Belief Revision and Epistemology

John L. Pollock
Anthony S. Gillies

Department of Philosophy
University of Arizona
Tucson, Arizona 85721

Abstract

Postulational approaches attempt to understand the dynamics of belief revision by appealing to no more than the set of beliefs held by an agent and the logical relations between them. It is argued there that such an approach cannot work. A proper account of belief revision must also appeal to the arguments supporting beliefs, and recognize that those arguments can be defeasible. If we begin with a mature epistemological theory that accommodates this, it can be seen that the belief revision operators on which the postulational theories are based are ill-defined. It is further argued that there is no way to repair the definitions so as to retain the spirit of those theory. Belief revision is better studied from within an independently motivated epistemological theory.

1. Belief Revision

It is an obvious fact about human epistemology, and presumably the epistemology of any sophisticated rational agent, that beliefs are not static. As time passes, new information is acquired, and beliefs change accordingly. *Belief revision* is the process of changing one's beliefs to reflect the acquisition of new information. A natural concern of epistemology is to give an account of *rational* belief revision. There are two ways one might proceed in investigating rational belief revision. The "postulational approach" proposes abstract general principles purportedly governing the process—in effect, axiomatizing or partially axiomatizing belief revision. By contrast, the "derivational approach" tries to *derive* a theory of belief revison from a more concrete epistemological theory.

A number of authors have favored the postulational approach.[1] The best known such theory is the AGM theory originated by Alchourrón, Gärdenfors, and Makinson (1985) and developed further by Gärdenfors (1988, 1992) and Alchourrón and Makinson (1982, 1985). The

AGM theory has been criticized in recent work by researchers in AI, notably by Boutilier, et al. (1998), and Darwiche and Pearl (1997). These criticisms, however, culminate in proposing extensions of the AGM theory (e.g., by proposing additional postulates)—and so can easily be seen to fit within the postulational approach. The purpose of this paper is to question the viability of that theory, and by implication to question the fruitfulness of the entire postulational approach. Our contention will be twofold. First, such theories of belief revision proceed at too high a level of abstraction, ignoring aspects of rational cognition without which it is impossible to formulate true principles of rational belief revision. A theory of belief revision should be driven by a more concrete epistemology. In the absence of an epistemological theory to generate a theory of belief revision, the latter will have be driven by nothing more than vague formal intuitions, and such intuitions are almost certain to get the details of belief revision wrong. Second, abstract theories of belief revision are not necessary anyway if we have a sufficiently detailed epistemological theory, because we can just apply the theory to see how to revise our beliefs. Of course, to serve this purpose the epistemological theory must be developed in sufficient detail to tell us precisely how belief revision should proceed in any given instance. Many epistemological theories are themselves too abstract to serve that purpose. The epistemological theory employed in this paper is that underlying the OSCAR architecture for rational agents (Pollock 1995). This theory is fully implemented in the AI system OSCAR and in any given case it makes a determinant decision about how an agent's beliefs should be revised.

2. The AGM Theory

The AGM theory begins by modeling a belief-state as a set of propositions (or sentences) closed under logical entailment. Gärdenfors (1988) acknowledges that the beliefs of real epistemic agents are not closed under logical entailment (symbolized as ‘ ’), but proposes (pg 9) that belief states so construed be understood as "an *idealization* of the psychological concept—an idealization that is judged in relation to the rationality criteria for the epistemological theory". He continues, "a rational state of belief is one that is in equilibrium under all forces of internal criticism". Belief revision occurs in response to "epistemic inputs", which "can be thought of as the deliverances of experience or as linguistic (or other symbolic) information provided by other individuals (or machines)" (pg 7).

Much of the structure of the AGM theory results from distinguishing between three different cases of belief revision. Given a belief set K and a new piece of information encoded by a proposition P, let K^*P be the result of revising K by adding P and making whatever other changes that requires. Roughly, $K*P$ is K minimally changed to accommodate P. One central tenet of the AGM theory is that we must distinguish between the case in which P is logically consistent with K and the case in which it is not. It is claimed that when P is consistent with K, then we can simply add it to K and close the result under logical entailment. We define:

$$K+P = \{Q \mid K\cup\{P\} \quad Q\}.$$

'+' is the *expansion* operator. The claim is then:

$$\text{If } {\sim}P \notin K \text{ then } K*P = K+P.$$

A different kind of belief change consists of withdrawing belief in P. This is symbolized as '$K–P$'. '–' is the *contraction* operator. Because belief sets are closed under logical entailment, we cannot just remove P from K. P may be entailed by other members of K. E.g., suppose $Q,R \in K$ and $(Q\&R) \quad P$. Then in addition to removing P, we must remove either Q or R (or both). The main problem for a theory of belief revision in this case is to specify which beliefs should be removed along with P. The second central tenet of the AGM theory is a principle of "informational economy" according to which no belief is to be given up unnecessarily—the set of beliefs removed from K along with P must be minimal. This is captured by defining

$$K\perp P = \{K' \mid K' \subseteq K \,\&\, P \notin K' \,\&\, (\forall Q)[(Q \in K \,\&\, Q \notin K') \supset K' \cup \{Q\} \quad P]\}.$$

That is, the members of $K\perp P$ are maximal subsets K' of K not containing P and such that no member of K not in K' can be added without entailing P. $K–P$ might be a member of $K\perp P$. Alternatively, several members of $K\perp P$ might be "tied" as the best way to revise K by removing P, in which case $K–P$ is the intersection of those several members of $K\perp P$.

Contraction requires rejecting beliefs other than P, but there is a common conviction that in choosing which beliefs to reject, or eqivalently, choosing which members of $K\perp P$ to prefer, it is not logically determinate just how the change should be made. Gärdenfors has explored the possibility of ordering the members of $K\perp P$ in terms of a relation of *epistemic entrenchment.* Spohn (1987) has suggested an alternative ordering in terms of *epistemic plausibility*. These orderings may be logically determinate, but they could also be taken to reflect personal preferences.

The meat of the AGM theory results from the claim that $K*P$ and $K–P$ are interdefinable. Alchourrón, Gärdenfors, and Makinson endorse the *Harper Identity* according to which:

$$K - P = (K * {\sim}P) \cap K$$

and the *Levi Identity* according to which:

$$K * P = (K - {\sim}P) + P$$

The latter is core of the AGM theory, telling us that belief revision can be regarded as the composition of a contraction and an expansion. Alchourrón, Gärdenfors, and Makinson defend these two interdefinitions by proposing the following set of postulates (where a belief-set is any set of propositions closed under logical entailment):

(K*1) If K is a belief-set, so is $K*P$.
(K*2) $P \in K*P$.
(K*3) $K*P \subseteq K+P$.
(K*4) If ${\sim}P \notin K$ then $K+P \subseteq K*P$.
(K*5) $K*P = K_{\perp}$ iff ${\sim}P$ (where $K_{\perp}$ is the absurd belief-set—the set of all propositions).
(K*6) If $(P \equiv Q)$ then $K*P = K*Q$.
(K*7) $K*(P\&Q) \subseteq (K*P)+Q$.
(K*8) If ${\sim}Q \notin (K*P)$ then $(K*P)+Q \subseteq K*(P\&Q)$.

(K–1) If K is a belief-set, so is $K-P$.
(K–2) $K-P \subseteq K$.
(K–3) If $P \notin K$ then $K-P = K$.
(K–4) If not P then $P \notin (K-P)$.
(K–5) If $P \in K$ then $K \subseteq (K-P)+P$.
(K–6) If $(P \equiv Q)$ then $K-P = K-Q$.
(K–7) $(K-P) \cap (K-Q) \subseteq K-(P\&Q)$.
(K–8) If $P \notin K-(P\&Q)$ then $K-(P\&Q) \subseteq K-P$.

They then prove the following theorems:

Theorem 1: If '–' satisfies (K–1) – (K–4) and (K–6) and '*' is defined by the Levi Identity, then it satisfies (K*1) – (K*6).

Theorem 2: If the assumptions of theorem 1 are satisfied then (a) if (K–7) is satisfied, (K*7) is satisfied, and (b) if ($\underline{K}$–8) is satisfied, ($\underline{K}$*8) is satisfied.

Theorem 3: If '*' satisfies ($\underline{K}$*1) – ($\underline{K}$*6), and '–' is defined by the Harper Identity, then it satisfies ($\underline{K}$–1) – ($\underline{K}$–6).

Theorem 4: If the assumptions of theorem 3 are satisfied then (a) if ($\underline{K}$*7) is satisfied, ($\underline{K}$–7) is satisfied, and (b) if ($\underline{K}$*8) is satisfied, ($\underline{K}$–8) is satisfied.

Alchourrón, Gärdenfors, and Makinson conclude that these theorems strongly support the interdefinitions.

Many authors have found the AGM theory appealing, and it has formed the basis for an entire industry of articles on belief revision. However, it is our contention that the theory is based upon premature formalization without giving sufficient care to the meanings of the fundamental concepts of the theory. We find it quite difficult to make careful epistemological sense out of the theory, and our best effort at doing that has the consequence that virtually all the postulates ($\underline{K}$–1) – ($\underline{K}$–8), ($\underline{K}$*1) – ($\underline{K}$*8), and the Levi and Harper Identities, are false. To establish this we must lay some epistemological groundwork. That will be the subject of the next section.

3. Epistemology

It is our conviction that the postulational approach to belief revision places the belief revision cart squarely before the epistemological horse. It is a mistake to try to construct a general theory of belief revision in an epistemological vacuum. If we want to know how belief revision works, we should start with a mature epistemological theory and see what it implies about belief revision. The result might or might not accord with the AGM postulates, and if not then so much the worse for the postulates.

3.1 Nondoxastic Foundationalism

We will assume the epistemological theory developed in Pollock (1987, 1995). This theory is implemented in OSCAR. Pollock called this *direct realism*, but a more descriptive title might be *nondoxastic foundationalism*.[2] The basic assumption is that the source of our justified beliefs is perception, construed broadly. Perception provides us with *percepts*, which are representations of the world. Having a percept gives us a defeasible justification for believing the world is as the percept represents it. This is captured by saying that the percept provides a *defeasible reason* for the belief it supports. Note that this makes the *reason-for* relation a relation between mental states, not the contents of the mental states. This is important because we reason quite differently from different kinds of mental states with the same content, e.g., the percept of there being something red before me, the desire that there be something red before me, and the belief that there is something red before me.

Reasoning begins with percepts and moves to beliefs. Most reasoning is from beliefs to beliefs. When the belief that *P* provides a reason for believing *Q*, we can conveniently express this by saying that *P* is a reason for *Q*, but it must be borne in mind that it is really the beliefs, and not their contents, that are reasons for each other.

Gärdenfors talks about belief revision being driven by epistemic input. The only kind of epistemic input our epistemology recognizes is perception. Of course, much of our knowledge is the result of third-person reports, but that too comes via perception. Perception gives us a reason for thinking that it is reported (by a person S) that *P*. We have a reason (presumably inductive) for believing that reports (of certain sorts) are generally true, and that gives us a defeasible reason for believing that *P*. So information derived from third-person reports is in fact derived from perceptual input via a several-step argument.

3.2 Defeasible Reasoning

Beliefs are held on the basis of arguments, the ultimate premises of which are percepts. Arguments are constructed by stringing together reason-schemas. A set Γ of states is a reason for a belief B in a proposition *P* iff the pair $\langle\Gamma, B\rangle$ instantiates a reason-schema. If Γ is a set of belief states, we will also say that the set of propositions believed constitutes a reason for *P*.

There is a fundamental distinction between two kinds of reason-schemas—defeasible and non-defeasible. Reasoning in accordance with a defeasible reason-schema provides a presumption in favor of the conclusion, and justifies belief in the conclusion in the absence of conflicting information. However, associated with a defeasible reason-schema is a set of *defeaters*, belief in which can defeat the justification of the conclusion. Defeasible reason-schemas are those for which there are (potential) defeaters. But some reason-schemas do not have associated defeaters. For example, for any propositions *P* and *Q*, (*P*&*Q*) is a reason for *P*, and this inference is non-defeasible. Deductive inference is inference performed exclusively in terms of non-defeasible reason-schemas.

There are two kinds of defeaters for defeasible reason-schemas. *Rebutting defeaters* are reasons for denying the conclusion. For example, something's looking red to me constitutes a defeasible reason for believing that it is red. A rebutting defeater for this inference is any reason for thinking that the object is not red. It is of fundamental importance to the structure of defeasible reasoning that rebutting defeaters do not exhaust the class of defeaters. *Undercutting defeaters* attack the connection between the premises and the conclusion of a defeasible inference rather than attacking the conclusion itself. For example, if something looks red to me but I also know that it is illuminated by red lights and red lights can make things look red when they are not, this defeats the inference to the conclusion that the object is red. But notice that "x is illuminated by red lights and red lights can make things look red when they are not" is not a reason for thinking that x is not red, since red things look red in red light too. Instead, it attacks the connection between the object's looking red and its being red. Where *P* is a defeasible reason for *Q*, we can think of undercutting defeaters as reasons for denying that *P* wouldn't be true unless *Q* were true. More compactly, an undercutting defeater is a reason for thinking that P's being true does not guarantee that *Q* is true.

The logical machinery of defeasible reasons and defeaters makes it possible to reconstruct a lot of epistemologically complex reasoning. Pollock (1990) applies this framework to a varieity of probabilistic and inductive reasoning, including the statistical syllogism, direct inference, and statistical and enumrative induction. Pollock (1997 and 1998) applies it to reasoning about change and persistence, and causal reasoning. Pollock (1999 and 1999a) uses it to provide logical foundations for means-end reasoning. A salient feature of all of these reconstructions is that it is the undercutting defeaters associated with a defeasible reason-schema that provide most of the logical structure that makes complex defeasible reasoning work. For example, the statistical syllogism tells us, roughly, that "*x* is an *F*, and the probability of an *F* being a *G* is *r*" is a defeasible reason for believing "*x* is a *G*", provided *r* is high enough.[3] To illustrate, suppose we know that Boris' fingers are covered with spots, and it is extremely probable that if a person's fingers are covered with spots then they are suffering from myometatarsilitis. Then we have a reason for believing that Boris is suffering from myometatarsilitis. But suppose we also know that Boris does not have a fever, and the probability is low that a person has myometatarsilitis if his fingers are covered with spots but he does not have a fever. Then a second application of statistical syllogism gives us a reason for thinking that Boris is not suffering from myometatarsilitis. Thus we have two defeasible arguments with conflicting conclusions, as diagrammed in figure 1. The "fuzzy" arrows represent defeat relations. As diagrammed, this would be a case of collective defeat—each inference provides a rebutting defeater for the other, and no way of deciding between them has been provided. But intuitively, we want to give precedence to the second inference because it is based upon more information. This is accomplished by adopting the following undercutting defeater:

"$\underline{C}(\underline{c})$ & prob($\underline{A}/\underline{B}$&$\underline{C}$) $\neq$ prob($\underline{A}/\underline{B}$)" is an undercutting defeater for the inference from "$\underline{B}(\underline{c})$

& prob($\underline{A}$/$\underline{B}$) = $\underline{r}$" to "$\underline{A}$($\underline{c}$)".

With the addition of this undercutting defeater, we can reason as in figure 2, and draw the undefeated conclusion that Boris does not have myometatarsilitis.

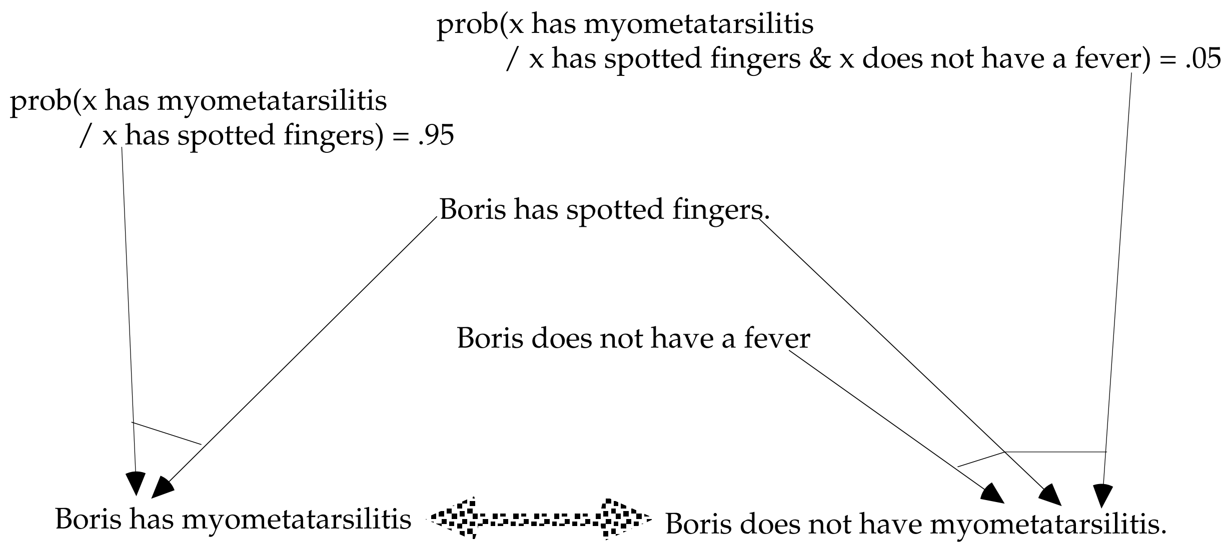

Figure 1. Collective defeat

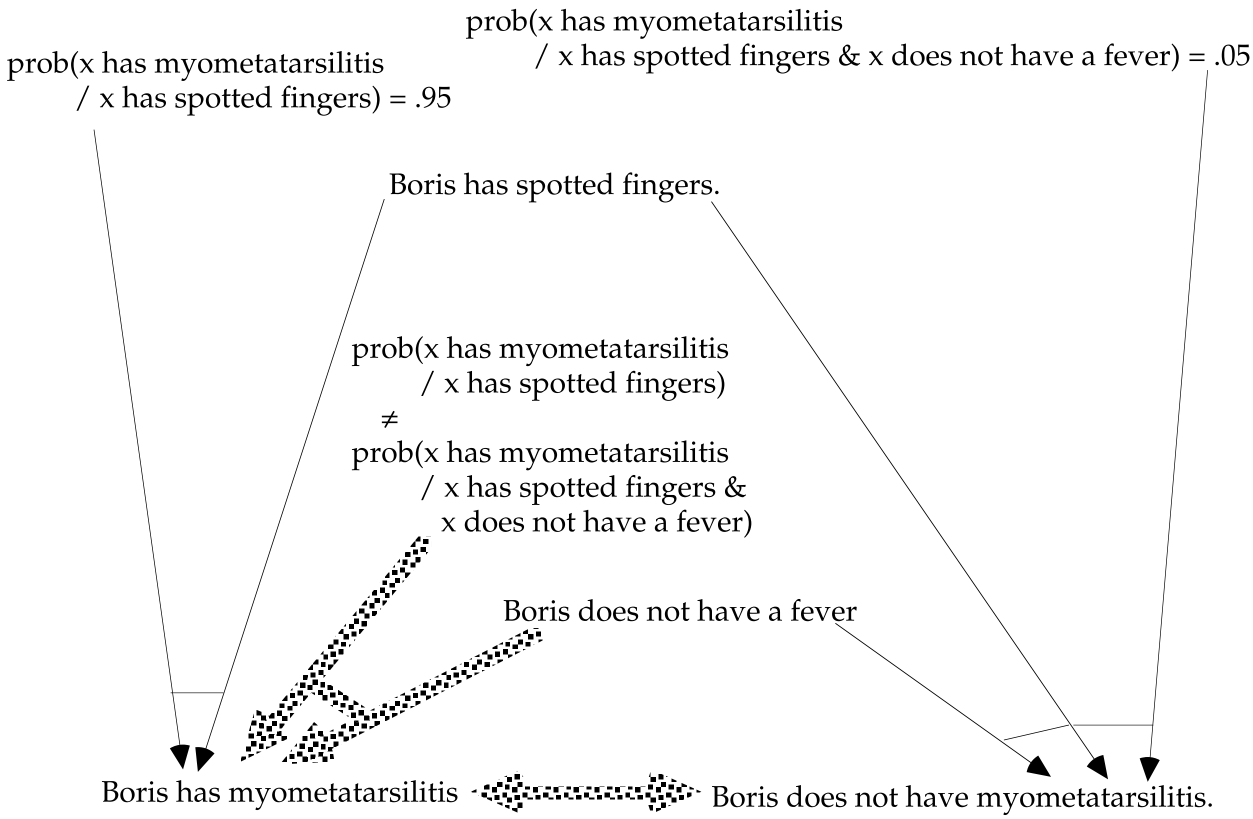

Figure 2. Resolution of the collective defeat

The preceding example illustrates an important point. Inconsistent conclusions generate rebutting defeat, but rebutting defeat is symmetrical. Unless there is something to break the symmetry, there is no way to choose between the inconsistent conclusions, and so the rational thing to do is remain agnostic and take both conclusions to be defeated. This is a case of collective defeat. For example, if I am in a windowless room, and Jones, whom I trust, comes into the room and tells me that it is raining, it is defeasibly reasonable for me to believe that it is raining. But if Smith, whom I trust equally, tells me that it is not raining, I should rationally withdraw the belief that it is raining, but I should not go further and conclude that it is not raining. I should remain agnostic in the absence of further evidence. There are, however, two ways the symmetry of the rebutting defeat can be broken. The simplest is that I may have better reasons for believing one of the inconsistent conclusions rather than the other. For example, if I trust Smith more than Jones, then I can continue to believe that it is raining. The other way of breaking the symmetry is to have independent defeat for one of the inconsistent conclusions. This is what undercutting defeaters provide.

3.3 Degrees of Justification

As indicated above, sometimes a dispute between *P* and ~*P* can be adjudicated by appealing to the fact that we have stronger support for one of the inconsistent conclusions rather than the other. How is degree of support measured? Arguments are constructed by stringing together inferences licensed by reason-schemas. Let us take an *argument-step* to be a triple ⟨premises, conclusion, reason-schema⟩. We can then provisionally define

> An *argument* *A* is a set of argument-steps such that for each $\langle \Gamma, P, R \rangle$ in *A*, for each *Q* in Γ, either *Q* is a percept or for some Λ and *S*, *A* contains an argument-step $\langle \Lambda, Q, S \rangle$.[4]

(We will adopt a more elegant representation of arguments in Section 4.) Reason-schemas have *reason-strengths* associated with them, which are real-numbers between 0 and 1. Non-defeasible reason-schemas always have reason-strength 1. Percepts also have strengths associated with them. The strength of a percept is the degree of justification it will confer on the conclusion that the world is as perceived (in the absence of defeaters). Thus, for example, a percept that something looks clearly red will have greater strength than a percept that something looks vaguely red (e.g., as seen in dim light). It is argued in Pollock (1995) that the degree of support that an argument provides for its conclusion is determined by its weakest link:

> The *strength* of an argument is the minimum of the strengths of the percepts and reason-schemas employed in it.[5]

If an argument is undefeated, then the degree of justification it confers on its conclusion is equal to the strength of the argument. If a conclusion is supported by several arguments, its degree of justification is the maximum of the strengths of the undefeated arguments supporting it.

Collective defeat can often be adjudicated by appealing to argument-strengths. Given otherwise undefeated arguments for *P* and for ~*P*, belief in *P* is justified if the argument for *P* is stronger than the argument for ~*P*. More generally, a defeater defeats an inference iff its degree of justification is greater than or equal to the minimum of the reason-strength and the degrees of justification of the premises of the inference.

It remains to say when an argument is undefeated. Defeat status is determined by argument strengths and their comparison to the degrees of justification of defeaters. We have principles describing how argument strengths are computed, and how the degrees of justification of defeaters must compare to the defeatees, so it is natural to suppose that we can just apply these principles recursively to compute defeat statuses and degrees of justification in general.

The proposal would be that we begin with percepts, which are assigned degrees of justification directly, and then apply the weakest link principle recursively to compute degrees of justification for the conclusions of arguments and the defeaters for arguments. Unfortunately, matters are not so simple. The proposed recursion assumes that the set of arguments can be ordered in such a way that each argument comes *after* all arguments for undercutting defeaters for its steps. Without this *linearity assumption*, there is no reason to expect the recursive computation to assign degrees of justification to every conclusion. And, unfortunately, the linearity assumption need not hold. Arguments can form a tangled web in which no such linearization is possible. To consider a very simple example, suppose Brown tells me that Smith is an unreliable informant, and Smith in turn tells me that Brown is an unreliable informant. Because we know that people are generally reliable informants, a person's telling us something gives us a reason (in accordance with the statistical syllogism) for believing what we are told. Thus we have the two interrelated arguments of figure 3. The problem is that each supports an undercutting defeater for the other, so there is no way to order them so that each one comes after all arguments supporting undercutting defeaters for it, and accordingly there is no way to compute degrees of justification recursively.

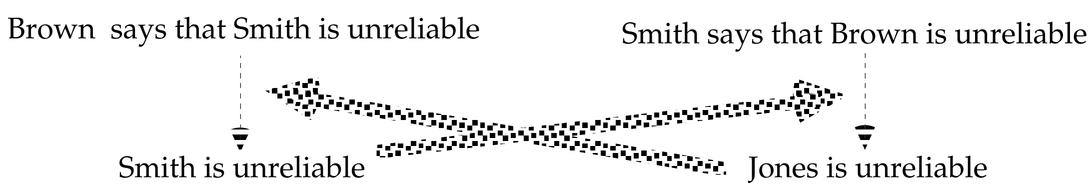

Figure 3. Mutual undercutting defeat

A solution to the general problem of computing degrees of justification must incorporate a solution to the problem of determining which arguments are defeated and which are undefeated. These are not really separable problems, because in accordance with the principles defended above, degrees of justification are a function in part of defeat statuses, but defeat statuses are also a function in part of degrees of justification. These two problems can be unified by identifying the degree of justification of a conclusion with the defeat status of the strongest undefeated argument supporting it. In other words, defeat statuses are not just "defeated" and "undefeated". Defeat statuses are numbers representing degrees of justification, and a defeated argument is one that is completely devoid of justification, i.e., one for which the defeat status is 0.

The problem of computing degrees of justification then becomes the problem of computing defeat statuses. A theory of defeasible reasoning must provide an analysis of defeat status when we are presented with a set of arguments some of which support defeaters for others. The proposal made in Pollock (1992 and 1995) was to define status-assignments to be assignments of defeat status to the conclusions of arguments, where the assignments are required to be consistent with the principles formulated above regarding degrees of justification. Then a conclusion is ruled undefeated iff it is undefeated in every status assignment.

More precisely, we collect arguments into an *inference graph*, where the nodes represent the conclusions of arguments, *support-links* tie nodes to the nodes from which they are inferred, and *defeat-links* indicate defeat relations between nodes. The diagrams of arguments given in Figures 1-3 can be regarded as diagrams of inference-graphs. The analysis is somewhat simpler if we construct the inference graph in such a way that when the same conclusion is supported by two or more arguments, it is represented by a separate node for each argument. So in an important sense, the nodes represent arguments rather than just representing their conclusions. The nodes of the inference-graph will be called *conclusions*. A conclusion α has a type ("percept", "inference", "desire", etc.), a propositional content prop(α) (the proposition *supported by* α), the set def(α) of nodes that are defeaters for α, the set basis(α) of nodes from which α is inferred, and the reason-scheme employed in the inference.

To define status-assignments, we first define the notion of a partial-status assignment, which assigns status-assignments to some subset of an inference-graph $\mathcal{G}$ in accordance with the following rules:

σ is a *partial-status-assignment* for an inference-graph $\mathcal{G}$ iff for each $\alpha \in \mathcal{G}$:

1. if α encodes a percept, $\sigma(\alpha)$ is the strength of the percept;
2. otherwise, if basis(α) = $\emptyset$, $\sigma(\alpha) = 0$;
3. otherwise, if for some $\delta \in$ def(α), either $\sigma(\delta) \geq$ reason-strength(α) or there is a $\beta \in$ basis(α) such that $\sigma(\delta) \geq \sigma(\beta)$, $\sigma(\alpha) = 0$;
4. otherwise, if σ assigns values to all members of def(α) and basis(α), $\sigma(\alpha)$ = the minimum of reason-strength(α) and the values of $\sigma(\beta)$ for $\beta \in$ basis(α).

σ is a *status-assignment* for an inference-graph $\mathcal{G}$ iff σ is a maximal partial-status-assignment for $\mathcal{G}$ (i.e., there is no partial-status-assignment σ^* for $\mathcal{G}$ such that $\sigma \subset \sigma^*$).

A conclusion α is *undefeated relative to* an inference-graph $\mathcal{G}$ iff for every status-assignment σ for $\mathcal{G}$, σ assigns a non-zero value to α.

Different status-assignments may assign different values to a conclusion, but it turns out that if the conclusion is undefeated, all status-assignments assign the same value. So we can define:

> If a conclusion α is defeated relative to an inference-graph $\mathcal{G}$, its degree of justification relative to $\mathcal{G}$ is 0. If it is undefeated, its degree of justification relative to $\mathcal{G}$ is the unique value assigned to it by every status-assignment.

Recall that conclusions represent arguments. If a proposition <u>*P*</u> is supported by more than one argument, then it will be supported by more than one conclusion in the inference-graph. Accordingly, we can define:

> A proposition <u>*P*</u> is justified to degree δ relative to an inference-graph $\mathcal{G}$ iff δ is the maximal γ such that there is an α of type "inference" for which prop(α) = <u>*P*</u> and α is justified to degree γ relative to $\mathcal{G}$.

3.4 Epistemological States

Philosophers tend to suppose that one's "epistemological state" is constituted by beliefs. But we can now see that to be an inadequate representation of the elements of the agent's psychological state that are relevant to epistemological concerns. Specifically, the focus on beliefs is too narrow. If an agent is rational, it believes all and only the propositions that are justified relative to the inference-graph recording all of its reasoning. Thus the beliefs correspond to undefeated conclusions. But we cannot describe all relevant aspects of an agent's epistemology just by focusing on the undefeated conclusions. Defeated conclusions also play a role in belief dynamics. To illustrate, consider the inference-graph diagrammed in figure four. If we regard Smith and Jones as equally reliable, and each accuses the other of lying, then we have a case of collective defeat. We should remain agnostic about whether either is a liar. But under these circumstances, it would also be unreasonable to believe that it is raining just because Smith says that it is. This illustrates that the conclusion that Smith is a liar can defeat the inference from Smith's saying that it is raining to its raining, even though the conclusion that Smith is a liar is defeated. What is crucial to this example is that the defeat is a case of collective defeat. There are two status-assignments for this inference-graph. One assigns "defeated" to "Smith is a liar" and "undefeated" to "Jones is a liar", and the other does the reverse. The assignment assigning "undefeated" to "Smith is a liar" assigns "defeated" to "It is raining". Thus there is an assignment in which "It is raining" is assigned "defeated", and so that conclusion is not justified. This shows that a conclusion can be defeated because there is a status-assignment assigning "defeated" to it, but it can retain the ability to defeat other inferences if there is also a status-assignment assigning "undefeated" to it. Such conclusions are said to be *provisionally defeated*. A conclusion is *defeated outright* if every status assignment assigns "defeated" to it. Conclusions that are defeated outright do not have the power to defeat other inferences.

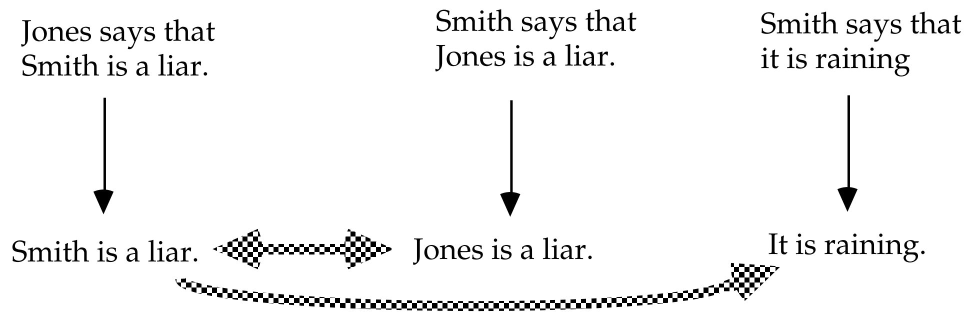

Figure 4. Provisional defeat

Provisionally defeated conclusions still play an important role in belief dynamics. If Smith says that it is raining *after* the initial interchange between Smith and Jones, it is the provisionally defeated conclusion that Smith is a liar that makes it unreasonable for us to revise our beliefs by concluding that it is raining. Thus more than beliefs (undefeated conclusions) must be included in a representation of an agent's epistemological state.

Only the conclusions of undefeated inferences should be believed, but the conclusions of defeated inferences can still be epistemologically relevant, and so must be encoded in a representation of one's epistemological state. In other words, the state must include *conclusions*, not all of which are beliefs. In fact, it looks like the inference-graph itself is an ideal representation of the agent's epistemological state, because it records all the conclusions drawn, the bases upon which they were drawn, and the defeat relations between them.

We do not claim that the inference-graph is a psychologically realistic representation of a datastructure actually present in the agent's cognitive structure. At the very least, it seems psychologically implausible that when we have multiple arguments supporting a single proposition, these are represented by distinct conclusions in the inference graph. It is, however, possible to reformulate the theory so that a conclusion may have multiple bases, and that has been done in Pollock (1995), providing what are called "and/or inference-graphs". Whether that is a representation of a psychologically real datastructure is an open question.

3.5 Epistemological Warrant

Inference-graphs record the current state of an agent's reasoning, and epistemic justification measures the epistemic status of an agent's beliefs relative to that current state. There is always more reasoning that can be done, even without new perceptual input, and sometimes it is useful to talk about what would be justified if the agent did further reasoning. Let $\mathcal{G}+$ be the result of extending $\mathcal{G}$ by adding nodes for all propositions $\underline{P}$ that can be inferred in a single step from nodes in $\mathcal{G}$. Then define:

$$\mathcal{G}_0 = \mathcal{G};$$
$$\mathcal{G}_{\underline{i}+1} = (\mathcal{G}_{\underline{i}})^{+};$$
$$\mathcal{G}_{\omega} = \quad \{\mathcal{G}_{\mathrm{i}} \mid \underline{i} \in \omega\}$$

$\mathcal{G}_{\omega}$ represents the result of doing all possible reasoning from the agent's current set of perceptual inputs. An agent can never actually be in that state, but this represents the ideal limit of the agent's reasoning. We can assess the epistemic status of a proposition relative to $\mathcal{G}_{\omega}$ in the same way we assess it relative to $\mathcal{G}$:

A proposition $\underline{P}$ is _warranted to degree_ δ relative to $\mathcal{G}$ iff $\underline{P}$ is justified to degree δ relative to $\mathcal{G}_{\omega}$.

Given a set $\mathcal{G}$ of conclusions, let $\mathcal{G}_{\text{percepts}} = \{\alpha \mid \alpha \in \mathcal{G}$ and α encodes a percept$\}$. All reasoning begins from percepts, so we have the following simple theorem:

Theorem: $\mathcal{G}_{\omega} = (\mathcal{G}_{\text{percepts}})_{\omega}$.

The set of justified beliefs reflects the current state of the agent's reasoning, which is never complete, so the set of justified beliefs will not be closed under logical entailment, and may not even be logically consistent. By contrast, however, the set of warranted propositions takes account of all possible reasoning, so:

Theorem: If Γ is a set of propositions warranted to at least degree δ and $\Gamma \quad \underline{P}$ then $\underline{P}$ is warranted to degree at least δ.

4. Belief Revision Reconsidered

Now let us reconsider theories of belief revision in light of the epistemological theory described in Section 3. A theory of belief revision is supposed to pertain to the way in which beliefs change as new beliefs are adopted. We are presumably not interested in beliefs *per se*. Our interest is in either justified beliefs or warranted propositions, and rational changes to them. If our interest is in justified beliefs, there are two ways we can be led to new beliefs encoding new information. The information can be provided by perception, or it can result from reasoning from previously held beliefs. In either case, new conclusions can become inferrable, and old conclusions may be defeated. However, it is doubtful that there is any mathematical theory of belief revision which models the dynamics of justified beliefs. The difficulty is that the set of justified beliefs can exhibit all kinds of (as yet undiscovered) logical incoherences because it represents an intermediate stage in reasoning. To get a mathematically interesting theory, we should at least let reasoning finish in order to clean up the set of justified beliefs as much as possible. In other words, a theory of belief revision should be concerned with warrant rather than justification. This accords nicely with Gärdenfors' remark that he is concerned with belief states that are "in equilibrium under all forces of internal criticism". This suggests defining, much in the spirit of Alchourrón, Gärdenfors, and Makinson:

> $\underline{K}$ is a *belief-set* iff there is an inference-graph $\mathcal{G}$ such that $\underline{K}$ is the set of all propositions warranted relative to $\mathcal{G}$.

'*' is the central AGM belief revision operator. Given a belief-set $\underline{K}$, $\underline{K}*\underline{P}$ is supposed to be the belief-set that results from adding the new information $\underline{P}$ and making whatever other adjustments to the belief-set (both additions and deletions) that this requires. In Section 2 we followed AGM in supposing uncritically that this makes sense. But now let us be more careful. What, exactly, is it to add a proposition to a belief-set? As remarked above, there are two ways of acquiring new justified beliefs—from perception, or from inference. But the set of warranted propositions already takes account of all possible inferences, so there is only one way to acquire new warranted propositons—through perception. There is an immediate problem here, however. Perception adds a percept to the inference-graph, not a belief, so the effect of perception on warrant cannot be represented as the addition of a proposition to a belief-set.

The whole point of perception is to make beliefs to the effect that the world is as perceived defeasibly reasonable. So we might construe adding a proposition $\underline{P}$ to consist of adding a percept α such that prop(α) = _P_. Most propositions cannot be the contents of percepts, so this severely restricts what kinds of propositions can be added, but we think that this is a sensible way of proceeding. However, it creates immediate problems for the AGM theory. For example, one of the simplest AGM postulates is the "success" postulate:

(_K_*2) $\underline{P} \in \underline{K}*\underline{P}$.

This says that if you add new information to the effect that _P_, then _P_ should be among the resulting warranted propositions. However, if adding _P_ consists merely of adding a percept defeasibly supporting an inference to _P_, there is no guarantee that other members of the belief-set may not provide a defeater for the inference to _P_. For example, I might have a percept of pink elephants but know perfectly well that I am hallucinating. Then adding the percept does not add the proposition that there are pink elephants in my vicinity to the set of warranted propositions. In this case $\underline{P} \notin \underline{K}*\underline{P}$.

We could get around this problem by understanding adding _P_ as consisting of adding a percept defeasibly supporting _P_ _under circumstances in which the inference to P is not defeated_. Then it is automatically the case that $\underline{P} \in \underline{K}*\underline{P}$. This has the effect, however, that '_K_*_P_' is defined for some choices of _K_ and not not for others. Specifically, it is undefined when _K_ contains an undercutting defeater for the inference from a percept of _P_ to _P_.

Perhaps this problem can be circumvented by understanding the input of new information more broadly. Gärdenfors remarks that much of our information comes to us from third person reports rather than directly from perception. We argued in section three that this is still based indirectly on perception and reasoning. Perception gives us a reason for thinking that something is being reported to us, and induction gives us a reason for thinking that what is reported to us is true. We might, however, understand adding _P_ as consisting of adding some percepts that make it defeasibly reasonable to believe _P_ either on the basis of perception or on the basis of report. But why stop there? Perhaps more generally we should think of adding _P_ as consisting of adding percepts that make it defeasibly reasonable to believe _P_ on the basis of some reasoning or other. It seems doubtful that _K_ could contain information defeating all possible bases for _P_, so if we are sufficiently noncommital about the basis for _P_, we can regard _P_ as undefeated in _K_*_P_.

This will not work, however. Different bases for _P_ will produce different sets of warranted propositions. If we are noncommital about the basis for _P_, we cannot determine which set of warranted propositions should be produced. In other words, _K_*_P_ will not be a belief-set in the sense defined above, and '*' will not be a revision _function_ at all. What is important here is that you cannot just add propositions to the set of warranted beliefs willy nilly, if the result of such an addition is supposed to make the added propositions warranted. You must add the proposition along with some (minimal) perceptual basis adequate to make it warranted.[6] However, there will generally be many (probably infinitely many) ways of doing this, and they will not result in the same overall set of warranted propositions. Thus there will be no unique set $\underline{K}*\underline{P}$ that results from adding $\underline{P}$ to $\underline{K}$ in this sense.

We could try various technical tricks for avoiding this problem. For example, we might let $\underline{K}*\underline{P}$ consist of the propositions that are common to all sets of warranted propositions that would result from the different minimal ways of adding a perceptual basis for $\underline{P}$. $\underline{K}*\underline{P}$ will no longer be a belief-set, but perhaps that is not a serious problem. A more serious problem is that all contingent members of $\underline{K}$ will be eliminated from $\underline{K}*\underline{P}$. Let $\underline{Q}$ be any contingent member of $\underline{K}$. Then one way of getting $\underline{P}$ warranted is by adding percepts supporting $\sim\underline{Q}$ and $(\sim\underline{Q} \supset \underline{P})$, but that will result in deleting $\underline{Q}$.

It seems to us that the best way around many of these difficulties is to add not just $\underline{P}$, but an entire argument (starting from percepts) that supports $\underline{P}$. On this account, belief revision becomes a function from Belief-Sets × Arguments to Belief-Sets rather than from Belief-Sets × Propositions to Belief-Sets.

This still will not quite do. The difficulty is that we have defined belief-sets to be sets of (warranted) propositions. The same set of propositions can be associated with different epistemological states in which the same propositions are warranted but have different degrees of warrant. Adding the same argument to these different epistemological states may produce different revised sets of warranted propositions. For example, suppose the original epistemological state includes an argument for $\sim\underline{P}$ with some degree of warrant δ, and we add an argument for $\underline{P}$ of strength δ^*. If the argument for $\underline{P}$ is not otherwise defeated by conclusions in the original epistemological state, then the revised set of warranted propositions will contain $\underline{P}$ iff $\delta^* > \delta$. So adding $\underline{P}$ to one such epistemological state may have the effect of inserting $\underline{P}$ into the revised set of warranted propositions, while adding $\underline{P}$ to another epistemological state containing the same warranted propositions but different degrees of warrant may fail to have that effect.

This indicates that simply looking at the set of warranted propositions does not tell us enough about the epistemological state represented to enable us to compute the effect of adding an argument for $\underline{P}$. We must adopt a richer characterization of epistemological states. First, we must at least attach degrees of warrant to the warranted propositions so that we can compute the result of adding defeaters.[7] Second, percepts are not beliefs, but they must be incorporated into the representation of the epistemological state as well. Perhaps the simplest way to do all this is to eschew belief-sets and identify the epistemological state with the inference-graph itself.

In Section 3, arguments were represented as sets of "argument-steps", the latter being triples ⟨premises, conclusion, reason-schema⟩. But notice that an essentially equivalent representation of arguments is to identify them with inference-graphs consisting of just the percepts used and the conclusions drawn in the argument. Let us say that an inference-graph $\mathcal{A}$ is an argument for $\underline{P}$ iff $\underline{P}$ is justified relative to $\mathcal{A}$ but not relative to any inference-graph that is a proper subgraph of $\mathcal{A}$.[8] Then if we also represent an epistemic agent's epistemological state as an inference-graph $\mathcal{G}$, the result of revising that state by incorporating an argument $\mathcal{A}$ for $\underline{P}$ is represented by the inference-graph $\mathcal{G}\cup\mathcal{A}$. We can then determine what propositions become warranted as a result of the revision and what propositions become unwarranted by employing our standard epistemological machinery and computing the status-assignments for $(\mathcal{G}\cup\mathcal{A})_{\omega}$.

Given an inference-graph $\mathcal{G}$, let $K_{\mathcal{G}}$ be the set of propositions warranted with respect to $\mathcal{G}$. $K_{\mathcal{G}}$ is a belief-set in the sense defined above. What we are proposing is that the revision operator '*', construed as operating on belief-sets, doesn't really make sense. Belief-sets do not contain enough information to allow us to determine what is warranted if we add an argument for a new conclusion. Instead of asking how $\underline{K}*\underline{P}$ differs from $\underline{P}$, what we must do is consider specific arguments $\mathcal{A}$ for $\underline{P}$, and ask how $K_{\mathcal{G}\cup\mathcal{A}}$ differs from $K_{\mathcal{G}}$. This is the only sensible way to talk about belief-revision.

The AGM Theory

Now let us reconsider the AGM theory in light of the preceding remarks about belief-revision. Suppose we begin with an epistemological state represented by an inference-graph $\mathcal{G}$ and a set of warranted propositions $K_{\mathcal{G}}$. We want to know what happens when we revise that epistemological state by adding an argument $\mathcal{A}$ for $\underline{P}$. Let $\underline{C}_{\mathcal{A}}$ be the set of all propositions supported by nodes of $\mathcal{A}$. This includes $\underline{P}$, and all the intermediate conclusions drawn in $\mathcal{A}$. The central idea behind the AGM theory is that there are two cases of belief revision to consider. First, $\underline{C}_{\mathcal{A}}$ may be consistent with $\underline{K}$. In that case, we simply add those conclusions to $K_{\mathcal{G}}$ and then add anything else that can be inferred from $K_{\mathcal{G}}\cup\underline{C}_{\mathcal{A}}$ by undefeated arguments. The supposedly more difficult case occurs when some subset of $\underline{C}_{\mathcal{A}}$ is inconsistent with $K_{\mathcal{G}}$. Then the proposal is that we first remove the elements of $K_{\mathcal{G}}$ inconsistent with that subset, and everything that depends upon them, and then add the elements of $\underline{C}_{\mathcal{A}}$ as in the first case. Formally, AGM attempted to capture this by defining two new operators:

$$\underline{K}+\underline{P} = \{\underline{Q} \mid \underline{K}\cup\{\underline{P}\} \quad \underline{Q}\} \quad \text{(expansion)}$$

$$\underline{K}-\underline{P} = \underline{K} \cap (\underline{K} * \sim\underline{P}) \quad \text{(contraction)}$$

and then proposing

$$\underline{K}*\underline{P} = (\underline{K} - \sim\underline{P}) + \underline{P}.$$

The definitions of expansion and contraction both require revision in light of the preceding observations. Because we cannot add $\underline{P}$ without adding an argument $\mathcal{A}$ for $\underline{P}$, it seems that the definition of expansion should at least be revised as follows:

$$\underline{K}+\mathcal{A} = \{\underline{Q} \mid \underline{K}\cup\underline{C}_{\mathcal{A}} \quad \underline{Q}\}.$$

However, this will not quite do. Even if $\mathcal{A}$ can be added to $\underline{K}$ without deleting any members of $\underline{K}$, we cannot form a belief-set just by closing $\underline{K}\cup\underline{C}_{\mathcal{A}}$ under logical entailment. The addition of $\mathcal{A}$ may make available undefeated defeasible arguments for other conclusions, which will also have to be incorporated into the set of warranted propositions.

Expansion is perhaps best viewed as a proposal regarding how to compute the set of warranted propositions relative to $\mathcal{G}\cup\mathcal{A}$. Rather than closing $\underline{K}\cup\underline{C}_{\mathcal{A}}$ under logical entailment, we must close it under all possible defeasible and non-defeasible reasoning, which is the same as constructing $(\mathcal{G}\cup\mathcal{A})_{\omega}$. What is novel about expansion is that in computing $\underline{K}_{(\mathcal{G}\cup\mathcal{A})}$, we can assume that the defeat-statuses of elements of the original inference-graph $\mathcal{G}$ remain unchanged, and so $K_{\mathcal{G}} \subseteq \underline{K}_{(\mathcal{G}\cup\mathcal{A})}$. This suggests that expansion should be considered a special case of belief revision, but not one that is captured by a special operator. We can regard a modified form of the AGM theory as proposing more simply that if $\underline{C}_{\mathcal{A}}$ is consistent with $K_{\mathcal{G}}$ then $K_{\mathcal{G}} \subseteq \underline{K}_{(\mathcal{G}\cup\mathcal{A})}$.

Regrettably, even this latter claim is false. Consistency alone is not sufficient to guarantee that $\mathcal{A}$ can be added to $\mathcal{G}$ without retracting any members of $K_{\mathcal{G}}$. The difficulty is that consistency only guarantees the absence of rebutting defeat. Adding $\mathcal{A}$ can generate arguments for undercutting defeaters for some of the arguments for members of $K_{\mathcal{G}}$, in which case the latter will no longer be in the set of warranted propositions after we add $\mathcal{A}$. There is a true claim lurking here however. If adding $\mathcal{A}$ to $\mathcal{G}$ does not allow us to construct arguments for any defeaters for members of $\mathcal{G}$, then $K_{\mathcal{G}} \subseteq K_{\mathcal{G}\cup\mathcal{A}}$. Precisely:

Theorem: If $(\mathcal{G}\cup\mathcal{A})_{\omega}$ does not contain any argument for a defeater for a member of $\mathcal{G}$ that is not already contained in $\mathcal{G}_{\omega}$, then $K_{\mathcal{G}} \subseteq K_{\mathcal{G}\cup\mathcal{A}}$.

This, we take it, is what is true of expansion.

Now let us turn to contraction. Contraction was defined in terms of '$\underline{K} * \sim\underline{P}$', which we have concluded to be ill-defined. The difficulty is the familiar one that we cannot add $\sim\underline{P}$ to an epistemological state without adding an argument for it as well. We could attempt to repair the definition in various ways to accommodate this, but it seems reasonably clear that this approach to the definition of contraction will not work anyway. $\underline{K}-\underline{P}$ is supposed to be the result of removing $\underline{P}$ and anything depending upon $\underline{P}$ from $\underline{K}$. However, removing $\underline{P}$ is quite different from adding an argument for $\sim\underline{P}$. The latter might, for example, enable the construction of an argument for an undercutting defeater for an argument supporting some member $\underline{Q}$ of $\underline{K}$, in which case $\underline{Q}$ would not be in ($\underline{K} * \sim\underline{P}$), and hence not in $\underline{K} \cap (\underline{K} * \sim\underline{P})$, even though the argument for $\underline{Q}$ does not depend on $\underline{P}$. It seems better to define contraction directly. Given an inference-graph $\mathcal{G}$, removing $\underline{P}$ should consist of removing all arguments for $\underline{P}$ that occur in $\mathcal{G}_\omega$. Recall that arguments for $\underline{P}$ are *minimal* inference-graphs supporting $\underline{P}$. So let us define contraction as a function on inference-graphs and propositions:

$$\mathcal{G} - \underline{P} = \mathcal{G}_\omega - \{\mathcal{A} \mid \mathcal{A} \subseteq \mathcal{G}_\omega \text{ \& } \mathcal{A} \text{ is an argument for } \underline{P}\}.$$

Then the corresponding belief-set is $\underline{K}_{(\mathcal{G}\,-\,\underline{P})}$. More generally, where $\underline{X}$ is a set of propositions, let us define:

$$\mathcal{G} - \underline{X} = \mathcal{G}_\omega - \{\mathcal{A} \mid \mathcal{A} \subseteq \mathcal{G}_\omega \text{ \& } \mathcal{A} \text{ is an argument for some member of } \underline{X}\}.$$

Now consider the central claim of the AGM theory, which is that belief revision can be composed into first a contraction and then an expansion (i.e., the Levi Identity). This was formulated as:

$$\underline{K}*\underline{P} = (\underline{K} - \sim\underline{P}) + \underline{P}.$$

We have concluded that none of these operators is well-defined, but perhaps the same idea can be expressed differently. The idea is that in computing $\underline{K}_{(\mathcal{G}\cup\mathcal{A})}$, where $\mathcal{A}$ is an argument for $\underline{P}$, we first perform a contraction to remove the elements of $\mathcal{G}$ inconsistent with $\mathcal{A}$, and then we add $\mathcal{A}$ by expansion. Let $\sim\Pi\underline{C}_\mathcal{A}$ be the negation of the conjunction of the members of $\underline{C}_\mathcal{A}$ The first step of this construction would yield $\mathcal{G} - \sim\Pi\underline{C}_\mathcal{A}$, and then the second step would yield $(\mathcal{G} - \sim\Pi\underline{C}_\mathcal{A})\cup\mathcal{A}$. It is then claimed (1) that the second step is a case of expansion, and (2) that $((\mathcal{G} - \sim\Pi\underline{C}_\mathcal{A})\cup\mathcal{A})_\omega = (\mathcal{G}\cup\mathcal{A})_\omega$.

Unfortunately, both of these claims fail for what should now be familiar reasons. First, this is only a case of expansion if adding $\mathcal{A}$ to $(\mathcal{G} - {\sim}\Pi\underline{C}_{\mathcal{A}})$ does not allow us to construct new arguments for defeaters for any members of $(\mathcal{G} - {\sim}\Pi\underline{C}_{\mathcal{A}})_{\omega}$. By deleting the arguments for ${\sim}\Pi\underline{C}_{\mathcal{A}}$, we guarantee that there are no rebutting defeaters, but we do not rule out the possibility of undercutting defeaters. This difficulty is easily circumvented. Instead of deleting arguments for ${\sim}\Pi\underline{C}_{\mathcal{A}}$, we want to delete arguments that, with the help of the members of $\underline{C}_{\mathcal{A}}$, support defeaters for elements of $\mathcal{G}$. There is an argument in $(\mathcal{G}\cup\mathcal{A})_{\omega}$ for a conclusion $\underline{D}$ iff there is an argument in $\mathcal{G}_{\omega}$ for $(\Pi\underline{C}_{\mathcal{A}} \supset \underline{D})$. So the contraction step should proceed by forming

$$\mathcal{G} - \{(\Pi\underline{C}_{\mathcal{A}} \supset \underline{D}) \mid \underline{D} \text{ is a defeater for a conclusion in } \mathcal{G}\}.$$

Call this '$\mathcal{G} \sim \mathcal{A}$'. If we then construct $(\mathcal{G} \sim \mathcal{A})\cup\mathcal{A}$, it is genuinely a case of expansion.

Is it true, however, that $((\mathcal{G} \sim \mathcal{A})\cup\mathcal{A})_{\omega} = (\mathcal{G}\cup\mathcal{A})_{\omega}$? Regrettably not. By removing the arguments for all defeaters for members of $\mathcal{G}$, we guarantee that adding $\mathcal{A}$ will not necessitate the deletion of any further members of $(\mathcal{G} \sim \mathcal{A})$. However, if the degrees of warrant are appropriate, adding $\mathcal{A}$ to $\mathcal{G}$ may instead leave some of those defeating arguments intact and instead necessitate the rejection of the conclusions the defeating arguments defeat. In constructing $\underline{K}_{(\mathcal{G}\cup\mathcal{A})}$, we may have to delete members of $K_{\mathcal{G}}$, but there is no guarantee that the right things to delete are the elements of the arguments for the conditionals $(\Pi\underline{C}_{\mathcal{A}} \supset \underline{D})$.

Our conclusion is that there is no apparent way to define expansion and contraction so that belief-revision can always be regarded as a sequential contraction and then expansion. The epistemological facts are just too complicated for that. It appears that the AGM theory is led astray by adopting too abstract a model of epistemological states. Very little epistemological mileage can be gotten out of just considering the set of believed propositions and logical relations between them. Epistemological relationships, and correspondingly the dynamics of belief revision, are based upon a much richer structure of reason-schemas, defeaters, and degrees of justification, and an abstract theory of belief revision, should such a thing be possible, must take account of all aspects of this structure.

Acknowledgements

This research was supported by NSF grants nos. IRI-9634106 and DGE-9355028.

References

Alchourrón, C. E., P. Gärdenfors, and D. Makinson: 1985, 'On the logic of theory change: Partial meet functions for contraction and revision', _Journal of Symbolic Logic_ **50**, 510-530.

Alchourrón, C. E., and D. Makinson: 1982, 'The logic of theory change: Contraction functions and their associated revision functions', _Theoria_ **48**, 14-37.

Alchourrón, C. E., and D. Makinson: 1985, 'On the logic of theory change: Safe contraction', _Studia Logica_ **44**, 405-422.

Boutilier, C.: 1994, 'Unifying default reasoning and belief revision in a modal framework', _Artificial Intelliegence_ **68**, 33-85.

Boutilier, C., N. Friedman, and J. Y. Halpern: 1998, 'Belief revision with unreliable observations', _Proceedings of AAAI-98._, 127-134.

Darwiche, A. and J. Pearl: 1997, 'On the logic of iterated belief revision', _Artificial Intelligence_ **89**, 1-29.

DePaul, M. (ed.): 1999, _Foundationalism_, Rowman and Littlefield, Savage, Maryland.

Furhmann, A.: 1991, 'Theory contraction through base contraction', _Journal of Philosophical Logic_ **20**, 510-530.

Gärdenfors, P.: 1988, _Knowledge in Flux_, MIT Press, Cambridge, MA.

Gärdenfors, P.: 1992, 'Belief revision: a vade-mecum', in A. Pettorossi (ed.), _Meta-Programming in Logic_, Springer, Berlin, 1-10.

Harper, W. L.: 1977, 'Rational conceptual change', in _PSA 1976, Volume 2._, Philosophy of Science

Association, East Lansing, MI, 462-494.

Levi, I.: 1988, 'Iteration of conditionals and the Ramsey Test', *Synthese* **76**, 49-81.

Pollock, J. L.: 1987, 'Defeasible reasoning', *Cognitive Science* **11**, 481-518.

Pollock, J. L.: 1990, *Nomic Probability and the Foundations of Induction*, Oxford University Press, New York, NY.

Pollock, J. L.: 1992, 'How to reason defeasibly', *Artificial Intelligence* **57**, 1-42.

Pollock, J. L.: 1995, *Cognitive Carpentry*, MIT Press, Cambridge, MA.

Pollock, J. L.: 1997, 'Reasoning about change and persistence: A solution to the frame problem', *Nous* **31**(2), 143-169.

Pollock, J. L.: 1998, 'Perceiving and reasoning about a changing world', *Computational Intelligence* **14**, 498-562.

Pollock, J. L.: 1998a, 'Degrees of justification', in Weingartner, Schurz, and Dorn (eds.), *The Role of Pragmatics in Contemporary Philosophy: Proceedings of the 20th International Wittgenstein Symposium*, Hoelder-Pichler Tempsky Publishers, Vienna.

Pollock, J. L.: 1999, 'Means-end reasoning', in R. Elio (ed.), the proceedings of the 1998 Vancouver Cognitive Science Conference, Oxford University Press, forthcoming.

Pollock, J. L.: 1999a, 'The logical foundations of goal-regression planning', *Artificial Intelligence*.

Spohn, W.: 1987, 'Ordinal conditional functions: A dynamic theory of epistemic states', in Harper and Skyrms (eds.), *Causation in Decision, Belief Change and Statistics, Volume 2*, Dorecht, Reidel, 105-134.

Notes

[1] Harper (1977), Stalnaker (1984), Levi (1988), Spohn (1987), Furhmann (1991), Boutilier (1994).

[2] For a general discussion of different kinds of foundationalist theories, see DePaul (1999).

[3] The correct statement of the principle requires further qualifications. See Pollock (1990) for the details.

[4] In Pollock (1995), arguments are taken to support sequents rather than propositions, but sequents and conditionals are interchangeable, so this definition will work equally well.

[5] For an alternative account, see Pollock (1998a).

[6] Furhmann (1991) makes a related observation.

[7] Spohn (1987) makes a similar observation.

[8] That is, where $\mathcal{G}$ is the ordered pair of the set $\underline{N}_{\mathcal{G}}$ of nodes and $\underline{L}_{\mathcal{G}}$ the set of links, $\mathcal{A}$ is a proper subgraph of $\mathcal{G}$ iff $\underline{N}_{\mathcal{A}} \subset \underline{N}_{\mathcal{G}}$ and $\underline{L}_{\mathcal{A}} \subset \underline{L}_{\mathcal{G}}$. We represent this by simply writing that $\mathcal{G} \subset \mathcal{A}$.