

WOLFGANG SPOHN

## A BRIEF COMPARISON OF POLLOCK'S DEFEASIBLE REASONING AND RANKING FUNCTIONS

**ABSTRACT.** In this paper two theories of defeasible reasoning, Pollock's account and my theory of ranking functions, are compared, on a strategic level, since a strictly formal comparison would have been unfeasible. A brief summary of the accounts shows their basic difference: Pollock's is a strictly computational one, whereas ranking functions provide a regulative theory. Consequently, I argue that Pollock's theory is normatively defective, unable to provide a theoretical justification for its basic inference rules and thus an independent notion of admissible rules. Conversely, I explain how quite a number of achievements of Pollock's account can be adequately duplicated within ranking theory. The main purpose of the paper, though, is not to settle a dispute within formal epistemology, but rather to emphasize the importance of formal methods to the whole of epistemology.

### 1. INTRODUCTION

Formal epistemology could have a better standing within philosophical epistemology than it actually has. One half is Bayesianism, i.e., probability theory, which is silent, though, on the most basic notion of philosophical epistemology, the notion of belief, and replaces it by many degrees of belief. The other half is a divided lot of theories, rather at home in computer science and hardly perspicuous for interested philosophers. Doxastic and epistemic logic as introduced by Hintikka (1962) is a common background to those theories, but an insufficient one as soon as changes of beliefs, inductive and defeasible reasoning and kindred things are at issue. Quite a number of those theories even originate from philosophers. Deplorably, however, they appear to proceed as separate research programs, hardly knowing of each other and rather trying to find allies in computer science. If this is the appearance, why should the epistemologist care to attend to that theoretical diversity?

It's not easy to join forces; after all, there is theoretical disagreement. But then there should at least be a joint market for internal exchange and with the external message that there, and only there, a lot of things are offered which the epistemologists urgently need.



*Synthese* **131**: 39–56, 2002.

© 2002 Kluwer Academic Publishers. Printed in the Netherlands.

When speaking so critically, I am mainly thinking of three theories, which definitely do not exhaust formal epistemology beyond probability even within philosophy,<sup>1</sup> but form a good part of it: namely, (1) John Pollock's theory of defeasible reasoning which he has been working to develop and augment for more than 30 years,<sup>2</sup> always in close contact to philosophical epistemology, but sadly in an imperspicuous formal shape; (2) the Alchourrón-Gärdenfors-Makinson (AGM) theory of belief revision,<sup>3</sup> started by Gärdenfors in the mid 1970's, originally associated rather with the logic of counterfactuals and the philosophy of explanation, meanwhile by far the largest in terms of man and paper power, but the least connected to philosophy;<sup>4</sup> and (3) the theory of ranking functions, which I developed in 1982/3 from Gärdenfors' theory, which, however, resembles probability theory much more than belief revision theory, which I have since tried to put to various philosophical uses, and which I distinguish here as a third strand for understandable reasons.<sup>5</sup>

Since belief revision theory and ranking functions are close neighbors, their relation is more or less fully understood.<sup>6</sup> Pollock and Gillies (2001) have commendably compared their theory with belief revision theory. This is highly illuminating. But, of course, the comparison is carried out from Pollock's perspective; I strongly hope for a reply. In this paper, I would like to add a comparison of Pollock's theory with ranking theory. The main aim is to help in creating and enriching the joint market we urgently need.

At first, I thought my comparison could get down to the formal level. However, this turned out to be unfeasible, due to the very different formal formats. Hence, I shall carry out the comparison at a more strategic level. This is not a mere expedient; it is rather only at this level where the more general motivations and conceptions behind the various theories become perspicuous.

In Sections 2 and 3 I shall give a brief informal sketch of Pollock's defeasible reasoning and of ranking theory. This brings to the fore, in Section 4, what I perceive to be the basic difference, namely that Pollock's theory is entirely computational, whereas mine is located at a regulative level. Section 4 goes on to argue that Pollock's theory cannot bridge this difference and that its normative condition is therefore deficient. Section 5 shows, by contrast, that ranking theory provides ample means to bridge the difference.

## 2. A BRIEF SKETCH OF POLLOCK'S THEORY

Pollock draws a large and detailed picture of doxastic states as huge nets or *graphs of inferences*, reasons, justifications, or *arguments*. Each argu-

ment consists of a set of *premises*, a *conclusion*, and an inference rule or *reason-schema* applied in the argument. There are two kinds of arguments. The first are *conclusive* or non-defeasible arguments, which we know well enough from deductive logic. The essence of Pollock's theory lies in the second kind, the *defeasible* arguments, which realize defeasible reason-schemata. They are not deductively valid, but only they get our impressive inductive machinery running.

That a reason-schema or the arguments realizing it are defeasible is to say that they have *defeaters*. Therefore they need to be amended by a specification of their defeaters. They come in two kinds. There are *rebutting* defeaters; these are arguments arriving at the opposite conclusion. And there are *undercutting* defeaters; they undermine the connection between premises and conclusion of the defeated argument and hence conclude that the conclusion might be wrong despite the premises being true.

Of course, defeating arguments may in turn be defeated in both ways. And all this may mesh with conclusive arguments. In this way, a big and complicated architecture of more or less provisional justifications emerges.<sup>7</sup>

This is the formal picture. What's the good of it? Well, it is to be filled, and that's what Pollock amply does. The picture has a start, a substance, a form, and a goal.

The *start* is *perception*. Without perception there are no premises to start with and no conclusions to arrive at. Perceptions form the indispensable base of the whole wide web of reasons and beliefs. But the base is defeasible, and it is governed by the following defeasible reason-schema: if  $p$  is a perceptible state of affairs, then infer  $p$  from "it looks as if  $p$ ". This argument step moves us from phenomenal premises to a conclusion about the external world.<sup>8</sup> Having arrived there, the inductive machinery can gather full speed.

The *substance* is provided by the many specific defeasible reason-schemata Pollock proposes.<sup>9</sup> That there is defeasible as well as conclusive reasoning is by now common place in epistemology. How profoundly our epistemological picture thereby changes is well understood by many less people. But Pollock is still more or less the only one making specific proposals for a constructive theory of defeasible reasoning.

We have already seen an example, the rule governing perception. It is defeasible and thus accompanied by potential defeaters. Pollock states two: one for the case where the subject perceives something, but believes to perceive something else, and one for the case that the subject believes to be in unreliable perceptual circumstances.

But there are quite a number of further defeasible inference rules. Just to give the flavor, the most important one for Pollock is the statistical syllogism: if  $G$  is projectible with respect to  $F$ , then infer  $Ga$  from  $Fa$  and “the probability of  $F$ ’s being  $G$  is greater than 0.5” (where the strength of the inference depends on that probability). Of course, this may be defeated, most importantly by a subproperty defeater: if  $G$  is projectible also with respect to  $H$ , then the above inference is undercut by  $Ha$  and “the probability of  $F$ -and- $H$ ’s being  $G$  differs from that of  $F$ ’s being  $G$ ”. There are rules for enumerative and statistical induction arriving at universal and probabilistic generalizations. And so on.

Then, there is a *form* that provides rules for combining the many arguments a subject has in mind into an integrated inference graph. Individual arguments have strengths, and a formal theory is required for specifying the strengths within a complex argument structure. Arguments can be defeated, and a formal theory is required for determining how the defeats run in the integrated graph. And so on.<sup>10</sup> However, without the substance of the specific reason-schemata this form would remain empty.

Finally, the *goal* of all this reasoning is to arrive at a set of justified and warranted beliefs. Prima facie, it is not at all clear whether the goal can be reached. There are two issues. First, one wonders whether, given a certain stage of reasoning with its multiply entangled defeatings, the various conclusions may be unambiguously said to be undefeated or to be defeated. All kinds of crazy situations are imaginable, and Pollock has long struggled with them, as his changing explications display. But his present theory of defeat status assignments seems to get the issue under control.<sup>11</sup>

A *justified* belief, then, is a conclusion that comes out undefeated under this theory. Justification is here still relative to a certain stage of reasoning. Rightly so, according to Pollock, because a subject is always at, and acts always out from, an unfinished state of reasoning. This suggests, however, that there is also a stronger notion of justification which Pollock calls ideal warrant. The defeat statuses may change in unforeseen ways as soon as a further argument is considered. Hence, the subject should stepwise extend his inference graph, until all (possibly infinitely many) arguments in principle available to him are considered. The second issue, then, is whether this process is at all well behaved. There is no guarantee. But Pollock defines a conclusion to be *ideally warranted* if it is unambiguously undefeated in the maximal inference graph in which the subject, *per impossibile*, takes into account the arguments available to him all at once. The stepwise extension of the inference graph, then, is well behaved if it eventually arrives at precisely the ideally warranted conclusions or

beliefs, and Pollock specifies conditions under which this process is so well behaved.<sup>12</sup>

In this way, Pollock draws a detailed dynamic picture of reasoning and its goal. In another sense, the picture is still static; the whole edifice of reasoning and thus the set of ideally warranted beliefs rest on a given input of perceptions. The question usually addressed in belief revision theory, however, is what happens to the beliefs when the input changes; only by answering this question do we acquire a fully dynamic picture. But Pollock has no difficulties in principle with this question; he can enrich the stock of perceptions and set his reasoning machinery in motion again, and then new sets of justified and ideally warranted beliefs will result.<sup>13</sup>

This may suffice as a description of Pollock's theory for the present purpose.

### 3. A BRIEF SKETCH OF RANKING FUNCTIONS

The theory of ranking functions starts where Pollock's theory ends; it is a theory about how to change beliefs in view of new pieces of information. One may well suspect, then, the theories to be incomparable; but wait and see.

At the outset, I should emphasize that one should not narrow down the reception of new information to perception. Initially, the information must flow through sensory channels, no doubt. But it should be possible to study the statics and dynamics of the beliefs in a certain restricted field or domain having a relative input boundary; there should be a general theory of how such a restricted field is acted upon by its boundary. Of course, changes of (e.g., new beliefs in) that boundary ultimately result from some perceptual input; but it should be possible to remain silent about the process leading from the absolute to the relative input. It would be awkward if every account of the restricted field has to turn on the complete doxastic architecture.<sup>14</sup> I think it is important for belief revision theories to maintain generality in this respect, and therefore I shall more neutrally speak of new pieces of information instead of perceptions. Hence I do not agree with the criticism of Pollock and Gillies (2001, sect. 4) that AGM belief revision theory does not start from scratch by totally neglecting perception.<sup>15</sup> This seems to be one major point of divergence between us.

Having cleared away this point, let us move to ranking theory. Its task is a very simple one. It is to characterize doxastic states in such a way that (a) they contain plain beliefs which can be true or false, and that (b) general and full dynamic laws can be stated for them. In these tasks it succeeds.

Task (a) suggests that we should consider belief contents *only* insofar they can be true or false, i.e., to conceive them as truth conditions or propositions. Thereby, one ignores all questions of syntactic structure, of logical equivalence, and of logical entailment, and trivializes the rationality constraints of consistency and deductive closure of the beliefs at a given time. Pollock does not assume this, but I do, not because I do not see a problem here, but because I don't have anything to say about the problem.<sup>16</sup>

Task (a) excludes probability theory. Having a subjective probability of 0.7, say, for some proposition is simply not something that can be true or false; only the belief, or the disbelief, in this proposition can. That subjective probability does not explain plain belief is highlighted, of course, by the famous lottery paradox. I don't want to claim that there is no good solution to the lottery paradox, but the mere fact that all attempts are debated heatedly and no solution is easily accepted shows that probability theory is, presently, not a good base to tackle task (a).

Doxastic logic is good enough for task (a); so may be various kinds of default logic, and so is definitely AGM belief revision theory. However, task (b) is to be achieved as well. There is no point here in discussing all the candidates for task (b). But clearly, doxastic logic is not good enough; it was never meant to take a dynamic perspective. And despite appearances AGM belief revision theory is not good enough, either. A full dynamics has to account for several or iterated belief changes. But standard AGM theory accounts generally only for one step; after that, belief states are characterized simply as belief sets and are dynamically as barren as doxastic logic.<sup>17</sup> Though the problem of iterated belief revision has been around since Harper (1976) and though there have been quite a number of attempts to solve it within the confines of AGM theorizing,<sup>18</sup> I cannot find any of these attempts to be satisfying.

The problem *is* solved by the theory of ranking functions which I have proposed in Spohn (1988).<sup>19</sup> I do not claim that it is the only solution or that any solution must be somehow equivalent to mine; this would be presumptuous. But I think my paper at least suggests that no weaker theory will do.

How does the theory work? *Ranks are grades of disbelief* (where I find it natural to take non-negative integers as grades, but other non-negative numbers would do as well). For a proposition to have rank 0 means to be not disbelieved, i.e., to be held true or neither true nor false. Having a rank larger than 0 means to be disbelieved; the larger the rank, the firmer the disbelief. Thus task (a) is achieved: that a proposition is believed according to a ranking function means that its negation has a rank  $> 0$ .

This characterization entails certain laws: the *law of negation* that each proposition or its negation (or both) receive rank 0 (they cannot be both disbelieved), and the *law of disjunction* that the rank of *A-or-B* is the minimum of the ranks of *A* and *B* (*A-or-B* cannot be more firmly disbelieved than either of its disjuncts, but also not less firmly than both disjuncts, because if both, *A* and *B*, are disbelieved, *A-or-B* must be so as well). Thus starts a formal theory.

The main point, though, is the definition of conditional ranks. The *rank of B given A* is the rank of *A-and-B* minus the rank of *A*. Equivalently, this is the *law of conjunction*: the grade of disbelief in *A-and-B* is the sum of the grade of disbelief in *A* and the conditional grade of disbelief in *B* given *A*. From there we may proceed to a notion of doxastic independence: *A* and *B* are independent iff the rank of *B* is not affected by conditioning it to *A* or non-*A*; similarly for conditional independence.

These notions behave in almost exactly the same way as their probabilistic counterparts.<sup>20</sup> The basic rule for probabilistic belief change is simple conditionalization according to which one moves to the probabilities conditional on the information received. This is generalized by Jeffrey conditionalization<sup>21</sup> that is unrestrictedly performable and thus defines a full dynamics within the realm of strictly positive probabilities. All this immediately carries over to ranking functions, and hence the rule of belief change in terms of ranking functions which I proposed in Spohn (1988, sect. 5) closely resembles Jeffrey conditionalization for probabilities. This solves task (b).

This very rough sketch of ranking theory may suffice for the sequel.

#### 4. SOME DIFFERENCES AND NO WAY TO BRIDGE THEM FROM THE COMPUTATIONAL SIDE

If belief revision theories start where Pollock's theory of defeasible reasoning ends, they seem to have different subject matters and thus to be independent and compatible; perhaps they might be combined. Alas, this is not the case. The theories compete, and we should get clear about how exactly they compete.

One point to observe is that the theories do not entirely diverge; they overlap in what Pollock calls ideally warranted beliefs. Pollock approaches them from below, as it were, as emerging from the closure of defeasible reasoning. Ranking theory, by contrast, takes the idealization for granted and tries to say how ideally warranted beliefs behave. At least to this extent they have the same subject matter and may, and indeed do, make different assertions about it.

The other, even more striking, point that casts doubt on combining the theories is that the theories follow entirely different methods. Pollock's theory is a decidedly computational theory. It provides a model of human computation that can actually be implemented on a computer. It is not intended as an empirical model; it is a model of how rationality is to work and hence a normative model. By contrast, ranking theory, like other theories of belief revision, is decidedly not a computational theory. This is clear already from its neglect of questions of deduction by its taking propositions and not sentences as objects of doxastic attitudes. It is rather about the dynamical structure of ideal warrant, independent of its computational accessibility. Therefore I call ranking theory a regulative theory.<sup>22</sup>

This difference in type creates a tension between the theories which shows most clearly in their normative status. Regulative theories are clearly normative, and what is going on in belief revision theory is a big abstract normative debate about the structure of ideal warrant, to use Pollock's term again. This discussion is largely formal. It derives desirable or undesirable consequences of certain assumptions, it proves completeness theorems, say, by showing that certain axioms on belief revision are equivalent to certain properties of a relation of doxastic entrenchment, and so on. But despite the formal appearance, the debate is basically normative, and thus beset with the difficulties all normative discussions have. It seeks secure grounds in intuitively compelling or commonly shared normative principles and often finds only assumptions which some find plausible and other do not; that seems unavoidable. Ranking theory is part of this debate, not driven by security, which often results in weakness, but rather by task (b) above, which calls for stronger assumptions. Whatever the merits of this debate, they are certainly not nil.

By contrast, I find purely computational theories normatively defective, they have no normative standards to appeal to, and in particular I do not see how Pollock's theory of defeasible reasoning could contribute to that normative debate about the dynamic structure of ideal warrant.

This is so because in Pollock's theory ideal warrant behaves just as determined by the computational rules; there is no independent judgment about ideal warrant. For instance, Pollock (1995, p. 140) observes that his notion of defeasible consequence (closely related to ideal warrant) satisfies the property of restricted or cumulative monotony. But why not rational monotony<sup>23</sup> (to which I adhere since its first appearance in Lewis (1972, p. 132, axiom (5)))? Why not say, whatever the computational rules, they must be so as to satisfy rational monotony? This is a perspective which, it seems, cannot be gained within Pollock's framework.



So, where do the normative issues reside in Pollock's theory? In the computational constructions, that is, in the specific inference rules and in the combination rules for integrating arguments into an inference graph. However, concerning the specific inference rules we are engaged in different issues. Of course, it is important to discuss the adequacy of, say, the statistical syllogism, but this is not a discussion about the general dynamics of belief. And concerning the combination rules, one may well ask for normative guidance.

The difficulty I am aiming at here concerns so-called admissible inference rules. These can have two statuses. In a purely computational theory, what is admissible can be judged only relative to the basic inference and the combination rules. The admissible rules state abbreviations, as it were, of several applications of basic rules. With a regulative theory in the background, by contrast, no reference to basic rules is needed. What is admissible is defined by the regulative theory.

It surely seems desirable to gain the second perspective on admissibility with respect to defeasible reasoning. The ways of induction are multifarious and do not seem easily systematized by a few basic rules. Hence, the discussion of specific rules, which is of course necessary and valuable, provides no security concerning their completeness. Similarly, direct arguments why combination rules should take this rather than that form leave us insecure; systematic arguments concerning the resulting structure would provide stronger support. For all this, an independent standard of admissibility would be most useful. And this is what is forthcoming from regulative theories, whereas I don't see how Pollock's theory could ever have a notion of admissible inference independently of his basic inference and his combination rules. This is just another way to express its normative defectiveness.

Pollock disagrees. Pollock, Cruz (1999, ch. 5) explain and defend their view that norms of rationality can only be procedural. The norms must be feasible, it must be within our power to follow them, and hence there can be no abstract norms about the resulting structure. Here is certainly our most basic difference.

It is not the place here to generally discuss the nature of norms. Let me only express my dissatisfaction by saying that Pollock's view leads, I find, to an impoverished normative life. Pollock likes to compare thinking with riding a bicycle. A certain way, which neither the pupil nor the teacher need to be able to make explicit, is the right way to ride a bike, and the same holds for thinking. But why shouldn't there be norms of the form "Build a bike!" This is a complicated matter, requiring thousands of manipulations, each governed by a procedural norm. We could not have any idea of these

many norms, and they would not make any sense, without the final aim, the overarching norm of building a bike.

Perhaps thinking is rather like building a bike. Let's consider the simplest and most basic example (though it's a bit inappropriate after I have declared my neglect of problems of deduction). Pollock would endorse the norm: "Eliminate known inconsistencies!", or rather norms for that elimination procedure. He would also endorse the norm: "Check for consistency (as hard as you can?)!", or rather procedural norms for this check. However, I find the primary norm is: "Avoid inconsistencies!", which is well motivated and justifies all the procedural norms, even if it itself is not procedural.<sup>24</sup>

My general concern concretizes in a particular concern about Pollock's combination rules. Pollock provides his inferences with strengths. Defeated arguments have strength 0, conclusive arguments have maximal strength, the other arguments are somewhere in between. He is not very specific about the strength of basic inference steps. E.g., he says only that the strength of the statistical syllogism monotonically depends on the relevant probability. Then he finds reason to reject probabilities as inference strengths and adopts two principles instead: that the strength of a complex argument is the minimum of the strengths of its premises and the strengths of its links (the weakest link principle), and that the strength of a set of arguments to the same conclusion is the maximum of the strengths of its members (the no-accrual-of-reasons principle).<sup>25</sup>

These principles may look plausible, and likewise the arguments Pollock adduces in their favor. But it is at best plausibility which is thus bestowed on the principles, and there is little insight into their behavior. Both points could be considerably improved by appealing to a suitable regulative level.

The situation reminds me, for instance, of the situation of early modal logic. Many modal axioms had been proposed, most of them were plausible, each of them was backed up by some argument. However, their consequences and their mutual fit were not well understood. There is no doubt that the situation massively improved with the appeal to a regulative level, i.e., the invention of modal semantics (though one may perhaps argue about its precise merits).

There is a much closer parallel. The basic criticism of early treatments of uncertainty in AI such as MYCIN and its successors was just this: that they distribute uncertainties according to some plausible and manageable rules, that implausibilities are somehow eliminated after discovery, and that all this ends in an ad hoc patchwork without any guidance.<sup>26</sup> This is the reason why Pearl (1988), for instance, met so much enthusiasm at

least among the more theoretically minded AI researchers. Pearl started from the best entrenched regulative theory we have, i.e., from probability theory, and showed how to make it computationally manageable, namely via the powerful and beautiful theory of Bayesian nets, and thus how to put it to use to the specific purposes of AI.<sup>27</sup> This is a lesson that is pertinent for us here as well.

##### 5. BUT SOME WAYS TO BRIDGE THE DIFFERENCES FROM THE REGULATIVE SIDE

Does the reverse balance do any better? Yes, I think it does.

A general point is that ranking functions are not computationally inaccessible. On the contrary, the theory of Bayesian nets applies *in full depth* just as well to ranking functions, due to the above-mentioned fact that conditional independence shows the same behavior with respect to probabilities as with respect to ranks. This is a benefit as big as briefly stated.

However, the bridge to the computational side should lead specifically into Pollock's field. It is obscure how it could do this. Pollock builds a rich and detailed theory of arguments, inferences, or reasons; that's its form and substance. But these terms are not even used at what I call the regulative level of belief revision theories. How could they regiment notions they don't use?

This is indeed the point where the main bulk of belief revision theory lost contact to philosophical epistemology (which Pollock always maintained). The objection does, however, not apply to ranking theory. Perhaps the notions of argument or inference cannot be abstracted from a computational framework. But since Spohn (1983) I am defending a non-computational notion of being a reason: *A* is a *reason* for *B* (or supports, confirms, speaks for *B*) iff *A* strengthens the belief in *B*, i.e., iff the degree of belief in *B* is higher given *A* than given *non-A* (or given nothing, this comes to the same), i.e., iff *A* is positively relevant for *B*. The latter phrase is immediately definable within probability theory,<sup>28</sup> and within ranking theory as well.<sup>29</sup> In Spohn (2001b) I try to argue that this explication is more adequate to the epistemological needs than the other kinds of explications in the field. For the present purpose, however, this argument need not be repeated or defended; it suffices to acknowledge that this explication is at least not less plausible than others.

On this basis, that's now the crucial point, the structure of conditionalization determines the structure of doxastic change as well as the structure of reasons. In this perspective, reasons not only serve to drive present infer-

ences to present ideally warranted beliefs, they also drive doxastic change. In this perspective, hence, belief revision does not start where defeasible reasoning ends, they are rather two sides of the same coin. And this builds my bridge into Pollock's field.

The structure of reasons thus understood closely resembles the structure of Pollock's defeasible reasoning. There are conclusive or deductive reasons that can never be defeated; indeed, they are the only undefeasible reasons. Other reasons are defeasible and they can be defeated in Pollock's two ways. If  $A$  is positively relevant and thus a reason for  $B$ , there can nevertheless be a  $C$  which is negatively relevant for  $B$  and thus a reason for  $non-B$ . Then  $C$  is a rebutting defeater. And if  $A$  is positively relevant for  $B$ , there can be a  $C$  such that given  $C$   $A$  is no longer positively relevant for  $B$ . Then  $C$  is an undercutting defeater (which may in turn be undercut by  $D$  when  $A$  is positively relevant for  $B$  given  $C$ -and- $D$ , and so on). Moreover, reasons have strengths: if  $A$  is a reason for  $B$ , one can define its strength, for instance, as the degree of belief in  $B$  given  $A$  minus the unconditional degree of belief in  $B$  (sense 1), or, alternatively, minus the degree of belief in  $B$  given  $non-A$  (sense 2).

There is no structural identity, far from that. But I think there is sufficient structural similarity for saying that Pollock and I are talking about the same subject matter, i.e., about the structure of defeasible reasoning, and that we have an argument, insofar as our structures differ.

This they do. One difference concerns the weakest link principle. It has two components that should be distinguished. On the one hand, it refers to the strength of the premises. Indeed, in deductive reasoning this is the only issue, and then it says: "The degree of support of the conclusion is the minimum of the degrees of support of its premises" (Pollock 1995, p. 99). This corresponds to the law of disjunction in ranking theory, which thus escapes the argument put forward by Pollock against probabilistic or similar interpretations of the degrees of support. Insofar we agree.

On the other hand, however, the weakest link principle defines how degrees of support propagate through chains of reasoning. Pollock assumes that inferences are transitive. If one has arrived at a certain conclusion, even with less than maximal support, he thinks one can proceed from this conclusion neglecting how one has arrived at it. I believe this is a misunderstanding of the nature of defeasible reasoning. According to my notion, reasons need neither be transitive, nor do chains of reasons conform to the weakest link principle. If degrees of belief are taken probabilistically, the strength of a chain of reasons (in sense 2) is simply the product of the strengths of the individual links (in sense 2), provided each link is independent of the previous ones. Without this independence the relation is

more complicated. If degrees of belief are ranks, the relation is again more complicated, even given the independence. Still, the laws for the strength of a chain of reasons are not ad hoc, but determined by the well justified properties of ranking functions.

A similar remark applies to Pollock's no-accrual-of-reasons principle. The principle is right in denying that the support a conclusion receives jointly from two arguments is not simply the sum of the supports it receives from the individual arguments. But it makes a strong assumption instead, namely that the joint support is the maximum of the individual supports. This violates intuition as well as my theory.

Let us look at the case of two agreeing witnesses discussed by Pollock (1995, pp. 101f.). Certainly, that *a* asserts *p* is a reason for *p*, and that *b* asserts *p* is so as well. And then, that both *a* and *b* assert *p* is usually, though not necessarily, all the more reason. I would say then that there are three reasons; two individual reasons and a joint one, which is often, but not necessarily stronger, and the strength of which depends on the strength of the one individual reason and the strength of the other reason given the first (which may differ from the unconditional strength of the other reason). Indeed, in probabilistic as well as in ranking terms, the strength of the joint reason (in sense 1) is simply the sum of these two strengths (in sense 1), a formula which is also intuitively appealing.

What could Pollock say? He might say that there is actually only one reason, namely the joint one, and that the individual witnesses do not give separate reasons. But this would amount to saying that two arguments are separate in the sense required by the no-accrual-of-reasons principle only in case it is satisfied; the principle would thus be rendered vacuous. He cannot say that there are only the two individual reasons, because then the principle would clearly be falsified. But he might say, as I did and he tends to do as well, that there are three reasons, that the joint reason is the strongest one, and that the principle is thus satisfied. But no. If the joint reason is weaker than one of the individual ones, the principle would be falsified again, since we would listen then only to the more reliable witness and neglect the other witness and their interference that somehow lowers their joint force. Moreover, we would still like to know how the strength of the joint reason depends on the individual reasons, a question which could be answered as above, but is not answered by Pollock.

So, one point is that we have specific arguments concerning specific features of reasons, arguments that I have only touched upon and not at all carried through. But the more important point, which is in line with my criticism in the previous section, is that there is a theory and an explication to guide my claims about reasons, whereas Pollock can adduce only intuit-

ive support for his claims. This is precisely the difference between having a purely computational theory and having a regulative theory.

However, Pollock does not only make structural claims about reasons with which ranking functions can compete. By proposing specific inference rules, he also makes specific claims about what is a reason for what. This is highly important, and it is regrettable that nothing of this kind can be found in the belief revision literature. It is equally important, say, as Carnap's attempt to strengthen the subjective interpretation of probability theory by further axioms for a priori probabilities such as the positive relevance axiom (saying that one case being so-and-so raises the probability for the next case being so-and-so again).

Of course, every proposal in that direction is debatable, and the grounds on which Pollock makes his proposals may not be fully clear. From the design stance his primary goal rather seems to be to get a specific inductive system going which must thus be equipped with the most plausible and most basic inference rules he can think of. But it is obvious that some things of this kind can be said, that as many as possible should be said, and that this is not a task to be left to the cognitive scientist, but a genuinely philosophical task to be dealt with from a normative or an a priori perspective. And even though these issues are poorly discussed within belief revision theory, there is rich literature in the relevant fields, i.e., in perception theory, on statistical inference, etc., on which Pollock draws and to which he has contributed.

However, these merits can be duplicated within ranking theory. Being dissatisfied with belief revision theory in this respect, I indeed started doing so.<sup>30</sup> The duplication cannot take the form of inference rules, it must be rather stated in terms of the reason relation introduced above. Pollock's inference rules thus turn into a priori constraints on doxastic states.

For instance, in Spohn (1997/98) I have argued for the following constraint which I call the Schein-Sein principle: "Given that person  $x$  observes at time  $t$  the situation in front of  $x$ , the fact that it looks to  $x$  at  $t$  as if  $p$  is an a priori reason for person  $y$  to assume that  $p$  (and vice versa)". And I emphasized there that the a priori is a defeasible one; there may be information changing the relation between  $p$  and "it looks to  $x$  at  $t$  as if  $p$ " (by conditionalization). Still, each doxastic subject  $y$  has to start from the relation as stated. All this is very similar to what Pollock has long since said.

Of course, there are nice differences again. The various readings of "look" make any claims of this kind are very delicate, and one must be very careful in which reading to make such claims, a fact Pollock is fully aware of. According to ranking theory, the reason relation is always symmetric,

i.e., defeasible support always works both ways. This seems to me to be a relevant observation even in relation to the self-application of the above constraint (in which  $x = y$ ). By contrast, I do not see how the conversion of Pollock's inference rule can be established within his system. Again, we both believe that such a priori constraints have a conceptual nature. Pollock does so because he generally adheres to a conceptual role semantics in which concepts are defined by their conceptual or inferential role. It is not fully clear to me, though, on the nature of which concepts Pollock bases his perceptual inference and the statistical syllogism. By contrast, I have a specific reason for the conceptual origin of the Schein–Sein principle,<sup>31</sup> while being doubtful about conceptual role semantics in general.<sup>32</sup> Now, however, my discussion is about to enter deep and general issues that are far beyond the scope of this paper.

Let me rather sum up. Our agreement on the philosophical importance of formal epistemology in general and defeasible reasoning in particular is overwhelming, and the agreement extends to many details. But I see deficiencies in the normative condition of Pollock's theory and have an alternative to propose from which quite a number of differences concerning details of defeasible reasoning ensue. There is a lot of substance for a continuation of the discussion.

#### ACKNOWLEDGEMENTS

I am grateful to John Pollock for a correspondence which helped me to adjust the comparison and to sharpen our differences and to Ludwig Fahrback for additional useful remarks.

#### NOTES

<sup>1</sup> Historical fairness would demand that I mention quite different strands like formal learning theory (see, e.g., Kelly 1996, and Kelly 1999 for its relation to ranking functions), and make finer distinctions. E.g., it would be inappropriate to count Rescher's hypothetical reasoning (1964) merely as a predecessor and Levi's embracive theorizing as manifested, say, in Levi (1991, 1996) simply as part of AGM belief revision theory. Shackle's functions of potential surprise should be mentioned as an early predecessor of my ranking functions (cf., e.g., Shackle 1961). And so forth. And, of course, the earlier centuries have produced rich, though inconclusive attempts at giving non-probabilistic accounts of induction.

<sup>2</sup> I shall refer only to Pollock (1990, 1995, 1998), Pollock, Cruz (1999), and Pollock, Gillies (2001).

<sup>3</sup> Cf., e.g., Gärdenfors (1988) and Rott (2001).

<sup>4</sup> For instance, AGM theorizing strictly avoids, it appears, using such crucial epistemological terms as "reason" or "justification".

<sup>5</sup> Cf., e.g., Spohn (1983, sect. 5.3, 1988, and 2001a).

<sup>6</sup> See Gärdenfors (1988, sect. 3.7) and in particular Spohn (1999). The relation is roughly the following: It is well known in AGM theory that a given entrenchment relation uniquely determines a specific behavior of single contractions (or revisions), and vice versa. In Spohn (1999) I prove that a given ranking function uniquely determines a specific behavior of iterated contractions and that the reverse determination is unique up to a multiplicative constant. This result has been earlier and independently obtained, though not published, by Matthias Hild.

<sup>7</sup> Cf. Pollock (1995, ch. 1 and sect. 3.1–3).

<sup>8</sup> Cf. Pollock, Cruz (1999, pp. 38ff.) and Pollock (1995, sect. 2.2).

<sup>9</sup> The fullest collection of inference rules may be found in Pollock (1990). See in particular pp. 339–344.

<sup>10</sup> Cf. Pollock (1995, sect. 3.3–8).

<sup>11</sup> Cf. Pollock (1995, sect. 3.6–9).

<sup>12</sup> Cf. Pollock (1995, sect. 3.10–11).

<sup>13</sup> Cf. Pollock, Gillies (2001, sect. 4).

<sup>14</sup> This picture is more explicitly worked out in the theory of Bayesian nets; cf. Pearl (1988). Roughly, one might present a complete doxastic state by a huge Bayesian net the input nodes of which would be perceptions. But one could as well consider only parts of that huge net and their boundaries. And for such partial nets the theory works just as well. The whole issue also reminds me of an old debate between Jeffrey (1965, ch.11) and Levi (1967), where Levi argued that one can dispense with generalized probabilistic conditionalization as conceived by Jeffrey, because each such conditionalization is ultimately based on a simple conditionalization with respect to some perceptual or phenomenal proposition, and where Jeffrey could reply that generalized conditionalization is valuable precisely because it need not rely on such further reductive claims. Not surprisingly, my sympathies are here with Jeffrey.

<sup>15</sup> I agree, however, with the criticism that the input for belief revision does not simply consist in propositions; for this purpose, pieces of information should rather be considered as propositions furnished with some input strength.

<sup>16</sup> This includes the suspicion that all the constructive attempts towards the deduction problem, as it is also called, are not very satisfying.

<sup>17</sup> Cf. Spohn (1988, sect. 2–3).

<sup>18</sup> Cf., e.g., Nayak (1994).

<sup>19</sup> There, I still called them ordinal conditional functions. “Ranking functions” is *much* nicer, as Goldszmidt, Pearl (1992) have convinced me.

<sup>20</sup> Cf. Spohn (1994).

<sup>21</sup> Cf. Jeffrey (1965, ch. 11).

<sup>22</sup> At earlier places I spoke of the distinction between syntactic notions like proof, provability etc. and semantic notions like logical truth, valid inference, etc. The present distinction between computational and regulative theories is obviously analogous. The present terms are more appropriate since it would be mystifying to call ranking theory, or probability theory, for that matter, semantic theories.

<sup>23</sup> Deductive logic always satisfies strengthening of the antecedent or monotony: if one can infer  $C$  from  $A$ , one can infer  $C$  also from  $A$ -and- $B$ . In defeasible reasoning this holds only restrictedly. Cumulative monotony requires as an additional assumption that one can



also (defeasibly) infer  $B$  from  $A$ , whereas rational monotony only requires that one cannot (defeasibly) infer non- $B$  from  $A$ .

<sup>24</sup> In (1998), however, Pollock is less strict. There, on p. 392, he speaks of the necessity of combining top-down epistemological theorizing (such as foundationalism, coherentism, or probabilism) and bottom-up theorizing (which insists on implementation) and of the necessity of fitting low-level theories into high-level structural theories of epistemic justification. But even there, theories of inductive reasoning are counted among the low-level theories. I find a lot of evidence for the fact that the top-down strategy is legitimate and most fruitful also with respect to inductive reasoning.

<sup>25</sup> Cf. Pollock (1995, sect. 3.4).

<sup>26</sup> Cf. Pearl (1988, sect. 1.2).

<sup>27</sup> See also the well accessible textbook by Jensen (1996).

<sup>28</sup> Indeed, positive relevance was already used by Carnap to explicate confirmation; cf. Carnap (1950, ch. VI) and in particular the preface to the second edition, pp. xv–xx.

<sup>29</sup> One of my basic criticisms of AGM belief revision theory is that it cannot provide such a notion of being a reason; cf. Spohn (1999, sect. 1).

<sup>30</sup> If I had earlier read the works of Pollock, I might have had less work.

<sup>31</sup> The reason is that I approached the Schein–Sein principle by thinking not about perception and skepticism, but rather about dispositions (cf. Spohn 1997/98). There I concluded that reduction sentences really take the form: given that  $x$  is put into water, the assumption that  $x$  dissolves is a defeasibly a priori reason for the assumption that  $x$  is soluble, and vice versa. Hence, the defeasible a priori is here embedded into the dispositional concept. Now carry this over to secondary qualities and generalize, and thus you arrive at the Schein–Sein principle.

<sup>32</sup> For me, concepts are intensions, or rather diagonal (or A- or primary) intensions (cf. Haas-Spohn, Spohn 2001), and it is still a deep problem how this idea, which is overall the dominating one, relates to inferential role semantics.

## REFERENCES

- Carnap, R.: 1950, *Logical Foundations of Probability*, Chicago University Press, Chicago, 2nd edn (1962).
- Gärdenfors, P.: 1988, *Knowledge in Flux*, MIT Press, Cambridge, MA.
- Goldszmidt, M. and J. Pearl (1992), 'Rank-Based Systems: A Simple Approach to Belief Revision, Belief Update, and Reasoning About Evidence and Actions', in *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA.
- Haas-Spohn, U. and W. Spohn: 2001, 'Concepts are Beliefs about Essences', in A. Newen, U. Nortmann, and R. Stuhmann-Laeisz (eds), *Building on Frege. New Essays on Sense, Content, and Concept*, CSLI Publications, Stanford, pp. 287–316.
- Harper, W. L.: 1976, 'Rational Belief Change, Popper Functions, and Counterfactuals', in W. L. Harper and C. A. Hooker (eds), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. I, Reidel, Dordrecht, pp. 73–115.
- Hintikka, J.: 1962, *Knowledge and Belief*, Cornell University Press, Ithaca, NY.
- Jeffrey, R. C.: 1965, *The Logic of Decision*, The University of Chicago Press, Chicago, 2nd edn (1983).
- Jensen, F. V.: 1996, *An Introduction to Bayesian Networks*, UCL Press, London.

- Kelly, K.: 1996, *The Logic of Reliable Inquiry*, Oxford University Press, Oxford.
- Kelly, K.: 1999, 'Iterated Belief Revision, Reliability, and Inductive Amnesia', *Erkenntnis* **50**, 11–58.
- Levi, I.: 1967, 'Probability Kinematics', *British Journal for the Philosophy of Science* **18**, 197–209.
- Levi, I.: 1991, *The Fixation of Belief and Its Undoing*, Cambridge University Press, Cambridge.
- Levi, I.: 1996, *For the Sake of Argument*, Cambridge University Press, Cambridge.
- Nayak, A. C.: 1994, 'Iterated Belief Change Based on Epistemic Entrenchment', *Erkenntnis* **41**, 353–390.
- Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA.
- Pollock, J. L.: 1990, *Nomic Probability and the Foundations of Induction*, Oxford University Press, Oxford.
- Pollock, J. L.: 1995, *Cognitive Carpentry*, MIT Press, Cambridge, MA.
- Pollock, J. L.: 1998, 'Procedural Epistemology – At the Interface of Philosophy and AI', in J. Greco and E. Sosa (eds), *The Blackwell Guide to Epistemology*, Blackwell, Oxford, pp. 383–414.
- Pollock, J. L. and J. Cruz: 1999, *Contemporary Theories of Knowledge*, 2nd edn, Rowman & Littlefield, Lanham, MD.
- Pollock, J. L. and A. S. Gillies: 2001, 'Belief Revision and Epistemology', *Synthese* (forthcoming).
- Rescher, N.: 1964, *Hypothetical Reasoning*, North-Holland, Amsterdam.
- Rott, H.: 2001, *Change, Choice, and Inference*, University Press, Oxford.
- Shackle, G. L. S.: 1961, *Decision, Order, and Time in Human Affairs*, Cambridge University Press, Cambridge, 2nd edn, 1969.
- Spohn, W.: 1983, *Eine Theorie der Kausalität*, unpublished Habilitationsschrift, Munich.
- Spohn, W.: 1988, 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in W. L. Harper and B. Skyrms (eds), *Causation in Decision, Belief Change, and Statistics*, Vol. II, Kluwer, Dordrecht, pp. 105–134.
- Spohn, W.: 1994, 'On the Properties of Conditional Independence', in P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher, Vol. 1: Probability and Probabilistic Causality*, Kluwer, Dordrecht, 1994, pp. 173–194.
- Spohn, W.: 1997/98, 'How to Understand the Foundations of Empirical Belief in a Coherentist Way', *Proceedings of the Aristotelian Society, New Series* **98**, 23–40.
- Spohn, W.: 1999, 'Ranking Functions, AGM Style', *Forschungsberichte der DFG-Forschergruppe Logik in der Philosophie*, No. 28, also in B. Hansson, S. Halldán, N.-E. Sahlin, and W. Rabinowicz (eds), *Internet-Festschrift for Peter Gärdenfors*, Lund, s.: <http://www.lu.se/spinning/>
- Spohn, W.: 2001a, 'Deterministic Causation', in W. Spohn, M. Ledwig, and M. Esfeld, (eds), *Current Issues in Causation*, Mentis, Paderborn, pp. 21–46.
- Spohn, W.: 2001b, 'Vier Begründungsbegriffe', in T. Grundmann (ed.), *Challenges to Traditional Epistemology*, Mentis, Paderborn, pp. 33–52.

Fachbereich Philosophie  
 Universität Konstanz  
 78457 Konstanz  
 Germany



