §1.4's rule for expansion

$$(+) \quad (\mathbf{B} + \mathrm{P}) = \{\emptyset : \{\mathbf{B} \cup \{\mathrm{P}\}\} \Rightarrow \emptyset\}.$$

This is the right-hand side of the Expansion-as-Deduction Theorem. Hence that Theorem shows that Expansion Postulates go hand in hand with (+), pinning down the fact that expansion works like set-theoretic union plus logic. Since the rule (+) is itself motivated by the principle of informational economy, it is nice to see likewise-motivated Postulates fit into place. That is what the Expansion-as-Deduction Theorem shows.

Second, Gärdenfors proves something interesting about revision:

Revision Theorem

If		(1) '−' obeys (−1)−(−4) and (−6)−(−8),
	&	(2) '+' obeys (+1)−(+6),
	&	(3) '*' obeys the Levi identity,
	then	(4) '*' obeys (*1)−(*8).

This motivates the overall Belief Model construction in several directions at once. It makes the Levi identity look right, shows how Postulates fit together as a group, and demonstrates that the model is internally coherent. This is exactly what you would expect when technically astute philosophers such as Carlos Alchourrón, Peter Gärdenfors, and David Makinson bring similar intuitions to bear on a particular formalism.

5.7 Linking States and Transitions: Conditional Belief

All in all, then, we've seen the following:

- The Belief Model boils down its transition theory to three types of shift in view: Expansion, Revision, and Contraction.
- In each case, the model claims that shift in view should be minimal, that it should involve the smallest epistemic change brought on by the context to hand.
- Whenever possible, the model adopts a purely logical and/or set-theoretic take on 'smallest change', which leads to the Levi identity and Postulates for shift-in-view.
- Those Postulates together with that identity jointly imply:
  - Expansion works like set-theoretic union plus deduction,

  and so
  - Revision collapses to contraction plus expansion.
- The Belief Model says contracting **B** by P involves movement to the belief set which consists in what is common to the most entrenched maximal non-P belief subsets of **B**.

On this sort of approach, intuitively, rational configurations of epistemic attitudes should be logically kosher; and rational shifts between such configurations of attitude should themselves be minimal.

Yet nothing has been said so far about whether these prescriptions are linked systematically. So we shall ask: does the Belief Model see any direct link from its theory of states to its transition theory or vice versa?

Yes, the model sees a direct link in both directions: from its theory of states to its transition theory, and from its transition theory to its theory of states. To see how this works, consider the frame

(B) S believes that—.

Various sentences can be used to fill it. Each of the following sentences, for instance:

- You struck the match.
- You caused the fire.
- Oswald didn't kill Kennedy.
- Someone else killed Kennedy.

Plugging any of these sentences into frame (B) yields an ordinary belief attribution:

- S believes that you struck the match.
- S believes that you caused the fire.
- S believes that Oswald didn't kill Kennedy.
- S believes that someone else killed Kennedy.

The Belief Model sees no direct link between such beliefs and rational shift in view.

But it does see a direct link between certain extraordinary beliefs and its transition theory. And these extraordinary beliefs are described when certain *conditional sentences* are put into (B). For instance, when either of the following two sentences is put into (B)

- If you struck the match, you caused the fire.
- If Oswald didn't kill Kennedy, someone else killed Kennedy.

the result is conditional-belief attribution:

S believes that if you struck the match, you caused the fire.
S believes that if Oswald didn't kill Kennedy, someone else did.

The Belief Model sees a very deep link between such conditional beliefs and rational shift-in-view. It uses its take on such conditional beliefs to build a two-way bridge linking belief sets on the one hand and rational state transitions on the other.

In turn this is because the model is heavily influenced by the most influential remark ever made on conditional belief. Almost a century ago the great philosopher F. P. Ramsey wrote:

> If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q...

Throughout formal epistemology this has become known as the 'Ramsey Test' for conditional belief. Gärdenfors describes it in the following way:

> In order to find out whether a conditional sentence is acceptable in a given state of belief, one first adds the antecedent of the conditional hypothetically to the given stock of beliefs. Second, if the antecedent together with the formerly accepted sentences leads to a contradiction, then one makes some adjustments, which are as small as possible without modifying the hypothetical belief in the antecedent, such that consistency is maintained. Finally, one considers whether or not the consequent of the conditional is accepted in this adjusted state of belief. The Ramsey test can be summarized by the following rule: accept a sentence of the form 'If A, then C' in a state of belief **B** if and only if the minimal change of **B** needed to accept A also requires accepting C. (Gärdenfors 1988: 147)

By the Belief Model's lights, of course, minimal change is partial-meet contraction as described by the model's transition theory, as characterized by its Postulates, and so forth.

Let **min-rev** be the process of minimal revision, whatever it turns out to be exactly. The Ramsey Test for conditional belief then becomes:

(RT) (A→C) belongs to an agents set of beliefs **B** iff C belongs to **min-rev**(**B** by A).

This principle builds a bridge between epistemic states and their rational revision, with the bridge being conditional belief. The Ramsey Test ensures that a given epistemic state will mark its own rational revision by the conditional beliefs within it.

5.8 A Matching Psychology: Bella

In Chapter 1 we noted that any full-dress theory of rationality will have three major moving parts: a theory of states, a transition theory, and a story about how the two fit together. The theoretical template is this

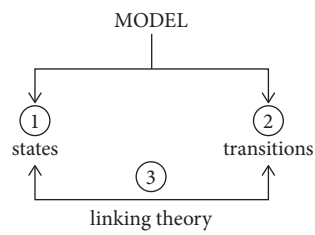


Figure 5.2

In this chapter we've seen how the Belief Model fills out the template

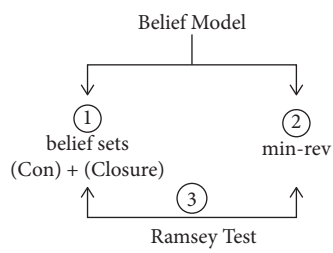


Figure 5.3

In Chapter 1 we also noted that it's best for three things to be true of a formal model of any target phenomena:

1. Basic facts in its target domain are explicitly marked by basic elements in the model;
2. Derivative facts in its target domain are explicitly marked by derivative elements in the model;
3. Derivative facts in the target domain are grounded in target basic facts so that grounding relations are mirrored by those between markers in the model.

When these things happen a model *metaphysically matches* its target domain; and when a model so matches its target domain one can read off from its surface how derivative and basic facts in that domain fit together. Question: what would it take for a target psychology to be metaphysically matched by the Belief Model?

Well, such a psychology would have to consist in basic states of mind marked explicitly by basic elements of the model; it would have to consist in derivative states of mind marked explicitly by derivative markers in the model; and these two aspects of the target domain would have to fit together grounding-wise as their markers do in the model. If we assume the model metaphysically matches a target psychology, therefore, we assume one can read off its surface what the basic facts are about rationality. So ask yourself this: what is the fundamental moving-part of the Belief Model?

Answer: the sentence.[5] The model groups collections of sentences together into belief sets. It then uses these sets to construe rational configurations of belief, and rational transitions between such configurations. The basic mental kind marked by the Belief Model is thus the propositional attitude of belief. Belief sets are the model's surface-level marker for collections of individual states of belief. And the model's revision function **min-rev** is its surface-level marker for rational shift in configurations of belief. The model's Ramsey Test is then its story about how its theory of states fits together with its transition theory. Conditional sentences in belief sets link input and output of **min-rev**. This is how the Belief Model fleshes out parts ❶-thru-❸ of the theoretical template with which we began this section.

For the model to metaphysically match a domain of fact, therefore, its target will need to consist in basic states of belief. Groups of those states will need to shift en masse, upon receipt of new input, as specified by **min-rev** function. And the **min-rev** function will need to work out in line with conditional belief. To see how this might be so, we shall augment an influential thought-experiment by Stephen Schiffer:

> Suppose we are faced with a rather unusual individual: Bella. She has beliefs in whatever sense we do, and those beliefs are psychologically basic to her mind. Bella thinks in a language-like internal system of representation—her language-of-thought—and she has a giant transparent head. Within her giant transparent head there is a transparent box marked 'B'. This is Bella's belief-box. She thinks in an internal language-of-thought; and

[5] Or better still: the sentence-letter, something which functions in in the model as a name for a normal sentence which might codify what one believed. We move back and forth without comment between talk of sentences like this and talk of letters which name them, since keeping track of the distinction between them is not necessary for our discussion.

> she lends the attitude of belief to a claim C by placing a C-meaning sentence of that language in her belief-box. We stipulate that whenever a C-meaning sentence is in Bella's belief-box she manifests the signature function of belief in C. This is just a fancy way of saying that Bella's belief-box is functionally individuated so that its cash value is identical to that of our states of belief.[6]

Let us suppose that Bella's language-of-thought is similar to the language used by the Belief Model—in both cases, say, a sentential language of propositional logic. Let us also suppose that sentences in Bella's belief-box—including conditional sentences—shift as a group in one step, exactly in line with the **min-rev** function. Then the model will metaphysically match Bella's psychology. We have reverse engineered her to ensure that is so.

The fundamental marking element in the Belief Model's theory of states is sentence membership in a belief set. This corresponds to the fundamental mental kind in Bella's psychology: the individual state of belief. The fundamental marking element in the model's transition theory is the **min-rev** function. This corresponds to how content-bearing sentences in Bella's belief-box shift upon perturbation. And the details of Bella's belief-transitions are marked exactly by conditional sentences in her belief box. This corresponds to the Ramsey-style role conditional sentences play in the Belief Model.

Further still—we may stipulate—Bella's psychology involves a pair of derivative propositional attitudes which are subject to epistemic appraisal: disbelief and suspended judgement. But these states only occur in Bella because and in virtue of her belief-box working a certain way. We stipulate that the disbelief-as-belief hypothesis is true of Bella:

(D-as-B) A subject S disbelieves Φ by believing $\neg\Phi$.

And we stipulate that the suspended-judgement-as-belief hypothesis is true of Bella:

(SJ-as-B) A subject S suspends judgement in Φ by failing to believe or disbelieve Φ.

This corresponds to the Belief Model's portrayal of disbelief and suspended judgement, since the model marks the former propositional attitude with a negated sentence in a belief set, and the latter propositional attitude with the absence of a sentence or its negation in a belief set.

By design, then, the Belief Model metaphysically matches Bella's psychology. It marks fundamental aspects of her mind with basic elements of its machinery, derivative aspects of her mind with derivative elements of its machinery, and ensures that fundamental and derivative facts in its target domain fit together exactly like their markers do in the model.

[6] (Schiffer, 1981, p. 212). The hero of Schiffer's thought-experiment is Hilarious Meadow rather than Bella. Schiffer notes—rather implausibly, in my view—that Meadow is no relative of Hilary Putnam or Hartry Field. See also (Fodor, 1975) for the language-of-thought hypothesis.

6

Critical Discussion of the Belief Model

6.1 Preview

The Belief Model is the best-known model of rational coarse-grained attitudes. Its mathematical properties are elegant, extensively investigated, and of permanent use in many areas of theoretical concern (e.g. the theory of data-management). The model has proved to be of intrinsic interest and practical worth. It is not though—or at least not as it stands, anyway—a good formalism of our rational coarse-grained attitudes. This chapter canvasses some of the main reasons why that is so. We'll not look at every serious worry for the Belief Model, since later chapters make clear that some of them cover popular models of fine-grained attitudes too. But we'll focus on some worries intrinsic to the Belief Model itself. These have to do with both its theory of states and its transition theory.

Our discussion will function not only as a critique of the Belief Model but also as a warm-up for tackling general aspects of rationality. Those aspects are to be emphasized repeatedly in later chapters of the book, so it's good to be clear about them here. In the next two sections we focus on the Belief Model's theory of states, and then we turn to its transition theory.

6.2 States in the Belief Model: Only One Coarse-Grained Attitude?

There is a curious asymmetry in the epistemology of coarse- and fine-grained attitudes. This asymmetry holds as much in informal discussions of the phenomena as it does in formal ones. The running motif in the literature on coarse-grained attitudes is that there is deep down only one such attitude: belief. It is not that philosophers reject the very idea of disbelief or suspended judgement—that would fly in the face of common sense, after all—it is rather that philosophers simply assume a reduction of these attitudes to belief. This means they assume, as we put in in §2.8, both the disbelief-as-belief hypothesis and the suspended-judgement-as-belief hypothesis:

(D-as-B) A subject S disbelieves Φ by believing $\neg\Phi$.
(SJ-as-B) A subject S suspends judgement in Φ by failing to believe or disbelieve Φ.

The Rational Mind. Scott Sturgeon, Oxford University Press (2020). © Scott Sturgeon.
DOI: 10.1093/oso/9780198845799.001.0001

Nothing like this happens in the epistemology of fine-grained attitudes. As we'll see in subsequent chapters, it is not generally assumed that one kind of fine-grained attitude, nor even a sub-collection of fine-grained attitudes, is explanatorily basic.

We may picture the relevant perspective on coarse-grained attitudes with a pair of grounding schemata:

Grounding-of-suspended-judgement schema

$$\mathrm{SJ}(\Phi)$$
$$\rightsquigarrow$$
$$\{\neg \mathrm{B}(\Phi) \,\&\, \neg \mathrm{DB}(\Phi)\}$$

Grounding-of-disbelief schema

$$\mathrm{DB}(\Phi)$$
$$\rightsquigarrow$$
$$\mathrm{B}(\neg \Phi)$$

The idea behind the first schema is this: suspended judgement is nothing over and above the absence of belief and disbelief. This is the suspended-judgement-as-belief hypothesis. The idea behind the second schema is analogous: disbelief is nothing over and above belief in negation. This is the disbelief-as-belief hypothesis. If both of these schemata are valid, if all of their instances are true, there is only one basic kind of coarse-grained attitude: belief. The Belief Model proceeds as if that is so.

Consider the analogue view in the epistemology of fine-grained attitudes. Within a probabilistic setting—of the kind we explained from scratch in Chapter 2—this would amount to the idea that for any real-number n (not less than 0 but less than .5):

Grounding of low credence schema

$$\mathrm{Cr}(\Phi) = n$$
$$\rightsquigarrow$$
$$\mathrm{Cr}(\neg \Phi) = (1{-}n).$$

The thought here would be that low credence is deep down nothing but high credence in negation. This is a credal analogue of the idea that disbelief is deep down nothing but belief in negation.

The analogue seems obviously wrong. Low credence doesn't look or feel like high credence in negation, and in the next section we'll see that low credence doesn't fully function like high credence in negation either. Instead low credence looks and feels (and functions) like a proprietary kind of attitudinal stance. Being one-quarter sure that it will snow, for instance, is not really the same thing as being three-quarters sure that it will not snow.[1] The latter take on the weather may itself be forced upon

[1] See Chapter 9 for a general argument that attitude and operator are not fully fungible.

one, rationally, when the former take on the weather is adopted, but the low-credence attitudinal stance seems to be something distinct from its high-credence analogue, psychologically speaking. And so it goes with other states of low credence. They seem to be metaphysically distinct from states of high credence, even states of high credence lent to the negation of their contents. The grounding of low credence schema is implausible on its face.

This fact casts doubt on the grounding of suspended judgement and the grounding of disbelief schemata. It is unsurprising, therefore, that neither of the latter schemata are plausible after reflection. To see this, recall the schema which grounds suspended judgement in belief:

Grounding-of-suspended-judgement schema

$$\mathrm{SJ}(\Phi)$$
$$\rightsquigarrow$$
$$\{\neg \mathrm{B}(\Phi) \,\&\, \neg \mathrm{DB}(\Phi)\}.$$

The lower condition here is clearly ***in***sufficient for its upper condition—or, put another way, the upper condition is clearly something over and above the lower one. When you fail to consider whether Caesar crossed the Rubicon carrying nuclear weapons, for instance—say because you lack the concept of a nuclear weapon—you fail to believe or disbelieve that Caesar crossed the Rubicon carrying nuclear weapons. This does not mean that you've suspended judgement in whether he did so. In these circumstances not only would it be the case that you have no view of the matter—belief, disbelief, or suspended judgement—it would also be the case that you lack the conceptual resources to have a view on the matter.

Suspended judgement is not the absence of belief and disbelief. It is the presence of a proprietary kind of neutral commitment, something more than a mere absence or lack. Suspended judgement is the propositional attitude of *committed neutrality*. It will be our job to flesh out what this comes to.[2]

Now recall the schema which grounds disbelief in belief in negation:

Grounding-of-disbelief schema

$$\mathrm{DB}(\Phi)$$
$$\rightsquigarrow$$
$$\mathrm{B}(\neg\Phi).$$

The idea here is that disbelief is nothing over and above belief in negation. But that too seems plainly false after reflection. The lower condition of the schema concerns the endorsement or acceptance of $\neg\Phi$. Disbelief in Φ does not intuitively concern the endorsement or acceptance of anything: it seems to involve the pushing-away of Φ in

[2] In the fall of 2002 I went through this material in a grad-seminar at Harvard. I confessed frustration at not having a good label for the conception of suspended judgment I was defending. After listening for a while Selim Berker offered the helpful phrase 'committed neutrality'. Thanks Selim!

thought. Disbelief seems to involve a proprietary type of attitudinal condemnation. This too will take some explaining.

But our initial impression in the area is clear enough: belief, disbelief, and suspended judgement are each their own kind of thing. None of them seems to reduce to the others. We need a philosophical conception of coarse-grained attitudes which underwrites this initial impression. Developing one is our next task.

6.3 Coarse-Grained Attitudes

Each of the three coarse-grained attitudes—belief, disbelief, and suspended judgement—seems to have a psychological life of its own. To unearth a conception of how this can be we shall first lay out a pair of thought experiments. Then we'll reflect on the conception of coarse-grained attitudes which falls out of them if one is a functionalist about the attitudes (in line with the Cash-Value Assumption of Chapter 1). The initial thought experiment is

THREE DOORS

> You are told to exit a room by one of three doors: the **L**eft door, the **M**iddle door, or the **R**ight door. Hence there are four things that you might do: exit by the left door **L**, exit by the middle door **M**, exit by the right door **R**, or fail to act. So we can define a condition C by appeal to a double failure:
>
> $$C =_{df.} (\neg \mathbf{L} \,\&\, \neg \mathbf{R}).$$
>
> C is the condition of failing to leave by the left and failing to leave by the right door. There are two ways to get into C: you can do so by exiting via the middle door **M**, you can do so failing to exit at all.

Intuitively, a similar structure plays out with coarse-grained attitudes. Suppose you are given some data relevant to claim Φ. You are instructed to examine the data carefully and adopt the take on Φ which best reflects their relevance concerning whether or not Φ is true. In the event, there are three coarse-grained options before you: belief, disbelief and suspended judgement. Hence there are four psychological possibilities to hand. You might come to believe Φ after studying the data. You might come to disbelieve Φ after doing so. You might come to suspend judgement in Φ after studying the data. Or you might fail to adopt a view about Φ at all (say because you get distracted and start thinking about something else). There are two ways, therefore, to fail to believe Φ while failing to disbelieve Φ: this happens when you suspend judgement in Φ, and also when you have no Φ-commitment whatsoever. While going through the data and trying to decide how they relate to Φ's truth-value—as well as before the question of that truth-value has even occurred to you—you have no Φ -commitment at all. In neither case do you suspend judgement in Φ. In neither of them do you adopt the attitude of committed neutrality to whether Φ is true.

Our second thought experiment is

THE JUDGE

> You are a pre-trial judge meant to assess evidence gathered by police. You must decide if the evidence makes for an initial indication of guilt, innocence, or neither. And there are three verdicts available to you: evidence indicates guilt, evidence indicates innocence, evidence does neither of these things. Suppose Mr. Big has been charged with stealing the cookies and you have been contacted by one of his powerful allies. You are urged to reach a verdict immediately so as to dampen public discussion of the case. But you haven't received the evidential dossier from the police yet. In the event, there are four options to hand. You might commit to the view that the evidence makes for an initial indication of guilt. You might commit to the view that the evidence makes for an initial indication of innocence. You might commit to the view that the evidence does neither of these things. Or you might steadfastly refrain from any such commitment at all. In the envisaged circumstances it is clear that you should opt for the last of these choices, since you've not received the evidential dossier from the police and thus have no clue about the relevant evidence. But you have four options before you rather than three.

The same structure plays out in the epistemology of coarse-grained attitudes. There are four attitudinal stances available. Intuitively:

> settled endorsement = belief
> settled denial = disbelief
> settled neutrality = suspended judgement
> unsettled stance = no epistemic attitude.

There is a clear psychological difference between these four scenarios. The first one involves the settled intellectual embrace of a claim, the kind of endorsement-like state which is incompatible with a fleeting or unstable take on its truth-value. The second scenario involves a settled intellectual pushing-away of a claim, a kind of rejection-like state likewise incompatible with a fleeting or unstable take on its truth-value. The third scenario involves settled intellectual neutrality about a claim, a kind of stable reserve also incompatible with any fleeting take on its truth-value. And the fourth scenario involves no epistemic attitude at all, for it involves too much psychological instability.

None of these states seem to inter-reduce. For example, disbelief does not seem to be any combination of belief, suspended judgement, or lack of commitment. And nor do any of the other coarse-grained states look to boil down to a combination of their coarse-grained cousins. Moreover, a lack of stable attitudinal commitment obviously does not reduce to any combination of stable commitments. So none of the four possibilities above reduce to the others.

We can make ready sense of this from a functionalist perspective. The following is one way to do so (there are many others). First, we align belief lent to a particular content with the strong and stable disposition to rely on that content in various ways. For instance, we align belief in Φ with the strong and stable disposition to take bets

on Φ, to say things like 'Yes, of course!' when asked whether Φ is true, to make use of Φ in our practical and theoretical reasoning, and so on.[3] This sort of functionalism about belief-in-Φ aligns that particular kind of belief with a constellation of 'signature functions'. In turn these functions are always indexed somehow to Φ being true. And each of these signature functions generates a particular disposition, namely, the one defined as a disposition to carry out the Φ-related signature function in question. These signature dispositions are constituent elements of belief-in-Φ. The state of believing Φ itself is identified as the manifestation of enough of these signature dispositions.[4]

Second, we align disbelief in a particular content with the strong and stable disposition to rule-out that content in various ways. For instance, we align disbelief in Φ with the strong and stable disposition to reject bets on Φ, to say things like 'No, don't be stupid!' when asked whether Φ is true, to rule out anything which clearly entails Φ in our practical and theoretical reasoning, and so on. This sort of functionalism about disbelief-in-Φ likewise aligns that particular kind of disbelief with a constellation of signature functions. In turn these functions are indexed to Φ being false; and once more each of them generates a particular signature disposition. These signature dispositions are thought of as constituent elements of disbelief-in-Φ. And the mental state in question is itself identified as the manifestation of enough of the Φ-relevant signature dispositions.[5]

Third, we align suspended judgement with the strong and stable disposition to *refrain* from activities the strong and stable presence of which make for belief or disbelief. This means we align suspended judgement in Φ with the strong and stable disposition to refrain from accepting or rejecting bets on Φ, the strong and stable disposition to say things like 'Oh, I haven't the slightest idea!' when queried about Φ's truth-value, the strong and stable disposition to refrain from using Φ or its negation in any sort of reasoning, and so on. The thought here is to align suspended judgement in Φ with a constellation of signature disposition. Possession of enough of them rules-out functioning as someone who believes or disbelieves in Φ.

And finally, we align attitudinal unsettledness with the absence of dispositions the strong or stable presence of which make for belief, disbelief or suspended judgement. In turn this means being unsettled about Φ is a kind of attitudinal lack about whether Φ is true. More specifically, it's the lack of dispositions the presence of which make for the endorsement of a content, the rejection of a content, or committed neutrality

[3] Chapter 13 sketches a view about how this works for humans despite two further things being true of us: states of believing are states of confidence deep down, and states of confidence do not figure in reasoning.

[4] We take no stand in this book on which pro-Φ functions, exactly, are constitutive of the attitudes? Some kind of functionalism about them will be assumed, but nothing need be specified about whether that functionalism is of the common-sense variety, the scientific variety, or something else yet again. For elegant discussions of functionalism see (Loar, 1981), (Lewis, 1966) and especially (Lewis, 1994), and (Schiffer, 1987).

[5] Something very like the take on disbelief found here often makes an appearance in philosophical logic. Those who make use of it normally do so to reduce or explicate negation. But nothing follows about the semantics of negation, or the nature of logic, from the view that disbelief has a psychological life of its own, that it is metaphysically distinct from belief in negation. For a useful guide to the relevant literature see (Ripley, 2011). See also (Rumfitt, 2000).

about a content. Attitudinal unsettledness involves functional instability. Such instability is the signature of an attitudinal lack, given functionalism about the attitudes.

These are the beginnings of a functional understanding of our coarse-grained attitudes. They underwrite the idea that such attitudes fail to inter-reduce, which in turn makes it reasonable to affirm our initial impression of the coarse-grained attitudes, namely, that belief is a proprietary kind of positive commitment, disbelief is a proprietary kind of negative commitment, and suspended judgement is a proprietary kind of committed neutrality. When it comes to the relation between belief, disbelief and suspended judgement, therefore, the most we can hope for is systematic alignment when the attitudes are fully rational.

For instance, it is plausible to suppose that the following schema is true of fully rational agents who consider whether or not Φ:

(a) $SJ(\Phi)$ iff $\neg B(\Phi) \,\&\, \neg DB(\Phi)$.

It seems plausible to suppose, that is to say, that such agents will suspend judgement exactly when they fail to believe and fail to disbelieve. But this sort of normative alignment between suspended judgement and the absence of belief and disbelief is itself compatible with anti-reductionism about suspended judgement. Just because fully rational suspended judgement (on a question being considered) lines up with the absence of belief and disbelief, after all, it does not follow that suspended judgement itself reduces to the absence of belief and disbelief. And we have seen good reason to think that it does not so reduce. We thus duly conclude that only an (a)-style correlation exists between them.

Similarly, it seems plausible to suppose that the following schema is true of fully rational agents like us:

(b) $DB(\Phi)$ iff $B(\neg\Phi)$.

It seems plausible to suppose, that is to say, that such agents will disbelieve exactly when they believe a negation. But this sort of normative alignment between disbelief and belief in negation is compatible with anti-reductionism about disbelief. Just because fully rational disbelief in humans lines up with belief in negation, after all, it does not follow that disbelief itself is grounded by belief in negation. And we have seen good reason to think that it is not so grounded. We duly conclude that only an (b)-style correlation exists between them.

We are left, then, with a pair of interim conclusions: coarse-grained attitudes stand on their own with respect to one another, and they are subject to proprietary epistemic norms. In this sense there is symmetry between coarse- and fine-grained attitudes: when it comes to how elements within a given attitudinal space relate to one another—i.e. how course-grained attitudes relate to one another, and how fine-grained attitudes relate to one another—anti-grounding seems the best bet. Both coarse- and fine-grained attitudes look to be on an explanatory par with their level-mates, and, for this reason, attitudes within a given level look to be subject to proprietary epistemic norms.

If this is right—and from now on we'll assume that it is—the Belief Model's theory of states fails to match its target domain. After all, the theory's approach to mental

states regards suspended judgement and disbelief as derivative phenomena, at best; but they are *bona fide* propositional attitudes which stand on all fours with belief. And they are subject to proprietary epistemic appraisal in light of the evidence. By marking them in a derivative way, the Belief Model's theory of states fails to match its target domain of fact.

To make it do so we'd need to add disbelief and suspended-judgement sets to the model's machinery; and we'd need to add rules for how the new sets should be structured. Then we could codify an agent's states with an ordered triple <**B, D, SJ**>. **B** would represent the agent's states of belief, **D** her states of disbelief, and **SJ** the contents about which the agent suspends judgement. And just as the consistency rule (Con) and the entailment rule (Entail) are general rules for the overall structure of a belief set, a metaphysically matching Belief-Model-style theory of states would need analogue rules to do with the structure of rational disbelief and suspended judgement.

Think of it this way. Disbelief and suspended judgement are propositional attitudes subject to epistemic appraisal. Belief-Model-style modelling of propositions takes place with sentences. Like their belief-theoretic counterparts, therefore, disbelief and suspended-judgement sets in a matching approach can be represented by sets of sentences. Just as a belief set **B** can be thought of as a set of sentences which express claims believed by a rational agent, disbelief set **D** can be thought of as a set of sentences which express claims disbelieved by a rational agent, and suspended-judgement set **SJ** can be thought of as sets of sentences which express claims in which an agent is neutrally committed. This much is easy when it comes to creating an AGM-style theory which matches our psychology. What is not so easy is specifying the norms which structure rational disbelief and suspended-judgement when we do so. What should they look like?

Well, recall the norms used in the Belief Model's theory of states:

(Con-for-belief)	Belief sets are logically consistent.
(Entail-for-belief)	For any belief set **B** and sentence S: if the members of **B** logically entail S, then S belongs to **B**.

The first rule ensures that as a matter of logic all the members of a belief set can be true together. The second rule ensures that belief sets are closed by logical implication: intuitively, this means that you cannot walk your way out of a belief set by chasing down its logical implications.[6] Each of these rules is meant to apply to a rational agent, and that is a plausible thought on its face. What is equally plausible, however, is that direct analogues of them should *not* apply to rational states of disbelief or suspended judgement.

To see why, consider the analogue rules for disbelief. Neither of them withstands a moment's reflection:

(Con-for-D)	Disbelief sets are logically consistent.
(Entail-for-D)	For any disbelief set **D** and sentence S: if the members of **D** logically entail S, then S belongs to **D**.

[6] See §5.2 for details.

Since disbelief is a kind of stable rejection of a claim, the attitude of disbelief can rationally occur when a claim turns out to be impossible. It might be explicitly inconsistent, for instance, or a truth-functional falsity, or a conceptually incoherent claim (e.g. the claim that a given shoe is red but uncoloured). In any of these cases a rational agent will disbelieve the claim to hand; but this means that the set which represents their states of disbelief will itself be inconsistent. Our approach should definitely permit rational agents to disbelieve such impossible claims. Any acceptable extension of the Belief Model to handle disbelief will thus allow disbelief sets to be inconsistent. The rule (Con-for-D) is a mistake.

Similarly, just because disbelieved contents logically imply a given claim, it does not follow that disbelievers should rule-out that claim. After all, a set of things rationally disbelieved will doubtless contain formal contradictions. Those contradictions will entail basically everything, including simple truth-functional tautologies. Disbelief sets are closed by implication, therefore, only if rational agents disbelieve simple truth-functional tautologies. But that does not look to be possible. We should thus reject the idea that disbelief sets are closed under implication.

What seems true is that rational disbelief is closed under some kind of negative reverse implication. Here is one way to spell out the idea:

(Rev-D-1) For any disbelief set **D** and sentence S: if S entails something in **D**—or entails that something in **D** is true—then S is in **D**.

Here is a more general way to spell out the idea:

(Rev-D-2) For any disbelief set **D** and sentences $S_1 \ldots S_n$: if $S_1 \ldots S_n$ jointly entail something in **D**—or jointly entail that something in **D** is true—then, $(S_1 \& \ldots \& S_n)$ is in **D**.

We do not need to explore how best to spell out the relevant idea; for it is clear enough that entailing something rejected is itself a way of meriting rejection. Any complete model of rational coarse-grained states should respect that insight.

It also seems clear that rational suspended judgement is not consistent. For instance the following rule seems plainly silly:

(Con-for-SJ) Suspended-judgement sets are logically consistent.

After all, suspended judgement is stable neutrality about the truth-value of a content. It would be daft to maintain neutrality toward a content C without also doing so toward its negation. The moment you ruled in or out ¬C, after all, you'd be in position to rule out or in C. Suspended-judgement sets should be *in*consistent sets par excellence, as contradictory as you please.

This means that the following rule is dead on arrival:

(Entail-for-SJ) For any suspended-judgement set **SJ** and sentence S: if members of **SJ** logically entail S, then S belongs to **SJ**.

Since suspended-judgement sets are logically inconsistent, they entail every claim: it is not logically possible for their members all to be true while a given claim is

false, since it is not logically possible for their members all to be true. If suspended-judgement sets were closed by implication, therefore—in line with the rule (Entail-for-SJ)—it would follow that such sets contain every claim whatsoever. But that is absurd. Fully rational agents do not to suspend judgement in everything.

A theory of coarse-grained states metaphysically matches its target domain of fact only if the approach deals explicitly with belief, disbelief and suspended judgement. Fully rational states of belief look to be consistent and closed by logical implication. In this way logic itself plays a crucial role in how fully rational belief is shaped. It is also plausible that two further norms link fully rational belief, disbelief, and suspended judgement. The first links suspended judgement to belief and disbelief about contents being considered:

(a) $SJ(\Phi)$ iff $\neg B(\Phi)$ & $\neg DB(\Phi)$.

And the second links disbelief to belief itself in beings like us:

(b) $DB(\Phi)$ iff $B(\neg\Phi)$.

These are plausible norms linking fully rational belief, disbelief and suspended judgement. If a model is to metaphysically match the links they detail, it must have explicit analogues of them amongst its marking elements.

6.4 The Belief Model's Transition Theory

Recall that the Belief Model's transition theory is built atop three ideas:

1. When shifting view, don't give up something unless forced to by logic;
2. When shifting view, don't take on something unless forced to by logic;
3. Whenever possible, measure the size of what is given up or taken on by inclusion relations between belief sets.

We saw in the previous chapter that these ideas shape the Belief Model's transition theory in several ways. They motivate the Levi identity aligning revision with a two-step process (of contraction and then expansion). They loosely motivate Postulates for modelling revision, expansion and contraction. And their net effect is meant to ensure that the model's transition theory captures something like *purely logical constraints* on the ordinary shift in rational opinion.

As we'll now see, however, there are no such constraints on ordinary rational shift-in-view. One cannot use loose intuition about informational economy, spelled out via logical-cum-set-theoretic machinery, in an effort to track what happens in the ordinary shift of rational opinion. To see why, recall it is by appeal to such machinery that the Belief Model motivates expansion postulates:

(+1) (**B**+P) is consistent and fully logical.
(+2) P is in (**B**+P)
(+3) Everything in **B** is also in (**B**+P)
(+4) If P is in **B**, then (**B**+P) = **B**

(+5) If everything in **B** is also in **B***, then everything in (**B**+P) is also in (**B***+P)
(+6) (**B**+P) is the smallest set satisfying (+1)—(+5).

But these Postulates do not accurately track rational expansion of opinion. They entail, for instance, that adding P to the intersection of two belief sets yields the same thing as taking the intersection of the states got by adding P to each of them. Or in other words: the Belief Model expansion postulates entail that the result of expanding what two belief sets have in common by P is itself identical to expanding each of the belief sets by P and then gathering up what they have in common. Or in still other words:

(☹) $[(\mathbf{B} \cap \mathbf{B}^{\star}) + \mathrm{P}] = [(\mathbf{B} + \mathrm{P}) \cap (\mathbf{B}^{\star} + \mathrm{P})]$.

This is a sad consequence of the Belief Model's transition theory. We may picture it this way

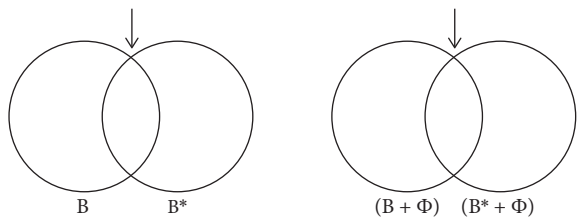


Figure 6.1

Adding Φ to here and then updating $\neq$ What you find here

It is easy to see that rational shift-in-view does not work this way. Just consider a case about Canada and its speakers.

The Canada Case

Let **B** equal the belief set got from the consequences of (i) and (ii):

(i) 99% of Canadians speak English and not French.
(ii) Montreal is in Canada.

Let **B*** equal the belief set got from the consequences of (i), (ii), and (iii):

(iii) 50% of Montreal Canadians speak French and not English.

Now let P be the claim that Pierre is a Montreal Canadian. Obviously, the intersection of **B** and **B*** is **B** itself:

(a) $(\mathbf{B} \cap \mathbf{B}^{\star}) = \mathbf{B}$.

So the expansion of each by P must yield the same state:

(b) $[(\mathbf{B} \cap \mathbf{B}^{\star}) + \mathrm{P}] = (\mathbf{B} + \mathrm{P})$.

But let E be the claim that Pierre speaks English and not French.

Intuitively, E belongs to (**B**+P). If one starts out rational, believes (i) and (ii) and their consequences, and nothing else relevant to the case, then, by coming to believe merely that Pierre is a Montreal Canadian (i.e. P), one should also come to believe that Pierre speaks English and not French. That is why E belongs to the expansion of **B** by P. In symbols, that is why

(c) $\mathrm{E} \in (\mathbf{B} + \mathrm{P})$.

But claim (c) and (b) together ensure

(d) $\mathrm{E} \in [(\mathbf{B} \cap \mathbf{B}^{\star}) + \mathrm{P}]$.

After all, claim (b) insists that the epistemic state got from expanding **B** by P is itself identical to the one got by expanding what's in common to **B** and $\mathbf{B}^{\star}$ by P. It follows from this thought and (c) that E belongs to the overall state got by expanding what's in common to **B** and $\mathbf{B}^{\star}$ by P, i.e. it follows that claim (d) is true.

On the other hand, it is obvious that E will *fail* to be in the expansion of $\mathbf{B}^{\star}$ by P. If one starts out with fully rational belief in (i), (ii) *and* (iii), plus their consequences, and nothing else relevant to the case, then, by coming to believe that Pierre is a Montreal Canadian—i.e. by coming to believe P—one should *not* end-up believing that Pierre speaks English and not French. That is why E will not belong to the expansion of $\mathbf{B}^{\star}$ by P, i.e. why the following claim is true

(e) $\mathrm{E} \notin (\mathbf{B}^{\star} + \mathrm{P})$.

But this claim entails that E cannot be in the intersection of a set **X** and ($\mathbf{B}^{\star}$+P). After all, to be in the intersection of two sets **X** and **Y**, a claim must start out in both **X** and **Y**. Claim (e) insists that E is not in the expansion of $\mathbf{B}^{\star}$ by P; so it cannot be in the intersection of that expansion and something else. For any set at all **X**, therefore, we have the following:

$$\mathrm{E} \notin [\mathbf{X} \cap (\mathbf{B}^{\star} + \mathrm{P})],$$

which guarantees that this claim is true:

(f) $\mathrm{E} \notin [(\mathbf{B} + \mathrm{P}) \cap (\mathbf{B}^{\star} + \mathrm{P})]$.

Thus we find a counter-example to the Sad Consequence mentioned earlier:

(☹) $[(\mathbf{B} \cap \mathbf{B}^{\star}) + \mathrm{P}] = [(\mathbf{B} + \mathrm{P}) \cap (\mathbf{B}^{\star} + \mathrm{P})]$.

This consequence of the expansion postulates is not right. It does not mark an aspect of rational expansion of belief. We know by claim (d) that E belongs to the left side of (☹), and by claim (f) that E does not belong to the right side of (☹). It follows that the left- and right-sides of (☹) are not the same: the former contains the belief that Pierre speaks English and not French, the latter does not. The Belief Model expansion postulates entail mistakenly that these epistemic states are identical. Hence the Canada Case shows that the model's expansion goes wrong, placing mistaken

demands on rational expansion. Or in other words: the model's transition theory mischaracterizes rational shift-in-view from suspended judgement to belief.

Moreover, the Canada Case makes clear *where* Belief Model expansion postulates go wrong. The case points to a major flaw in those postulates. Consider the main moving parts in the case:

(i) 99% of Canadians speak English and not French.
(ii) Montreal is in Canada.
(iii) 50% of Montreal Canadians speak French and not English.
P = The proposition that Pierre is a Montreal Canadian.
E = The proposition that Pierre speaks English and not French.

A fully rational agent who accepts (i), (ii), and P—but nothing else relevant to the issues at hand—thereby has enough information rationally to accept E. A fully rational agent who believes (i), (ii), P and (iii), however—but nothing else relevant to the issues at hand—does *not* thereby have enough information rationally to accept E. Yet the second fully rational agent has all the information possessed by the first fully rational agent, plus further information relevant to the issue at hand. But the extra information had by the second fully rational agent itself undermines information used by the first agent to accept E. That is why the second agent should not join the first in accepting E.

Think of it this way. The following is a perfectly rational line of thought:

'99% of Canadians speak English and not French; Montreal is in Canada; Pierre comes from Montreal; no other information to hand seems relevant. Conclusion: Pierre speaks English and not French.'

But the following is *not* a rational line of thought:

'99% of Canadians speak English and not French; Montreal is in Canada; Pierre comes from Montreal; 50% of Montreal Canadians speak French and not English; nothing else to hand seems relevant. Conclusion: Pierre speaks English and not French.'

What the Canada Case shows is that good reckoning isn't like logical proof.

Here's why: good reckoning can be spoilt by the introduction of new information which is logically consistent with information to hand. Logical proof cannot be spoilt in that way. If a conclusion C is established logically by premises $P_1 \ldots P_n$, then, that conclusion is established logically from those premises together with any further information added to them. A fortiori C is established logically from the premises in question plus information which doesn't conflict with those premises.

This is very important. The key fact here can be unearthed by considering two arguments:

	(A)		(B)
1a	*All* Texans have a Texas accent.	1b	*Most* Texans have a Texas accent.
2	Guido's a Texan	2	Guido's a Texan
C	∴ Guido has a Texas accent.	C	∴ Guido has a Texas accent.

The two arguments share a conclusion, of course, and each of them provides intuitive support for that shared conclusion. But the two arguments support their conclusion in strikingly different ways. The premises of argument (A) link to the conclusion C in what we might term 'an unbreakable way', while the premises of argument (B) do not do anything like that. It is possible to expand (B)-premises with new information—adding nothing which makes for logical conflict—yet break the overall argumentative support for conclusion C. That would happen were you to learn, for instance, that Guido was not raised in Texas but in some place with a differing kind of accent. Although (B)'s premises do support C, they allow for consistent expansion which wipes out that support. (A)'s premises do not allow for this sort of wipe-out. They logically support C no matter how they are expanded.

Think of it this way. Suppose you build a column of claims. First you list a set of claims **S** and then draw a line under those claims. Then under that line you list **S**'s logical consequences L(**S**). Next you build a new column just to the right, only this time you start with the members of **S** plus a new claim N, drawing a line under the embellished collection of premises. Under that new line you then list the logical consequences of the expansion of **S** by N. Since that expansion contains everything in **S**, you are guaranteed that everything below your first line appears below your second, i.e. you are guaranteed that L(**S**+N) contains L(**S**)

$$\begin{array}{ccc} S\ \left\langle \begin{matrix} C_1 \\ C_2 \\ . \\ . \\ . \end{matrix} \right\rangle & \subseteq & \left\langle \begin{matrix} N \\ C_1 \\ C_2 \\ . \\ . \end{matrix} \right\rangle S + N \\ \hline L(S)\ \left\langle \begin{matrix} C_1 \\ C_2 \\ . \\ . \\ . \end{matrix} \right\rangle & \subseteq & \left\langle \begin{matrix} N \\ C_1 \\ C_2 \\ . \\ . \end{matrix} \right\rangle L(S + N) \end{array}$$

Figure 6.2

Rational expansion of opinion is not generally like that. Start a column with a set of premises **S**, draw a line and then list what is rationally supported by **S** (either logically or non-logically). Then start a new column with the members of (**S**+N). There will be no guarantee that what is below your first line appears below your second

$$\begin{array}{ccc} S\ \left\langle \begin{matrix} \Phi_1 \\ \Phi_2 \\ . \\ . \\ . \end{matrix} \right\rangle & \subseteq & \left\langle \begin{matrix} N \\ \Phi_1 \\ \Phi_2 \\ . \\ . \end{matrix} \right\rangle I(S + N) \\ \hline I(S)\ \left\langle \begin{matrix} I_1 \\ I_2 \\ . \\ . \\ . \end{matrix} \right\rangle & ...?... & \left\langle \begin{matrix} \Delta_1 \\ \Delta_2 \\ . \\ . \\ . \end{matrix} \right\rangle \end{array}$$

Figure 6.3

As one says in this area of inquiry, logical proof is *strictly increasing* (aka *monotonic*).[7] Intuitively this means that one never has to retract something logically proved simply because one has added information to one's premises. But we have seen that rational expansion of opinion is not strictly increasing: it is *non*-monotonic.

We can make this precise in the following way. Let us say that

a rule for shift-in-view **R** is *strictly increasing* (or *monotonic*) $=_{df.}$ when a belief set **B** is contained in another **B***, then, for any information I, the result of **R**-revising **B** by I is itself contained in the result of **R**-revising **B*** by I.

In other words, a rule for fully rational shift-in-view is strictly increasing when its picture appears this way

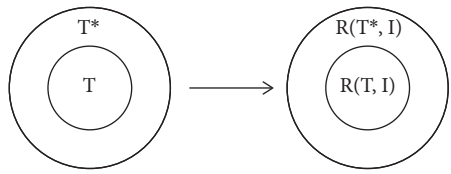


Figure 6.4

Belief Model expansion is strictly increasing. In this respect it works like logical proof. But rational expansion of opinion is not strictly increasing. As the Canada Case makes clear—and countless others like it—rational movement from suspended judgement to belief is a non-strictly-increasing process.

This is our first major lesson about the transition theory for coarse-grained attitudes:

Lesson 1: Rational expansion is non-monotonic.

This conflicts with the fifth expansion postulate of the Belief Model's transition theory:

(+5) If everything in **B** is also in **B***, then everything in (**B**+P) is also in (**B***+P)

But that doesn't seem right at all. Rational expansion does not intuitively work like logical proof. Sometimes when expanding rationally one has to *take things back*; and that is true even when new information is compatible with information already in one's possession.

For this reason, rational expansion is not like logical proof: it is non-monotonic rather than strictly increasing. Rationally accepting something about which you were previously neutral can oblige shrinkage as well as inflation of opinion. After all, new information can override old information. This is sometimes put by saying that

[7] The monotonicity of a function turns on whether it preserves the inclusion relation. An arithmetical function of natural numbers, for instance, is monotonic when the following is true: whenever a set of natural numbers S is a subset of another set of natural numbers S*, the set got by applying **f** to S's members is itself a subset of the set got by applying **f** to S*'s members.

new information can 'defeat' old information, and that's why fully rational expansion is sometimes called a 'defeasible' process. The Belief Model's transition theory mistakenly treats it as an indefeasible process.

How might a defender the model push back against this sort of criticism?

Well, he cannot disallow critique of the Belief Model's transition theory based on thought experiment or intuition. Discussion exactly like that surrounding our Canada Case was used by proponents of the Belief Model to motivate its postulates in the first place. Consider the follow proposal about rational expansion:

$$(*M) \quad \text{If } \mathbf{B_1} \subseteq \mathbf{B_2}, \text{ then } (\mathbf{B_1}{*}P) \subseteq (\mathbf{B_2}{*}P).$$

This rule echoes the Belief Model's postulate (+5), a postulate heartily endorsed by Gärdenfors. But just as (+5) insists that fully rational shift-in-view—from suspended judgement to belief—is itself monotonic, (*M) insists that fully rational shift from rejection to belief, or vice versa, is likewise monotonic. Gärdenfors accepts the first of these ideas but rejects the second. Yet his reason for rejecting (*M) is like ours for rejecting (+5). He constructs thought-experiments to put pressure on the idea that revisions are monotonic.[8] Of course I agree that this can be done: strong intuition about particular cases indicates clearly that fully rational revision is not like logical proof. It is non-monotonic or defeasible. By similar reasoning, however, strong intuition about situations like the Canada Case shows clearly that fully rational expansion is likewise non-monotonic or defeasible. Postulate (+5) is plainly wrong.

Here is a final way to see the bother. Recall the third and fourth postulate for rational revision:

($*_3$) Everything in ($\mathbf{B}{*}P$) is also in ($\mathbf{B}{+}P$)
($*_4$) If $\neg P$ is not in $\mathbf{B}$, then everything in ($\mathbf{B}{+}P$) is also in ($\mathbf{B}{*}P$)

As Gärdenfors notes, Postulates for revision jointly entail

(g) When $\neg P \notin \mathbf{B}$, ($\mathbf{B}{*}P$) = ($\mathbf{B}{+}P$).

By the time Gärdenfors gets around to proving this in his book, however, he's already proved that Belief Model expansion is monotonic! His overall position is thus committed to the view that all counter-examples to (*M) occur when the epistemic shift of opinion involves accepting something previously rejected. Why on Earth should *that* be? So far as I can see there is no rhyme nor reason to this constellation of commitments. We need a better theory of rational shift in coarse-grained attitudes.

Recall the guiding ideas behind the Belief Model transition theory:

- When shifting view, don't give up something unless forced to by logic;
- When shifting view, don't take on something unless forced to by logic;
- Whenever possible, measure the 'size' of what is given up or taken on by inclusion relations between belief sets.

[8] (Gärdenfors 1988: 59ff).

None of these ideas is correct. Often a fully rational agent should give up a belief consistent with all else she believes, often she should accept things not entailed by her new information, and minimal epistemic disturbance should definitely *not* be seen through the lens of logic and/or set theory.

At bottom the Belief Model transition theory is meant to capture something like 'purely logical' constraints on rational shift-in-view. But there is basically no reason to suppose that there *are* purely logical constraints on rational shift-in-view. The underlying assumptions of the theory seem deeply flawed.