This is a revised version of a paper of the same name appearing in *Artificial Intelligence.*
version of 10/2/01

Defeasible Reasoning with Variable Degrees of Justification

John L. Pollock
Department of Philosophy
University of Arizona
Tucson, Arizona 85721
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

Abstract

The question addressed in this paper is how the degree of justification of a belief is determined. A conclusion may be supported by several different arguments, the arguments typically being defeasible, and there may also be arguments of varying strengths for defeaters for some of the supporting arguments. What is sought is a way of computing the "on sum" degree of justification of a conclusion in terms of the degrees of justification of all relevant premises and the strengths of all relevant reasons.

I have in the past defended various principles pertaining to this problem. In this paper I reaffirm some of those principles but propose a significantly different final analysis. Specifically, I endorse the weakest link principle for the computation of argument strengths. According to this principle the degree of justification an argument confers on its conclusion in the absence of other relevant arguments is the minimum of the degrees of justification of its premises and the strengths of the reasons employed in the argument. I reaffirm my earlier rejection of the accrual of reasons, according to which two arguments for a conclusion can result in a higher degree of justification than either argument by itself. This paper diverges from my earlier theory mainly in its treatment of defeaters. First, it argues that defeaters that are too weak to defeat an inference outright may still diminish the strength of the conclusion. Second, in the past I have also denied that multiple defeaters can result in the defeat of an argument that is not defeated by any of the defeaters individually. In this paper I urge that there are compelling examples that support a limited version of this "collaborative" defeat.

The need to accommodate diminishers and collaborative defeat has important consequences for the computation of degrees of justification. The paper proposes a characterization of degrees of justification that captures the various principles endorsed and constructs an algorithm for computing them.

Keywords: defeasible, defeat, justification, reasoning, nonmonotonic

---

This work was supported by NSF grants nos. IRI-9634106 and IRI-IIS-0080888.

1. Introduction

I have argued at length elsewhere that a rational agent operating in a complex environment must reason about its environment defeasibly.[1] For example, perceptual input is not always accurate, so an agent forming beliefs about its environment on the basis of its sensors must be prepared to withdraw such beliefs in the face of further information. A sophisticated agent should be able to discover generalizations about its environment by reasoning inductively, but inductive reasoning is defeasible—new evidence may overturn earlier generalizations. Because perception only enables an agent to monitor small parts of its environment at any one time, in order to build a coherent world model the agent must combine conclusions drawn on the basis of different perceptual experiences occurring at different times. But that requires a defeasible assumption that the world does not change too rapidly, so that what was perceived a moment ago is still true. The ability to maneuver through a rich environment requires an agent to be able to reason about the causal consequences of both its own actions and other events that it observes. That requires a solution to the frame problem, and it is generally agreed that any such solution will require defeasible reasoning.[2] I have also argued that planning with incomplete information requires a defeasible assumption that different plans do not destructively interfere with each other.[3] Although most of these "theoretical" observations are fairly obvious, they have not had much impact on the actual practice of AI, because for the most part people have not tried to build autonomous agents of sufficient sophistication to encounter these problems. However, that is changing and we are getting close to the point where we can construct practical agents that will not work satisfactorily unless their designers take these observations to heart.

The OSCAR architecture for rational agents is based upon a general theory of defeasible reasoning. OSCAR implements the system of defeasible reasoning described in Pollock (1995).[4] That system in turn derives from thirty years of theorizing in philosophical epistemology. The basic idea is that the agent constructs arguments using both deductive and defeasible reason-schemes (inference-schemes). The conclusions of some of these arguments may constitute defeaters for steps of some of the other arguments. Given the set of interacting arguments that represent the agent's epistemological state at a given time, an algorithm is run to compute degrees of justification, determining which arguments are undefeated and the level of support they provide their conclusions. What the agent should believe at any particular time are the conclusions of the undefeated arguments. The hard part of a theory of defeasible reasoning is to give an account of which arguments are undefeated. This is a topic I have addressed before (my 1987, 1994, 1995), but some of the considerations adduced later in this paper have convinced me of the need to revisit it.

My analysis will turn on the taxonomy of defeaters that I introduced in my (1970) and (1974) and that has been endorsed by most subsequent work on defeasibly reasoning (see Prakken and

---

[1] The argument spans three decades. My most recent papers in this vein is Pollock (1998) and (1998b), but see also Pollock (1974), (1987), (1990), (1995), and Pollock and Cruz (1999).

[2] I have proposed and implemented a solution to the frame problem in my (1998).

[3] A system of planning based upon this observation was described in my (1998b), and has been implemented in OSCAR.

[4] For a presentation of OSCAR and its current capabilities, see the OSCAR web page at http://oscarhome.soc-sci..edu/ftp/OSCAR-web-page/OSCAR.htm.

Vreeswijk 2002 and Chesñevar, Maguitman, and Loui 2000). According to this taxonomy, there are two importantly different kinds of defeaters. Where P is a defeasible reason for Q, R is a *rebutting defeater* iff R is a reason for denying Q. All work on nonmonotonic logic and defeasible reasoning has recognized the existence of rebutting defeaters, but there are other defeaters as well. For instance, suppose x looks red to me, but I know that x is illuminated by red lights and red lights can make objects look red when they are not. Knowing this defeats the defeasible reason, but it is not a reason for thinking that x is *not* red. After all, red objects look red in red light too. This is an *undercutting defeater*. Undercutting defeaters attack the *connection* between the reason and the conclusion rather than attacking the conclusion directly. For example, an undercutting defeater for the inference from x's looking red to x's being red attacks the connection between "x looks red to me" and "x is red", giving us a reason for doubting that x wouldn't look red unless it were red. I will symbolize the negation of "P wouldn't be true unless Q were true" as "$P \otimes Q$". A shorthand reading is "P does not guarantee Q". If Γ (a set of propositions) is a defeasible reason for P, then where $\Pi\Gamma$ is the conjunction of the members of Γ, any reason for believing "$\Pi\Gamma \otimes P$" is a defeater. Thus I propose to characterize undercutting defeaters as follows:

> If Γ is a defeasible reason for P, an *undercutting defeater* for Γ as a defeasible reason for P is any reason for believing "$(\Pi\Gamma \otimes P)$".

Are there any defeaters other than rebutting and undercutting defeaters? A number of authors have advocated *specificity defeaters* (e.g., Touretzky 1984, Poole 1988, Simari and Loui 1992). These have been formulated differently by different authors, but the general idea is that if two arguments lead to conflicting conclusions but one argument is based upon more information than the other then the "more informed" argument defeats the "less informed" one.

A phenomenon like this is common in several different kinds of probabilistic reasoning. To illustrate, consider the statistical syllogism. The statistical syllogism can be formulated as follows (see my 1990):

(SS) If $r > 0.5$, then "Fc & $\text{prob}(G/F) \geq r$" is a defeasible reason for believing "Gc", the strength of the reason depending upon the value of r.

When reasoning in accordance with (SS), there is a kind of "total evidence requirement" according to which we should make our inference on the basis of the most comprehensive facts regarding which we know the requisite probabilities. This can be accommodated by endorsing the following undercutting defeater for (SS):

> "Hc & $\text{prob}(G/F\&H) \neq \text{prob}(G/F)$" is an undercutting defeater for (SS).

I refer to these as *subproperty defeaters.*[5]

Early work in AI on defeasible reasoning tended to concentrate on examples that were

[5] I first pointed out the need for subproperty defeaters in my (1983). Touretzky (1984) subsequently introduced similar defeaters for use in defeasible inheritance hierarchies.

instances of the statistical syllogism (e.g., the venerable "Tweety" arguments), and this led people to suppose that something like subproperty defeat was operative throughout defeasible reasoning. However, I do not see any reason to believe that. There are several specific kinds of defeasible reasoning that are subject to total-evidence requirements. The statistical syllogism is one. Direct inference (discussed in section eight below) is another. Don Nute has pointed out to me that various kinds of legal reasoning and deontic reasoning are subject to a similar requirement. These can all be accommodated by acknowledging similar subproperty defeaters (which are undercutting defeaters) for the defeasible inference-schemes involved in the reasoning. To the best of my knowledge, there has never been an intuitive example of specificity defeat presented anywhere in the literature that is not an example of the operation of the total-evidence requirement in one of these special varieties of defeasible inference, and the latter are all instances of undercutting defeat. Accordingly, I will assume that undercutting defeaters and rebutting defeaters are the only possible kinds of defeaters.

I have defended the preceding remarks at length in numerous publications over the past 30 years, so for the purposes of this paper I will regard them as ancient history and take them for granted without further discussion. They are not the topic of this paper. This paper is about how to compute defeat statuses. The literature on defeasible and nonmonotonic reasoning contains numerous proposals for how to do this.[6] The current version of OSCAR computes defeat statuses in the manner described in my (1994) and (1995).[7] If we ignore the fact that some arguments provide stronger support for their conclusions than other arguments, we can describe OSCAR's defeat status computation as follows. 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. These links relate their *roots* to their *targets*. The root of a defeat-link is a single node, and the root of a support-link is a set of nodes. 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. For example, consider the inference-graph diagrammed in figure one, which represents two different arguments for (*P*&*Q*) given the premises, *P*, *Q*, *A*, and (*Q* → (*P*&*Q*)). The nodes of such an inference-graph represent arguments rather than just representing their conclusions. In such an inference-graph, a node has at most one support-link. When it is unambiguous to do so, I will refer to the nodes in terms of the conclusions they encode.

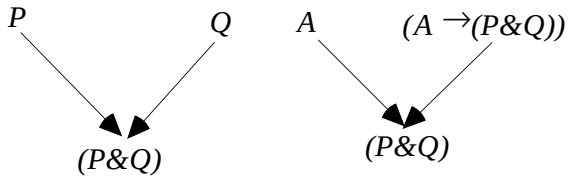


Figure 1. An inference-graph

[6] Two good surveys are Prakken and Vreeswijk (2002) and Chesñevar, Maguitman, and Loui (2000).

[7] For comparison with default logic and circumscription, see my (1995), chapter three. For comparison with more recent systems of defeasible argumentation, see Prakken and Vreeswijk (2002).

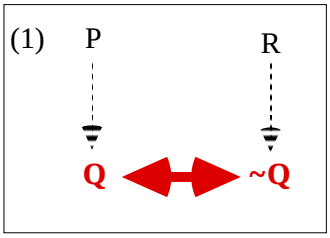
The *node-basis* of a node is the root of its support-link (if it has one), i.e., the set of nodes from which the node is inferred in a single step. If a node has no support-link (i.e., it is a premise) then the node-basis is empty. The *node-defeaters* are the roots of the defeat-links having the node as their target. We define:

A node of the inference-graph is *initial* iff its node-basis and list of node-defeaters is empty.

The *defeat status* of a node is either "defeated" or "undefeated". It is initially tempting to try to characterize defeat statuses recursively using the following three rules:

(1) Initial nodes are undefeated.
(2) A node is undefeated if all the members of its node-basis are undefeated and all node-defeaters are defeated.
(3) A node is defeated if either some member of its node-basis is defeated or some node-defeater is undefeated.

However, this recursion turns out to be ungrounded because we can have nodes of an inference-graph that defeat each other, as in inference-graph (1), where dashed arrows indicate defeasible inferences and heavy arrows indicate defeat-links. In computing defeat statuses in inference-graph (1), we cannot proceed recursively using rules (1)–(3), because that would require us to know the defeat status of *Q* before computing that of ~*Q*, and also to know the defeat status of ~*Q* before computing that of *Q*. The problem is more generally that a node *P* can have an inference/defeat-descendant that is a defeater of *P*, where an inference/defeat-descendant of a node is any node that can be reached from the first node by following support-links and defeat-links. I will say that a node is *P-dependent* iff it is an inference/defeat-descendant of a node *P*. So the recursion is blocked in inference-graph (1) by there being *Q*-dependent defeaters of *Q* and ~*Q*-dependent defeaters of ~*Q*.



Inference-graph (1) is a case of "collective defeat". For example, let *P* be "Jones says that it is raining", *R* be "Smith says that it is not raining", and *Q* be "It is raining". Given *P* and *Q*, and supposing you regard Smith and Jones as equally reliable, what should you believe about the weather? It seems clear that you should withhold belief, accepting neither. In other words, both *Q* and ~*Q* should be defeated. This constitutes a counter-example to rule (2). So not only do rules (1)–(3) not provide a recursive characterization of defeat statuses — they are not even true. The failure of these rules to provide a recursive characterization of defeat statuses suggests that no such characterization is possible, and that in turn suggested to me (in my 1994, 1995) that rules

(1)–(3) might be used to characterize defeat statuses in another way. Reiter's (1980) default logic proceeded in terms of multiple "extensions", and "skeptical default logic" characterizes a conclusion as following nonmonotonically from a set of premises and defeasible inference-schemes iff it is true in every extension. The currently popular stable model semantics (Dung 1995) is derived from this approach. There are simple examples showing that these semantics are inadequate for the general defeasible reasoning of epistemic agents (see section two), but the idea of having multiple extensions suggested to me that rules (1)–(3) might be used to characterize multiple "status assignments". On this approach, a status assignment is an assignment of defeat statuses to the nodes of the inference-graph in accordance with three simple rules:

> An assignment σ of "defeated" and "undefeated" to a subset of the nodes of an inference-graph is a *partial status assignment* iff:
> 1. σ assigns "undefeated" to any initial node;
> 2. σ assigns "undefeated" to a node α iff σ assigns "undefeated" to all the members of the node-basis of α and all node-defeaters of α are assigned "defeated"; and
> 3. σ assigns "defeated" to a node α iff either some member of the node-basis of α is assigned "defeated", or some node-defeater of α is assigned "undefeated".
>
> σ is a *status assignment* iff σ is a partial status assignment and σ is not properly contained in any other partial status assignment.

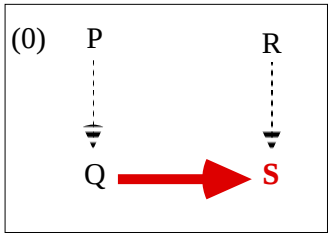
My proposal was then:

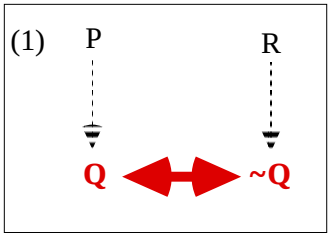
> A node is undefeated iff every status assignment assigns "undefeated" to it; otherwise it is defeated.
>
> Belief in *P* is justified for an agent iff *P* is encoded by an undefeated node of the inference-graph representing the agent's current epistemological state.

2. Examples

The ultimate test of a semantics is that it validates the intuitively right reasoning. It is human intuitions about correct reasoning that we want to capture. (See Pollock & Cruz (1999), chapter six, for further discussion of this.) With this in mind, it will be useful to have a number of examples of inference-graphs to test the analysis I will propound below. I will assume throughout this section that all initial nodes (premises) have the same degree of justification, and all reason-schemes have the same reason-strength.

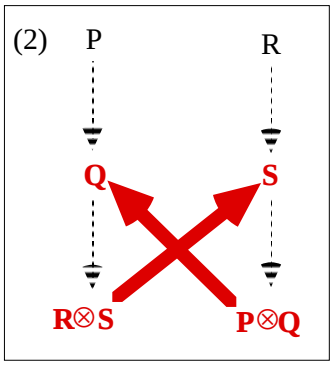
The simplest case is inference-graph (0). Presumably, any semantics for defeasible reasoning will yield the result that *S* is defeated, and *P*, *Q*, and *R* are undefeated.

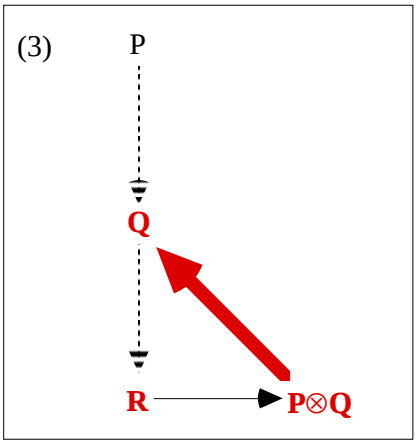




Inference-graph (1) illustrates "collective defeat". It was discussed above, and again all semantics for defeasible reasoning yield the result that *Q* and ~*Q* are defeated, but *P* and *R* are undefeated.

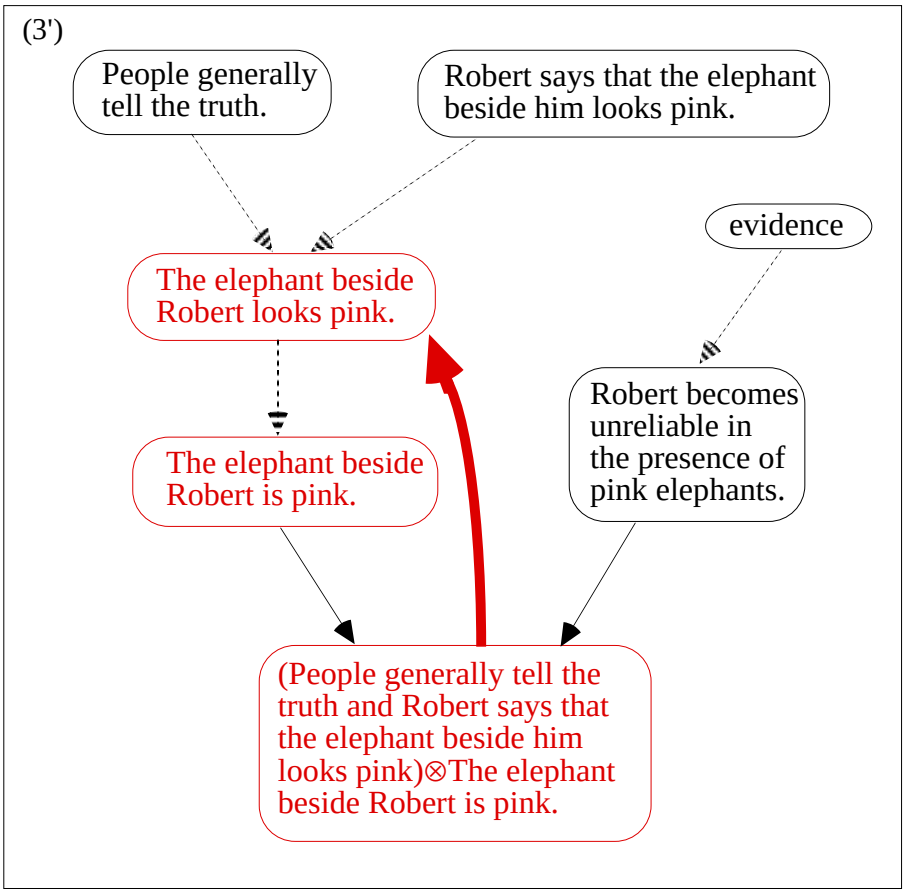
Inference-graph (2) is another instance of collective defeat, differing from inference-graph (1) in that the defeaters are undercutting defeaters rather than rebutting defeaters. The result should be that *Q*, *S*, *R*⊗*S*, and *P*⊗*Q* are defeated, and *P* and *R* are undefeated. For an example, let *P* = "Jones says Smith is untrustworthy", *R* = "Smith says Jones is untrustworthy", *Q* = "Smith is untrustworthy", *S* = "Jones is untrustworthy". The semantics of section one produces two status assignments, one in which *Q* and *R*⊗*S* are assigned "defeated" and all other nodes are assigned "undefeated", and one in which *S* and *P*⊗*Q* are assigned "defeated" and all other nodes are assigned "undefeated".

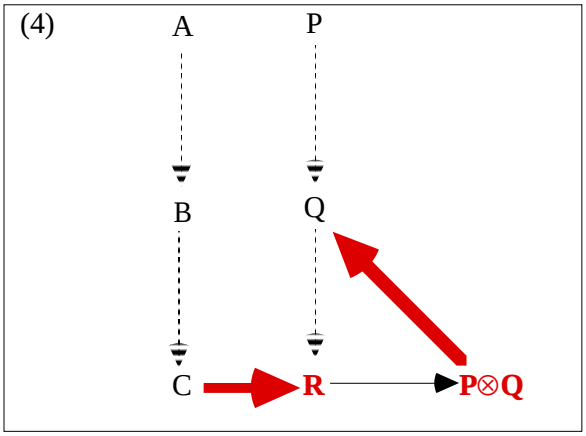




Inference-graph (3) is a simple example of a "self-defeating" argument. A partial status assignment must assign "undefeated" to *P*. If it assigned "undefeated" to *Q* then it would assign "undefeated" to *R* and *P*⊗*Q*, in which case it would have to assign "defeated" to *Q*. So it cannot assign "undefeated" to *Q*. If it assigned "defeated" to *Q* it would have to assign "defeated" to *R* and *P*⊗*Q*, in which case it would have to assign "undefeated" to *Q*. So that is not possible either. Thus a partial status assignment cannot assign anything to *Q*, *R*, and *P*⊗*Q*. Hence there is only one status assignment (i.e., maximal partial status assignment). Accordingly, *P* is undefeated and the other nodes are defeated. An intuitive example having approximately the same form is shown in inference-graph (3'). Inference-graphs (3) and (3') constitute intuitive counterexamples to default logic (Reiter 1980) and the stable model semantics (Dung 1995) because there are no extensions. Hence on those semantics, *P* has the same status as *Q*, *R*, and *P*⊗*Q*. It is perhaps more obvious that this is a problem for those semantics if we imagine this self-defeating argument being embedded in a larger inference-graph containing a number of otherwise perfectly ordinary

arguments. On these semantics, all of the nodes in all of the arguments would have to have the same status, because there would still be no extensions. But surely the presence of the self-defeating argument should not have the effect of defeating all other (unrelated) arguments.

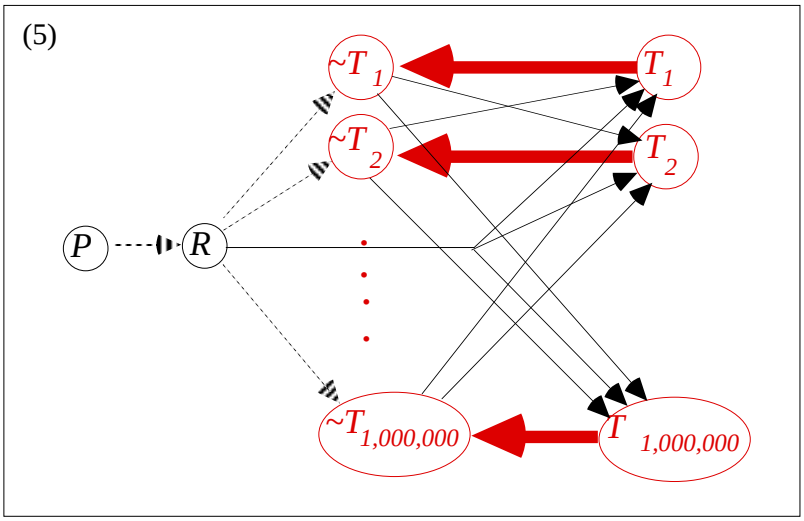




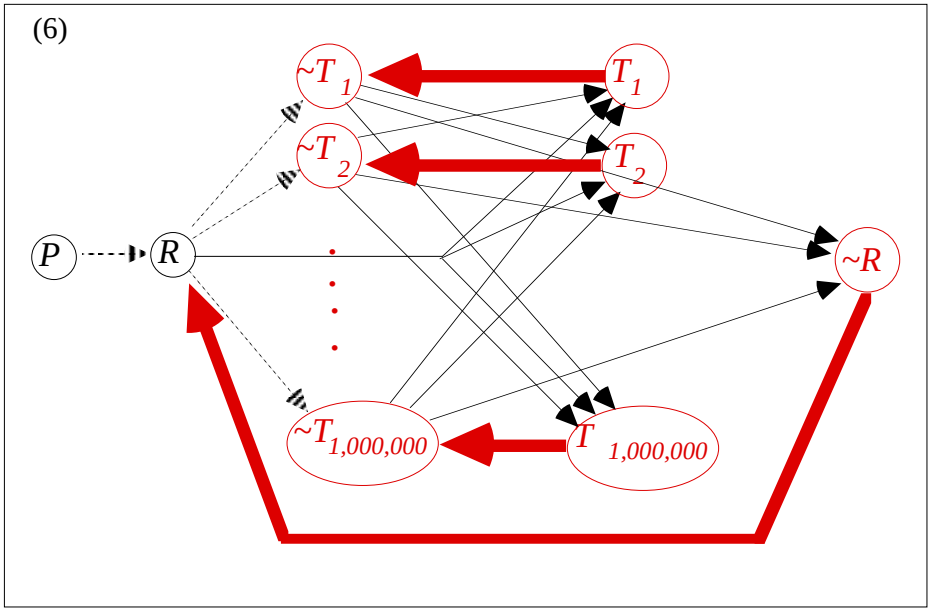
Inference-graph (4), illustrates that self-defeat can be "cancelled" by external defeat. Here *R* is defeated by *C*, so *P*⊗*Q* is defeated and *Q* is undefeated. Accordingly, there is just one status assignment, assigning "undefeated" to *A*, *B*, *C*, *P*, and *Q*, and "defeated" to *R* and *P*⊗*Q*.

Inference-graph (5) illustrates the so-called "lottery paradox" (Kyburg 1961). Here *P* reports a

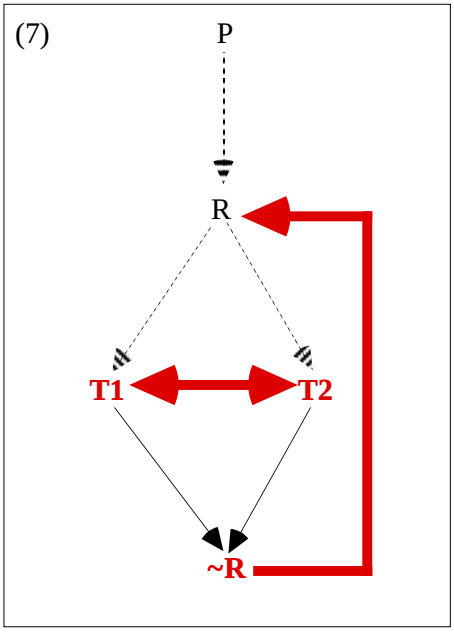
description (e.g., a newspaper report) of a fair lottery with one million tickets. *P* constitutes a defeasible reason for *R*, which is the description. In such a lottery, each ticket has a probability of one in a million of being drawn, so for each *i*, the statistical syllogism gives us a reason for believing $\sim T_i$ ("ticket *i* will not be drawn"). The supposed paradox is that although we thusly have a reason for believing of each ticket that it will not be drawn, we can also infer on the basis of *R* that some ticket will be drawn. Of course, this is not really a paradox, because the inferences are defeasible and this is a case of collective defeat. This results from the fact that for each *i*, we can infer T_i from the description *R* (which entails that some ticket will be drawn) and the conclusions that none of the other tickets will be drawn. This gives us a defeating argument for the defeasible argument to the conclusion that $\sim T_i$, as diagrammed in inference-graph (5). The result is that for each *i*, there is a status assignment on which $\sim T_i$ is assigned "defeated" and the other $\sim T_j$'s are all assigned "undefeated", and hence none of them are assigned "undefeated" in every status assignment.



I believe that all semantics for defeasible reasoning get the lottery paradox right. A more interesting example is the "lottery paradox *paradox*", diagrammed in inference-graph (6). This results from the observation that because *R* entails that some ticket will be drawn, from the collection of conclusions of the form $\sim T_i$ we can infer $\sim R$, and that is a defeater for the defeasible inference from *P* to *R*. This is another kind of self-defeating argument. Clearly, the inferences in the lottery paradox should not lead us to disbelieve the newspaper's description of the lottery, so *R* should be undefeated. Circumscription (McCarthy 1986), in its simple non-prioritized form, gets this example wrong, because one way of minimizing abnormalities would be to block the inference from *P* to *R*. My own early analysis (Pollock 1987) also gets this wrong. This was the example that led me to the analysis of section one. That analysis gets this right. We still have the same status assignments as in inference-graph (5), and $\sim R$ is defeated in all of them because it is inferred from the entire set of $\sim T_i$'s, and one of those is defeated in every status assignment.



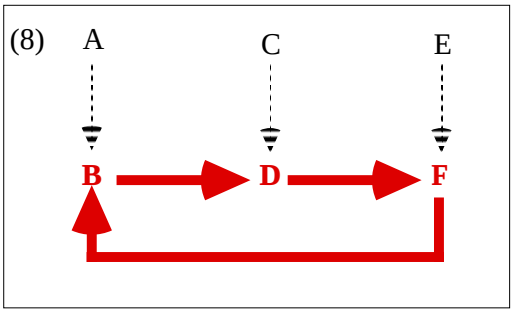
It will be convenient to have a simpler example of an inference-graph with the same general structure as the lottery paradox paradox. For that purpose we can use inference-graph (7). Here *P* and *R* should be undefeated, but *T1*, *T2*, and **~***R* should be defeated.



A final example that creates interesting problems involves "odd-length defeat cycles", as in inference-graph (8). For example, let *A* = "Jones says that Smith is unreliable", *B* = "Smith is unreliable", *C* = "Smith says that Robinson is unreliable", *D* = "Robinson is unreliable", *E* = "Robinson says that Jones is unreliable", *F* = "Jones is unreliable". Intuitively, this should be another case of collective defeat, with *A*, *C*, and *E* being undefeated and *B*, *D*, and *F* being defeated. The semantics of section one does yield this result, but it does it in a peculiar way. *A*, *C*, and *E* must be assigned "undefeated", but there is no consistent way to assign defeat statuses

of *B*, *D*, and *F*. Accordingly, there is only one status assignment (maximal partial status assignment), and it leaves *B*, *D*, and *F* unassigned. We get the right answer, but it seems puzzling that we get it in a different way than we do for even-length defeat cycles like that in inference-graph (1). This difference has always bothered me.

(8)



3. Varying Degrees of Justification

The preceding account of defeat status assumes that all arguments support their conclusions equally strongly. However, this assumption is unrealistic. For example, increasing the degrees of justification of the premises of an argument may increase the degree of justification of the conclusion, and increasing the strengths of the reasons employed in the argument may increase the degree of justification of the conclusion. This phenomenon has been ignored in most AI work on defeasible and nonmonotonic reasoning, but it is of considerable importance in applications of such reasoning. For example, in Pollock (1998) I discussed temporal projection, wherein it is assumed defeasibly that if something is true at one time it is still true at a later time. This is, in effect, a defeasible assumption that fluents are stable, tending not to change truth values as time passes. The stability of a fluent is measured by the probability ρ that if it is true at time *t* then it is still true at time *t*+1. More generally, if it is true at time *t*, then the probability of its being true at $t+\Delta t$ is $\rho^{\Delta t}$. So the strength of the defeasible expectation supported by temporal projection is a monotonic decreasing function of the time interval. This can be captured in a system of defeasible reasoning by employing a reasoning scheme of the following sort:

> "*P*-at-t" is a defeasible reason for believing "*P*-at-($t+\Delta t$)", the strength of the reason being a monotonic decreasing function of Δt.[8]

The decreasing strength is important in understanding perceptual updating, wherein on the basis of new perceptual experience the agent overrides temporal projection and concludes that the fluent has changed truth value. Perception is not infallible, so perception should provide only a defeasible reason for believing that the environment is as represented by the percept.[9] Suppose an object looks red at one time t_1 and blue at a later time t_2. The agent should assume

[8] The reason-schema proposed in Pollock (1998) involves some additional qualifications, but they are not relevant to the present discussion.

[9] This is discussed in detail in Pollock (1998).

defeasibly that the object is initially red, but should also conclude defeasibly that it changes color later and is blue at t_2. The object's being red at t_1 provides a defeasible reason for expecting it to be red at t_2, and its looking blue at t_2 provides a defeasible reason for thinking it blue and hence not red at t_2. If these reasons were of the same strength, there would be no basis for preferring one conclusion to the other and the agent would be unable to draw a justified conclusion about the color of the object at t_2. The situation is resolved by noting that the reason for thinking the object is still red at t_2 is weaker than the reason for thinking it was red at t_1, and hence weaker than the reason for thinking the object is blue (and so not red) at t_2. Because the agent has a stronger reason for thinking the object is blue at t_2 than for thinking it is red at t_2, that becomes the justified conclusion and the agent is able to conclude that the object has changed color.

The preceding example illustrates the importance of incorporating an account of degrees of justification into a system of defeasible reasoning. There are many other examples illustrating the same point. For instance, in the statistical syllogism the strength of the reason is a function of the probability employed. Autonomous agents capable of engaging in sophisticated defeasible reasoning must accommodate varying degrees of justification. The question addressed in this paper is how the degree of justification of a conclusion should be determined. A conclusion may be supported by several different arguments. The arguments are typically defeasible, and there may also be arguments of varying strengths for defeaters for some of the supporting arguments. What is sought is a way of computing the "on sum" degree of justification of a conclusion. It is clear that three variables, at least, are involved in determining degrees of justification. The reason-strengths of the reason-schemes employed in the argument are relevant. The degrees of justification of the premises are relevant. And the degrees of justification of any defeaters for defeasible steps of the argument are relevant. Other variables might be relevant as well. I am going to assume that reason-strengths and degrees of justification are measurable as non-negative extended real numbers (i.e., the non-negative reals together with ∞). The justification for this assumption will be provided in section nine.

4. Argument-Strengths

For the sake of completeness, this section and the next repeat arguments given in my (1995). Let us begin by looking at arguments for which we have no arguments supporting defeaters. Let the *strength of an argument* be the degree of justification it would confer on its conclusion under those circumstances. A common and seductive view would have it that argument strength can be modeled by the probability calculus. On this view, the strength a conclusion gains from the premises can be computed in accordance with the probability calculus from the strength of the reason (a conditional probability) and the probabilities of the premises. I, and many other authors, have argued against this view at length, but it has a remarkable ability to keep attracting new converts.

There are a number of familiar arguments against the probabilistic model. The simplest argument proceeds by observing that the probabilistic model would make it impossible to be justified in believing a conclusion on the basis of a deductive argument from numerous uncertain premises. This is because as you conjoin premises, if degrees of support work like probabilities, the degree of support decreases. Suppose you have 100 independent premises, each highly

probable, having, say, probability .99. According to the probability calculus, the probability of the conjunction will be only .37, so we could never be justified in using these 100 premises conjointly in drawing a conclusion. But this flies in the face of human practice. For example, an engineer building a bridge will not hesitate to make use of one hundred independent measurements to compute (deduce) the correct size for a girder. I have discussed this issue at length elsewhere (Pollock 1987, 1995, Pollock and Cruz 1999), so in this paper I am just going to assume that deductive arguments provide one way of arriving at new justified conclusions on the basis of earlier ones. A corollary is that the probabilistic model is wrong.[10]

If deductive reasoning automatically carries us from justified conclusions to justified conclusions, then the degree of support a deductive argument confers on its conclusion cannot decrease as the number of premises increases. The degree of justification for the conclusion must be no less than that of the most weakly justified premise. This is the *Weakest Link Principle for Deductive Arguments*, according to which a deductive argument is as good as its weakest link. More precisely:

> The argument strength of a deductive argument is the minimum of the degrees of justification of its premises.

This formulation of the weakest link principle applies only to deductive arguments, but we can use it to obtain an analogous principle for defeasible arguments. If P is a defeasible reason for Q, then we can use conditionalization to construct a simple defeasible argument for the conclusion $(P \rightarrow Q)$, and this argument turns upon no premises:

Suppose P
Then (defeasibly) Q.

Therefore, $(P \rightarrow Q)$.

As this argument has no premises, the degree of support of its conclusion should be a function of nothing but the strength of the defeasible reason. The next thing to notice is that any defeasible argument can be reformulated so that defeasible reasons are only used in subarguments of this form, and then all subsequent steps of reasoning are deductive. The conclusion of the defeasible argument is thus a deductive consequence of the premises together with a number of conditionals justified in this way. By the weakest link principle for deductive arguments, the degree of support of the conclusion should then be the minimum of (1) the degrees of justification of the premises used in the argument and (2) the strengths of the defeasible reasons:

> The argument strength of a defeasible argument is the minimum of the strengths of the defeasible reasons employed in it and the degrees of justification of its premises.

This is the general *Weakest Link Principle*. The problem of computing argument strengths is thus computationally simple.

---

[10] The same objection can be leveled against the Dempster-Shafer theory (Dempster 1968, Shafer 1976).

5. The Accrual of Reasons

If we have two independent reasons for a conclusion, does that make the conclusion more justified than if we had just one? It is natural to suppose that it does,[11] but upon closer inspection that becomes unclear. Cases that seem initially to illustrate such accrual of justification appear upon reflection to be better construed as cases of having a single reason that subsumes the two separate reasons. For instance, if Brown tells me that the president of Fredonia has been assassinated, that gives me a reason for believing it; and if Smith tells me that the president of Fredonia has been assassinated, that also gives me a reason for believing it. Surely, if they both tell me the same thing, that gives me a better reason for believing it. However, there are considerations indicating that my reason in the latter case is not simply the conjunction of the two reasons I have in the former cases. Reasoning based upon testimony is a straightforward instance of the statistical syllogism. We know that people tend to tell the truth, and so when someone tells us something, that gives us a defeasible reason for believing it. This turns upon the following probability being reasonably high:

(1) prob(*P* is true / *S* asserts *P*).

Given that this probability is high, I have a defeasible reason for believing that the president of Fredonia has been assassinated if Brown tells me that the president of Fredonia has been assassinated.

In the discussion of the weakest link principle, I urged that argument strengths do not conform to the probability calculus. However, that must be clearly distinguished from the question of whether probabilities license defeasible inferences. In fact, I think that a large proportion of our defeasible inferences are based upon probabilities. Such inferences proceed in terms of the statistical syllogism, described in section one and formulated as principle (SS). When we have the concurring testimony of two people, our degree of justification is not somehow computed by applying a predetermined function to the latter probability. Instead, it is based upon the quite distinct probability

(2) prob(*P* is true / S_1 asserts *P* and S_2 asserts *P* and $S_1 \neq S_2$).

The relationship between (1) and (2) depends upon contingent facts about the linguistic community. We might have one community in which speakers tend to make assertions completely independently of one another, in which case (2) > (1); and we might have another community in which speakers tend to confirm each other's statements only when they are fabrications, in which case (2) < (1). Clearly our degree of justification for believing *P* will be different in the two linguistic communities. It will depend upon the value of (2), rather than being some function of (1).

It is important to distinguish *epistemic reasoning* — reasoning about what to believe — from *practical reasoning* — reasoning about what actions to perform. These two kinds of reasoning

---

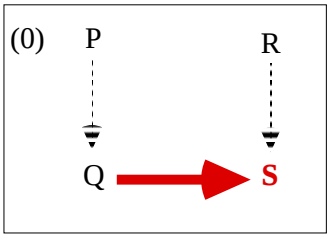
[11] See, for example, Verheij (1996).

have quite different logical properties, as is illustrated at length in Pollock and Cruz (1999). Something like the accrual of reasons seems to hold for practical reasoning. Normally, two independent reasons for choosing a particular action provide a stronger justification for choosing it than either reason by itself. But we cannot conclude from this that the same thing is true of epistemic reasoning.

All examples I have considered that seem initially to illustrate the accrual of reasons in epistemic reasoning turn out in the end to have this same form. They are all cases in which we can estimate probabilities analogous to (2) and make our inferences on the basis of the statistical syllogism rather than on the basis of the original reasons. Accordingly, I doubt that epistemic reasons do accrue. If we have two separate undefeated arguments for a conclusion, the degree of justification for the conclusion is the maximum of the strengths of the two arguments. This will be my assumption.

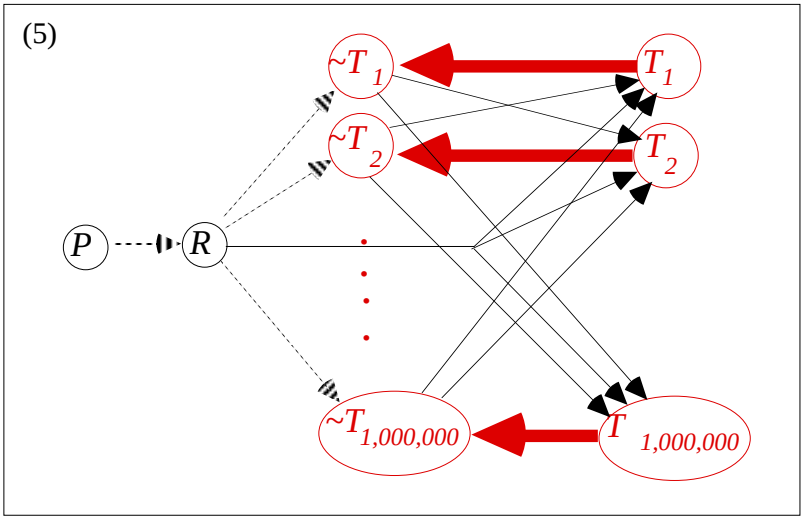
6. The Influence of Defeaters

Thus far I have considered how reason-strengths and the degrees of justification of premises affect the degree of justification of a conclusion. The third variable we must consider is the presence of arguments supporting defeaters. Suppose we have only two arguments to consider, and the conclusion of one of them is a defeater for the final step of the other, as diagrammed in inference-graph (0). How should this affect the degree of justification of *S*?



It seems clear that if the argument strength of the left argument (for *Q*) is as great as that of the right argument (for *S*), then the degree of justification of *S* should be 0. But what if the argument strength of the left argument is less than that of the right argument? In my (1995), I maintained that defeat was an all-or-nothing matter, and hence weaker defeaters leave arguments unaffected. In the scenario just described, this has the consequence that the degree of justification of *S* is the same as the argument strength of the right argument. However, there are some examples that now convince me that this is incorrect. The simplest examples have to do with biased lotteries. To see how these examples work, recall the earlier analysis of reasoning about fair lotteries. Consider a fair lottery consisting of 1 million tickets, and suppose it is known that one and only one ticket will win. Observing that the probability is only .000001 of any particular ticket being drawn given that it is a ticket in the lottery, it seems initially reasonable to employ the statistical syllogism and accept the conclusion regarding any particular ticket that it will not be drawn. This reasoning is completely general and applies to each ticket. However, these conclusions conflict jointly with something else we are justified in believing, viz., that some ticket will be drawn. We cannot be justified in believing each member of an explicitly contradictory

set of propositions, and we have no way to choose between them, so it follows intuitively that we are not justified in believing of any ticket that Jones did not hold that ticket. The formal reconstruction of this reasoning proceeds by observing that this is a case of collective defeat. For each *n*, the statistical syllogism provides a defeasible reason for believing "$\sim T_n$". But for each *k*, we have an equally strong defeasible reason for believing each "$\sim T_k$". We know that some ticket will be drawn. Thus we can construct the counterargument diagramed in inference-graph (5) for the conclusion that "T_n" is true. Our reason for believing each "$\sim T_k$" is as good as our reason for believing "$\sim T_n$", so we have as strong a reason for "T_n" as for "$\sim T_n$". Hence our defeasible reason for the latter is defeated and we are not justified in believing "$\sim T_n$".



Next, consider lottery 2, which is a biased lottery consisting of just ten tickets. The probability of ticket 1 being drawn is .000001, and the probability of any other ticket being drawn is .111111. It is useful to diagram these probabilities as in figure 2. In *this* lottery, it seems reasonable to infer that ticket 1 will not be drawn, because the probability of any other ticket being the drawn is more than 100,000 times greater. This can be justified as follows. As before, we have a defeasible reason for believing "$\sim T_n$", for each *n*. But these reasons are no longer of equal strength. Because ticket 1 is much less likely to be drawn than any other ticket, we have a much stronger reason for believing that ticket 1 will not be drawn. As before, for $n > 1$, we have the counterargument diagramed in inference-graph (5) for "T_n", and that provides as good a reason for believing "T_n" as we have for believing "$\sim T_n$". Thus the defeasible reason for "$\sim T_n$" is defeated. But we do not have as good a reason for believing "T_1" as we do for believing "$\sim T_1$". An argument is only as good as its weakest link, and the counterargument for "T_1" employs the defeasible reasons for "$\sim T_n$" for $n > 1$. These reasons are based upon lower probabilities (of value .888889) and hence are not as strong as the defeasible reason for "$\sim T_1$" (based upon a probability of value .999999). Thus, although we have a reason for believing "T_1", we have a better reason for believing "$\sim T_1$", and so on sum we are justified in believing the latter.

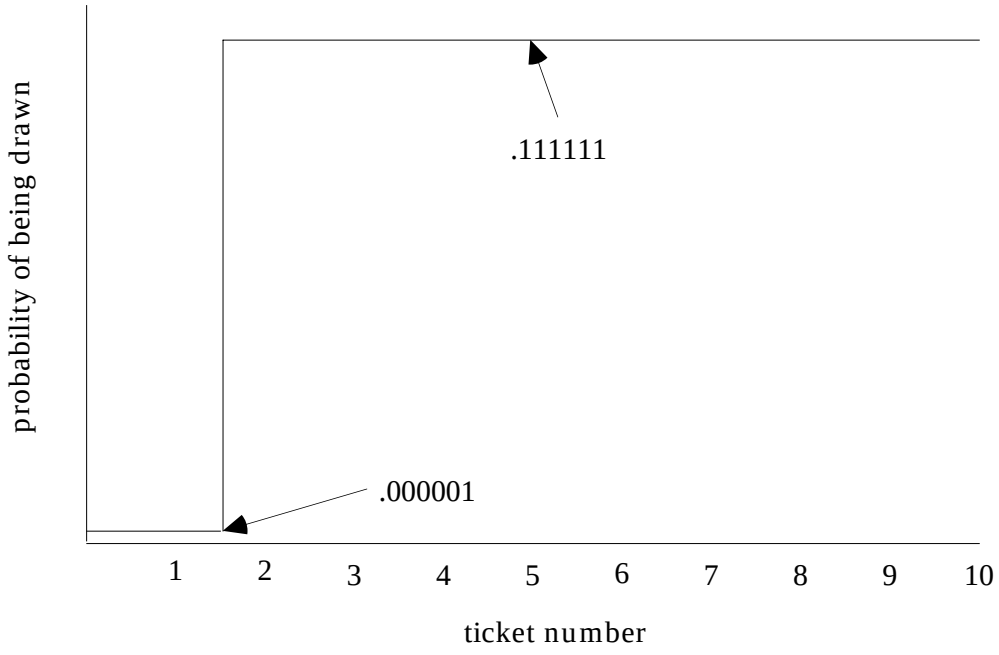


Figure 2. Lottery 2.

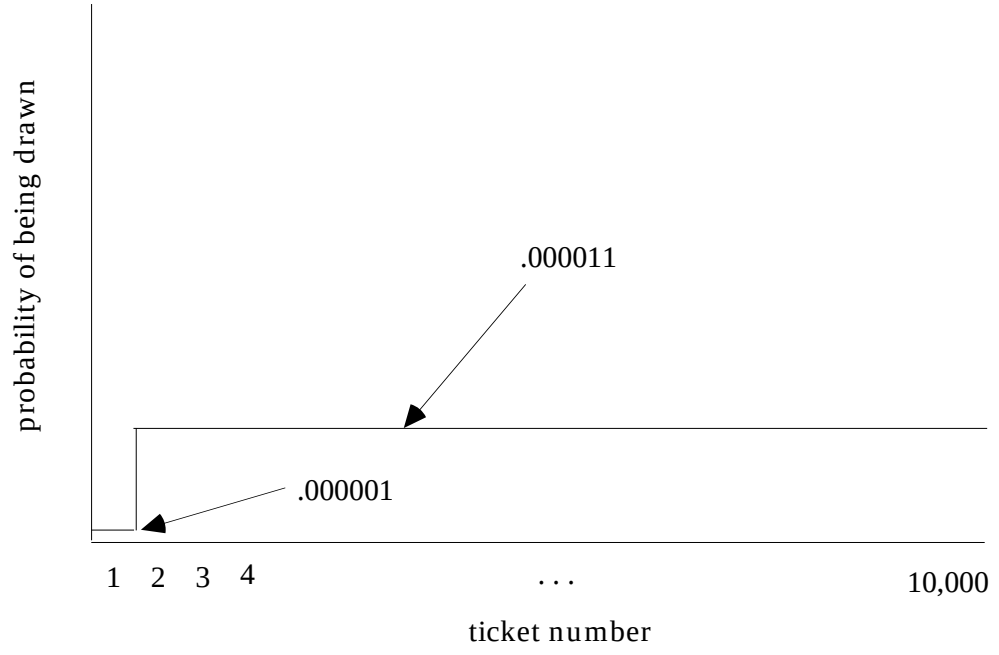


Figure 3. Lottery 3.

Now contrast lottery 2 with lottery 3, which consists of 10,000 tickets. In lottery 3, the probability of ticket 1 being drawn is still .000001, but the probability of any other ticket being drawn is .000011. This is diagramed as in figure 3. It may still be reasonable to infer that ticket 1 will not be drawn, but, and this is the crucial observation, the justification for this conclusion does not

seem to be nearly so strong. This is because although we have the same defeasible argument for "$\sim T_1$", the reasons involved in the counterargument for "T_1" are now much better, being based upon a probability of .999989. They are still not strong enough to defeat the argument for "$\sim T_1$" outright, but they seem to weaken the justification. Thus the degree of justification for "$\sim T_1$" is lower in lottery 3 than it is in lottery 2. The difference between lottery 2 and lottery 3 seems to illustrate that defeaters that are too weak to defeat a conclusion outright may still lower the degree of justification. In other words, they act as *diminishers*.

In my (1990) I argued that reasoning similar to this treatment of biased lotteries is what underlies statistical induction. In statistical induction, having observed a sample of *A*'s and found that the proportion of members of the sample that are *B*'s is *r*, we infer defeasibly that the probability of an arbitrary *A* being a *B* is approximately *r*, i.e., lies in an interval [*r*–δ,*r*+δ] around the observed relative frequency *r*. The logic of the reasoning that allows us to conclude this is the same as that involved in reasoning about biased lotteries. We know that the actual probability *p* is in the interval [0,1]. This is like knowing that some ticket will be drawn. If the sample is large, then for each choice of *p* the probability of getting the observed relative frequency is low given that *p* is the actual probability, but for some choices of *p* (those further from *r*) it is lower than for others. So we can reason as in the biased lottery that the actual probability does not lie on the tails of the bell curve.

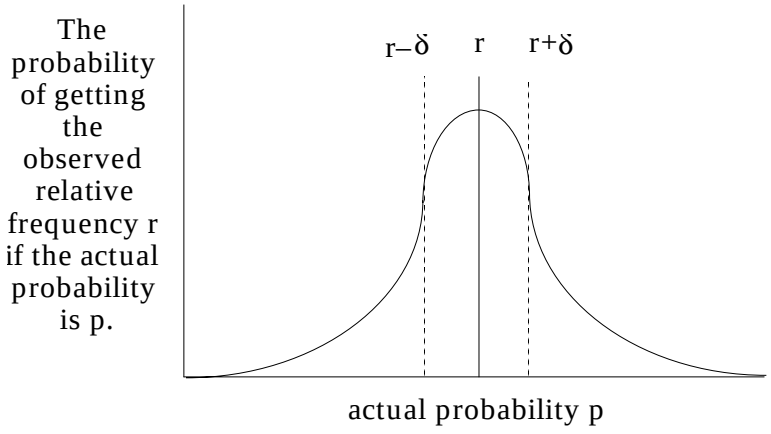


Figure 4. Statistical Induction

Biased lotteries and statistical induction illustrate that given an undefeated argument for *P* and an otherwise undefeated weaker argument for ~*P*, the degree of justification for *P* should be the argument strength of the first argument decremented by an amount determined by the argument strength of the second argument. That is, there should be a function *J* such that given two arguments that rebut one another, if their strengths are *x* and *y*, the degree of justification for the conclusion of the former is $J(x,y)$. *J* must satisfy the following conditions:

(3) $J(x,y) \leq x$
(4) $J(x,0) = x$
(5) if $y \geq x$ then $J(x,y) = 0$
(6) if $x \geq z$ and $w \geq y$ then $J(x,y) \geq J(z,w)$.

This leaves undetermined how $J(x,y)$ and $J(z,w)$ compare in cases in which $x \geq z$ and $w < y$. To resolve these cases, let us replace (3) and (6) by a stronger assumption:

(7) If $\varepsilon \geq 0$, $J(x+\varepsilon,y+\varepsilon) = J(x,y)$.

This is a "linearity assumption". It tells us that increasing the argument strength and the defeat strength by the same amount ε leaves the resulting degree of justification unchanged. With this further assumption, it becomes determinate how $J(x,y)$ and $J(z,w)$ compare, for any choice of x,y,z,w. Let us define:

$$x \sim y = \begin{cases} x - y \text{ if } y < x \\ 0 \text{ otherwise} \end{cases}$$

Then we have in general:

Theorem 1: If (4),(5), and (7) hold, $J(x,y) = x \sim y$.

Proof: If $x \leq y$, then by (5) $J(x,y) = 0 = x \sim y$. Suppose instead that $x \geq y$. Then by (7) and (4), $J(x,y) = J(x–y,0) = x–y = x \sim y$. ■

So I will assume:

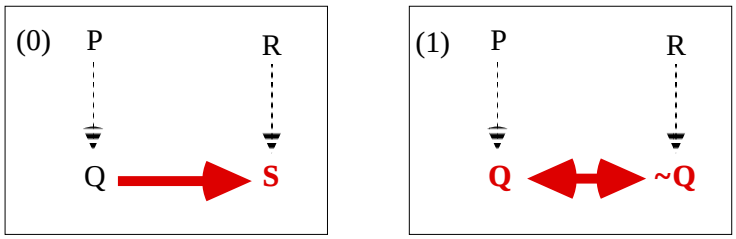
(8) Given an otherwise undefeated argument of strength x supporting P, and an otherwise undefeated argument of strength y supporting $\sim P$, and no other relevant arguments, the degree of justification of P is $x \sim y$.

The argument for diminishers is based on intuitive examples. In particular, I have appealed to biased lotteries. If one ticket is *much* more improbable than the others, it seems reasonable to conclude that it will not be drawn. But if it is only minutely less probable, that does not seem reasonable. In general, if the argument for a defeater is only minutely weaker than the argument for a conclusion, it does not seem reasonable to regard the conclusion as unscathed. These observations seem to me to be unassailable.

7. Computing Defeat Statuses

How should diminishers affect our defeat status computation? I assume that rather than merely assigning "defeated" or "undefeated", a status assignment should now assign numerical degrees of justification $j(P)$ to nodes. How is $j(P)$ determined? In the following examples, unless I explicitly say otherwise, I assume throughout that the reason-strengths are at least as great as the degrees of justification of the initial nodes so that they can be ignored in computing degrees of justification. If we consider a very simple inference-graph like (0), it seems clear that we should have $j(Q) = j(P)$, $j(S) = j(R) \sim j(Q)$. This is in accordance with principle (8). So if $j(P) \geq j(R)$, S is

defeated, otherwise it is diminished.



Consider the marginally more complicated inference-graph (1). Here there seem to be two possibilities regarding how the degrees of justification are to be computed:

(a) We could have $j(Q) = j(P) \sim j(\sim Q)$ and $j(\sim Q) = j(R) \sim j(Q)$.
(b) We could have $j(Q) = j(P) \sim j(R)$ and $j(\sim Q) = j(R) \sim j(P)$.

These seem to be the only two possibilities for this simple inference-graph. However, (a) is not a genuine possibility. We have the following theorem:

Theorem 2: If $j(P) > j(R)$, $j(Q) = j(P) \sim j(\sim Q)$ and $j(\sim Q) = j(R) \sim j(Q)$ then $j(Q) = j(P)$ and $j(\sim Q) = 0$.

Proof: Suppose $j(Q) \neq j(P)$. As $j(Q) = j(P) \sim j(\sim Q)$, $j(Q) < j(P)$, so $j(\sim Q) \neq 0$. Then as $j(\sim Q) = j(R) \sim j(Q)$, $j(\sim Q) = j(R) - j(Q) \leq R$. By assumption, $j(P) > j(R)$, so $j(P) > j(\sim Q)$, and hence $j(Q) = j(P) \sim j(\sim Q) = j(P) - j(\sim Q) = j(P) - j(R) + j(Q)$. Thus $j(R) = j(P)$, which is impossible given the assumption that $j(P) > j(R)$. So by reductio, $j(Q) = j(P)$. As $j(P) > j(R)$, $j(P) > 0$. But $j(Q) = j(P) \sim j(\sim Q)$, so $j(\sim Q) = 0$. ■

Thus (a) is incompatible with diminishing. That leaves only (b) as a possible computation of defeat statuses in inference-graph (1). That is,

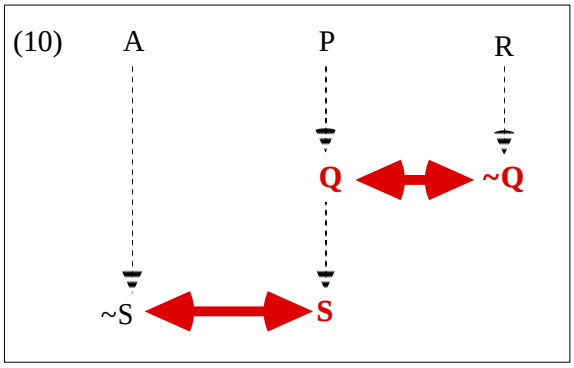
$j(Q) = j(P) \sim j(R)$ and $j(\sim Q) = j(R) \sim j(P)$.

This means that in computing $j(Q)$, we take the strength of the argument supporting it, in this case determined by $j(P)$, and then subtract the strength of the argument supporting the defeater, i.e., $j(R)$. We do not subtract the strength of the defeater itself, i.e., we do not subtract $j(\sim Q)$.

If we apply (b) to the case in which $j(P) = j(R)$, we get a single status assignment in which $j(Q) = j(\sim Q) = 0$. This has a surprising consequence. The semantics for defeasible reasoning described in section one, as well as default logic, the stable model semantics, circumscription, and almost all standard semantics for defeasible reasoning and nonmonotonic logic, support what I have called (1987) "presumptive defeat".[12] For example, consider inference-graph (10). A defeated conclusion like Q that is assigned "defeated" in some status assignment and "undefeated" in another retains the ability to defeat. In the case of inference-graph (10) this has the consequence

[12] The only semantics I know about that does not support presumptive defeat are certain versions of Nute's (1992) defeasible logic. See also Covington, Nute, and Vellino (1997), and Nute (1999).

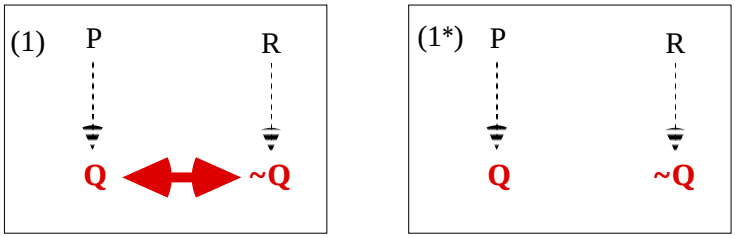
that *S* is assigned "defeated" in those status-assignments in which *Q* is assigned "defeated", but *S* is assigned "undefeated" and **~***S* is assigned "defeated" in those status-assignments in which *Q* is assigned "undefeated". Touretzky, Horty, and Thomason (1987) called this "ambiguity propagation", and Makinson and Schlechta (1991) called such arguments "Zombie arguments" (they are dead, but they can still get you). However, computation (b) precludes presumptive defeat. It entails that there is a single status assignment in which $j(S) = j(Q) = j(\sim Q) = 0$, and $j(\sim S) = j(A)$. So *Q*, **~***Q*, and *S* are all defeated, and **~***S* is undefeated. Is this the right answer? Consider an example: Jones and Smith hear one weather forecast, but disagree about whether rain was predicted (*Q*). *Q* gives one reason to believe it will rain (*S*). You didn't hear the first forecast, but you hear another forecast (*A*), which says it will not rain. Should you believe **~***S*? I have some inclination to think you should, but I don't find the answer obvious, nor have I ever found another example of presumptive defeat where the answer is obvious. At a recent workshop on defeasible reasoning held at the University of Georgia, I found there to be no consensus among the participants as to the intuitive status of presumptive defeat.



In the absence of clear intuitions, how can we decide whether presumptive defeat is a genuine phenomenon of defeasible reasoning, or an undesirable artifact of existing semantics? The preceding considerations constitute a rather strong argument for the conclusion that presumptive defeat is incompatible with diminishing. If it is granted that a correct semantics must accommodate diminishing, this means in turn that most standard semantics for defeasible reasoning produce incorrect assessments of inference-graph (10) even when all premises have the same degree of justification and all reason-strengths are the same. It is generally assumed that by ignoring variations in degrees of justification and reason-strengths we are simplifying the problem of constructing a semantics for defeasible reasons, but the present considerations indicate that by doing so we may also obscure phenomena that have important implications for the semantics even in this simplified case.

The fact that we get a single assignment of defeat statuses in inference-graph (1) suggests that we are not really computing status assignments. Rather, we are computing degrees of justification directly. The appeal to status assignments was motivated by the thought that we could not compute degrees of justification recursively because a node *P* of an inference-graph can have *P*-dependent defeaters (i.e., defeaters that are inference/defeat-descendants of *P*.) But what the present approach may yield is a way of doing a different kind of recursive computation of degrees of justification. In inference-graph (1), we cannot compute $j(Q)$ without having a value for **~***Q*, and we cannot compute $j(\sim Q)$ without having a value for *Q*. So we cannot do a

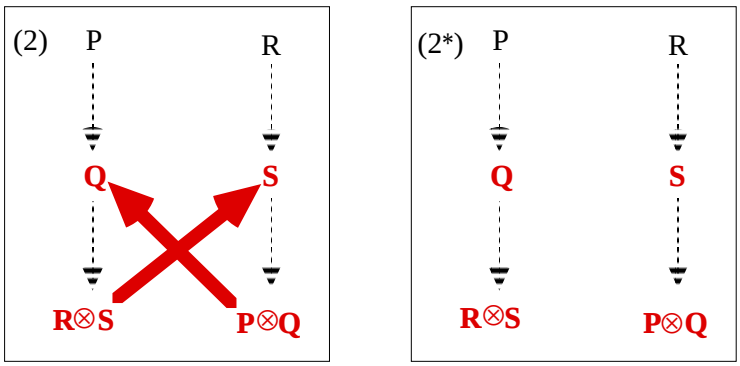
straightforward recursive computation of degrees of justification. However, the values required for Q and $\sim Q$ need not be their degrees of justification. Indeed, they cannot be because those would be zero for each. In computing $j(Q)$, what we want to subtract from $j(P)$ is not $j(\sim Q)$ but rather a measure of the strength of the argument for $\sim Q$. The value of the latter should be $j(R)$. This measure can be obtained by removing the mutual defeat between the two arguments, as in inference-graph (1*), and computing $j(Q)$ and $j(\sim Q)$ in the new inference-graph. Call those values $j^*(Q)$ and $j^*(\sim Q)$. Then returning to inference-graph (1), the values we subtract when we compute $j(Q)$ and $j(\sim Q)$ are $j^*(\sim Q)$ and $j^*(Q)$. That is, $j(Q) = j^*(Q) \sim j^*(\sim Q) = j(P) \sim j(R)$, and $j(\sim Q) = j^*(\sim Q) \sim j^*(Q) = j(R) \sim j(P)$. In constructing inference-graph (1*), we remove two defeat-links. Each link has the characteristic that removing it results in Q no longer having a Q-dependent defeater, and also in $\sim Q$ no longer having a $\sim Q$-dependent defeater.



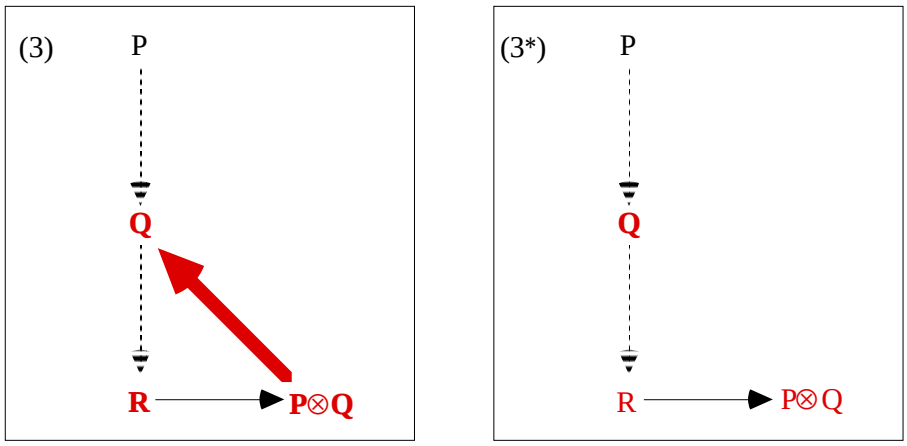
Now consider inference-graph (2). Again, there seem to be two ways the computation of degrees of justification might go. Presumably $j(R \otimes S) = j(Q)$ and $j(P \otimes Q) = j(S)$. Then we might have either:

(a) $j(S) = j(R) \sim j(R \otimes S) = j(R) \sim j(Q)$ and $j(Q) = j(P) \sim j(P \otimes Q) = j(P) \sim j(S)$; or
(b) $j(S) = j(R) \sim j(P)$ and $j(Q) = j(P) \sim j(R)$.

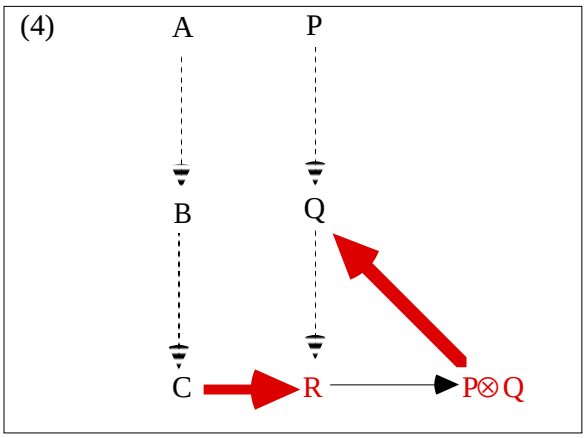
As in theorem 1, (a) is incompatible with diminishing, so (b) seems to be the only possibility. This means that in computing $j(S)$ we begin with the strength of the argument supporting S, i.e., $j(R)$, and then subtract the strength of the argument supporting the defeater $R \otimes S$ would have in the absence of the defeater $P \otimes Q$ that is obtained from S. We compute $j(Q)$ analogously. This is the same thing as computing $j(R \otimes S)$ and $j(P \otimes S)$ in inference-graph (2*) (call the resulting values $j^*(R \otimes S)$ and $j^*(P \otimes Q)$) and then setting $j(Q) = j^*(Q) \sim j^*(P \otimes Q)$ and $j(S) = j^*(S) \sim j^*(R \otimes S)$. Again, the defeat-links we remove in constructing inference-graph (2*) are those such that if we remove either, Q no longer has a Q-dependent defeater, and similarly for S.



Consider inference-graph (3). Presumably $j(P\otimes Q) = j(R) = j(Q) = 0$. We can get that by ruling that $j(Q) = j(P) \sim j(P)$. So here we begin with the strength of the argument supporting Q, i.e., $j(P)$, and then subtract the strength the argument supporting $P\otimes Q$ would have in the absence of the defeater (itself) that is obtained from Q. This is the value assigned to $(P\otimes Q)$ in inference-graph (3*). Again, the defeat-link we remove in constructing (3*) is the only defeat-link whose removal results in Q no longer having a Q-dependent defeater.

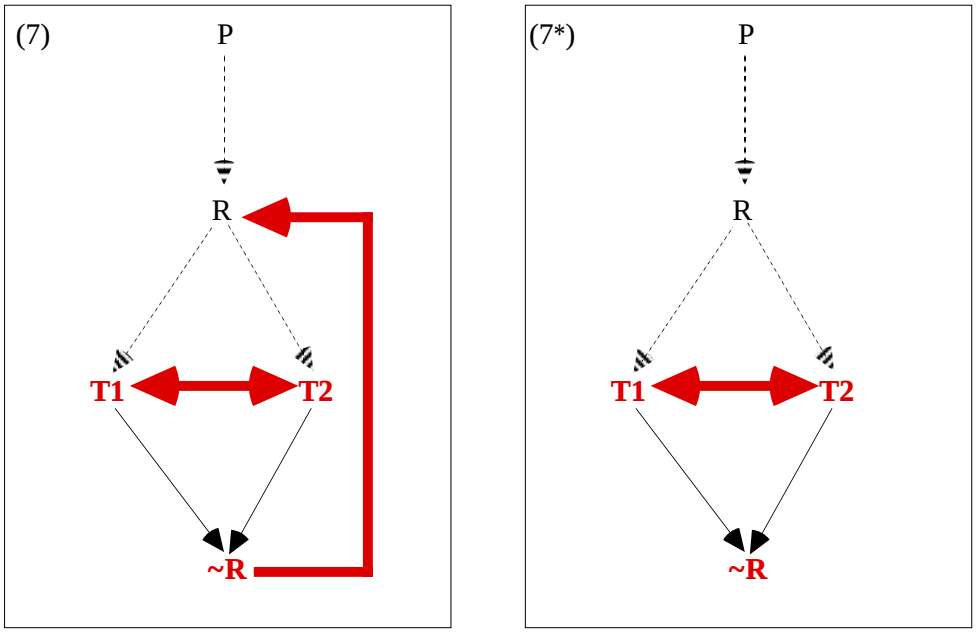


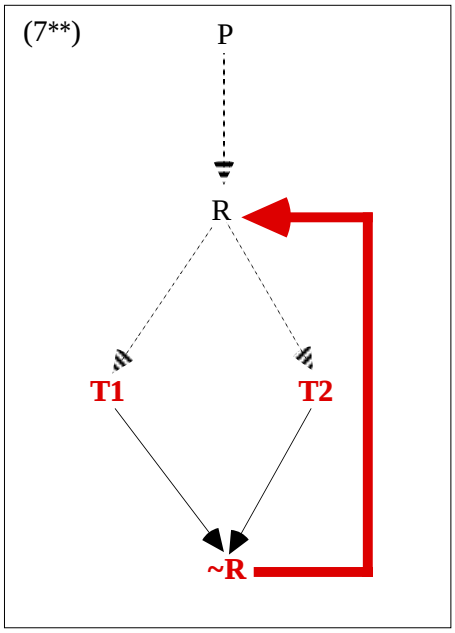
Inference-graph (4) is analogous. In (4) we compute $j^*(P\otimes Q)$ by removing the defeat-link between $P\otimes Q$ and Q. So $j^*(P\otimes Q) = j(P) \sim j(A)$. That has the result that if $j(A) = j(P)$ then $j^*(P\otimes Q) = j^*(R) = 0$ and so $j(Q) = j^*(Q) \sim j^*(P\otimes Q) = j(P)$. I will continue to refer to $j^*(P\otimes Q)$ as the argument-strength of $(P\otimes Q)$, but note that this is a somewhat more complex notion than the argument-strength discussed in section four. In the present sense, argument-strengths take account not only of the strengths of the premises and inference-schemes but also of the effect of those defeaters that are independent of the node being evaluated.



Finally, consider the lottery paradox paradox, in the guise of inference-graph (7). We construct (7*) by removing the only defeat-link whose removal results in R no longer having an R-dependent defeater. In (7*), the triangle consisting of R, $T1$ and $T2$ is analogous to inference-graph (1), and so $j^*(T1) = j^*(R) \sim j^*(R) = 0$, and $j^*(T2) = j^*(R) \sim j^*(R) = 0$. $j^*(\sim R) = \min\{j^*(T1), j^*(T2)\} = 0$. Then $j(R)$

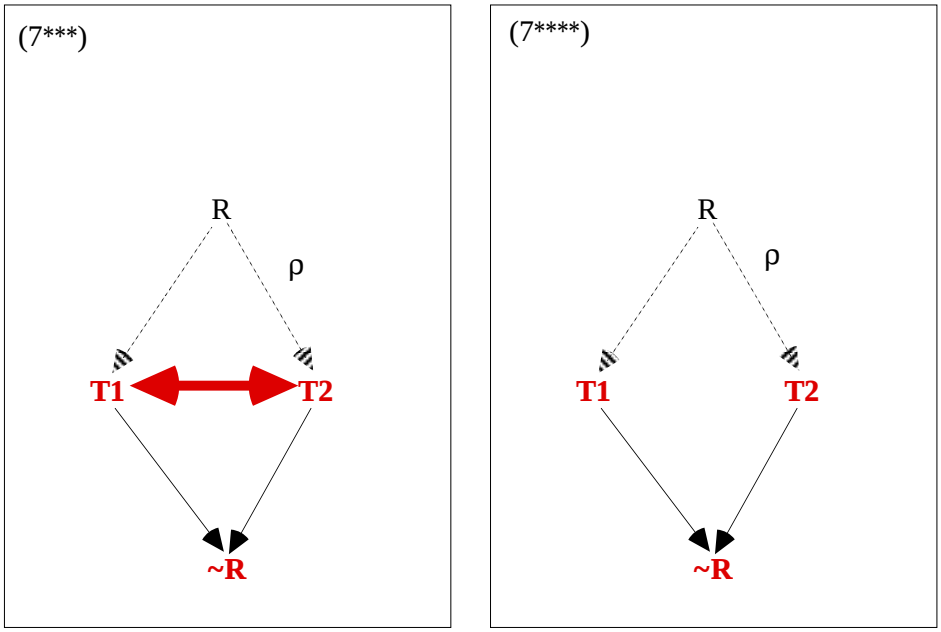
$= j^*(R) \sim j^*(\sim R) = j(P)$.





When we turn to the computation of $j(T1)$ and $j(T2)$ in (7), we encounter an additional complication. Removing the defeat-links whose removal results in *T1* no longer having a *T1*-dependent defeater produces inference-graph (7**). On the strength of the preceding examples we might expect that $j(T1) = j^{**}(T1) \sim j^{**}(T2)$. However, inference-graph (7**) is analogous to inference-graph (3). Accordingly, $j^{**}(T1) = j^{***}(T2) = j^{**}(R) = j^{**}(\sim R) = 0$. This produces the intuitively correct answer that $j(T1) = 0$, but it seems to do so in the wrong way. To see this, let us modify the example slightly by taking it to represent a biased lottery. More precisely, let us suppose that the reason-strength of the reason supporting the inference of *T1* from *R* is at least

as great as the degree of justification of *P* (and hence of *R*), but suppose the reason-strength ρ of the reason supporting the inference of *T2* from *R* is less than the degree of justification of *P*. In this case *T2* should be defeated, and hence *~R* should be defeated, but *T1* should only be diminished. However, this change does not affect the computation of degrees of justification in inference-graph (7**). We still get that $j^{**}(T1) = j^{**}(T2) = j(R) = j^{**}(\sim R) = 0$.



The source of the difficulty is that the computation of degrees of justification is not being done recursively. We should first compute the degree of justification of *R*, as above. Then holding that degree of justification fixed we should go on to compute the degrees of justification of its inference-children, in this case *T1* and *T2*. This amounts to regarding *R* as an initial node whose degree of justification is that computed in inference-graph (7). This yields inference-graph (7***). The degrees of justification for *T1* and *T2* are then those computed in inference-graph (7***). That computation proceeds by constructing inference-graph (7****) and computing $j^{****}(T1) = j(R)$, $j^{****}(T2) = \rho$, and so $j(T1) = j^{***}(T1) = j^{****}(T1) \sim j^{****}(T2) = j(R) \sim \rho$. Analogously, $j(T2) = j^{****}(T2) \sim j^{****}(T1) = 0$. Continuing recursively, $j(\sim R) = \min\{j(T1), j(T2)\} = 0$.

The defeat status computation that emerges from these examples proceeds in accordance with two rules. We begin with an inference-graph *G*. Where φ is a node of *G*, let $j(\varphi, G)$ be the degree of justification of φ in *G*. The first rule governs the case in which *P* has no *P*-dependent defeaters. In that case, the computation proceeds in accordance with the originally proposed rules (1)–(3). This has the effect of computing degrees of justification in terms of the degrees of justification of the basis of *P* and the defeaters of *P*. The second rule governs the case in which *P* has *P*-dependent defeaters. In that case the computation of the degree of justification of *P* proceeds instead in terms of the argument-strength for *P* and maximal argument-strength of the *P*-dependent defeaters. Those argument-strengths are computed by constructing a new inference-graph G_P by removing each defeat-link of *G* whose removal results in *P* no longer having a *P*-dependent defeater. More generally, there can be parallel routes from one node to another, with the result that we must delete multiple defeat-links to ensure that *P* no longer has a

P-dependent defeater. So let us define:

> A defeat-link is *P-critical* in an inference-graph G iff it is a member of some minimal set of defeat-links such that the removal of all the links in the set results in *P* no longer having a *P*-dependent defeater.

Let G_P be the inference-graph that results from removing all *P*-critical defeat-links from G and making all *P*-independent nodes φ initial with $j(\varphi,G_P) = j(\varphi,G)$. The argument strengths in G are then the degrees of justification in G_P.

Putting this altogether, the two rules are as follows (where $\max(\varnothing) = 0$) :

(DJ1) If *P* is inferred from the basis $\{B_1,...,B_n\}$ in an inference-graph G in accordance with a reason-scheme of strength ρ, $D_1,...,D_k$ are the defeaters for *P*, and no D_i is *P*-dependent, then

$$j(P,G) = \min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim \max\{j(D_1,G),...,j(D_k,G)\}.$$

(DJ2) If *P* has *P*-dependent defeaters $D_1,...,D_k$ in G and G_P results from deleting all *P*-critical defeat-links from G and making all *P*-independent nodes φ initial with $j(\varphi,G_P) = j(\varphi,G)$, then

$$j(P,G) = j(P,G_P) \sim \max\{j(D_1,G_P),...,j(D_k,G_P)\}.$$

The general idea is that the computation of degrees of justification is made recursive by appealing to argument strengths rather than degrees of justification in ungrounded cases. Argument strengths are computed by computing degrees of justification in simpler inference-graphs from which the source of ungroundedness has been removed. The result is a recursive computation of degrees of justification.

8. Collaborative Defeat

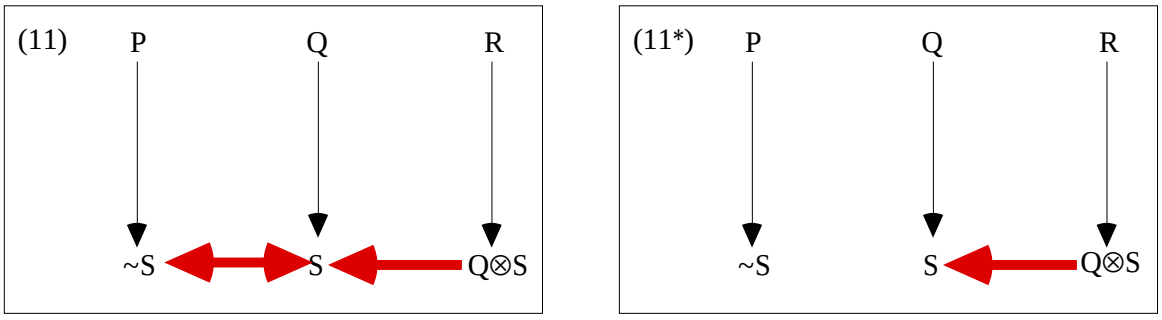
The rules (DJ1) and (DJ2) have the effect of dividing the defeaters of *P* into two sets — those that are *P*-dependent and those that are not. These two sets of defeaters then affect $j(P,G)$ separately. Those that are *P*-independent will diminish the value of *P* in G_P, reducing $j(P,G_P)$, and then those that are *P*-dependent will further diminish the value of *P* in G. So if D_1 is the most strongly supported defeater that is *P*-independent, and D_2 is the most strongly supported defeater that is *P*-dependent, the result will be that $j(P,G) = j(P,G_P) \sim j(D_2,G_P) = (\min\{\rho,j(B_1,G_P),...,j(B_n,G_P)\} \sim j(D_1,G_P)) \sim j(D_2,G_P)) = (\min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim j(D_1,G)) \sim j(D_2,G_P))$ $= \min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim (j(D_1,G) + j(D_2,G))$. In this way, we can replace the sequential application of (DJ2) and (DJ1) by the application of a single principle:

(DJ) If *P* is inferred from the basis $\{B_1,...,B_n\}$ in an inference-graph G in accordance with a reason-scheme of strength ρ, $D_1,...,D_k$ are the *P*-independent defeaters for *P*, $D_{k+1},...,D_m$ are the *P*-dependent defeaters of *P*, and G_P results from deleting all *P*-critical defeat-links from G and making all *P*-independent nodes φ initial with $j(\varphi,G_P) = j(\varphi,G)$, then $j(P,G) = \min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim [\max\{j(D_1,G),...,j(D_k,G)\} + \max\{j(D_{k+1},G_P),...,j(D_m,G_P)\}]$.

This produces a kind of double counting of defeaters. It has the consequence that although no single defeater may be sufficient to defeat the inference to *P*, two defeaters can accomplish that by acting in unison. I will call this *collaborative defeat*.[13]

Collaborative defeat might seem suspect. However, I will now argue that there are examples which illustrate that it actually occurs in defeasible reasoning. As an autobiographical note, I first became convinced of the need for collaborative defeat on the basis of examples like those I will present below. This was long before I had a theory of diminishers that entailed collaborative defeat. My expectation was that I would have to construct an initial theory to accommodate diminishers, and then embellish it in some rather ad hoc way to make it compatible with collaborative defeat. Thus I am delighted to find that collaborative defeat simply falls out of my theory of diminishers without any ad hoc treatment.

To set the stage, note that you cannot have two rebutting defeaters such that one is *P*-dependent and the other is not. Rebutting defeaters of *P* are always *P*-dependent. This is because rebutting defeat is symmetrical, so if you can follow defeat-links in one direction, you can follow them back again. Thus you can have one defeater that is *P*-dependent and another that is not only if one of them is an undercutting defeater. The general form is that of inference-graph (11). To compute $j(S,G_{11})$, we construct inference-graph (11*). $j(S,G_{11}) = j(Q,G_{11}) \sim j(R,G_{11})$, and so $j(S,G_{11}) = (j(Q,G_{11}) \sim j(S,G_{11})) \sim j(P,G_{11}) = j(Q,G_{11}) \sim (j(P,G_{11}) + j(R,G_{11}))$.



To confirm that the computation in inference-graph (11) is correct, we must consider examples that mix rebutting defeat and undercutting defeat. One place in which this is common is in inferences involving a total-evidence requirement. Suppose the total evidence consists of $P_1,...,P_n$, and this supports the conclusion *Q*. Typically, some of our evidence will be somewhat better justified than the rest. Suppose the maximally justified subset of our evidence consists of $P_1,...,P_k$. This may support a conflicting conclusion *R*. If the rest of the evidence is only slightly less justified, we will want to draw the conclusion *Q* rather than *R*. The logic of this is that we will have three arguments. The first argument infers *Q* from $P_1,...,P_n$. The second argument infers *R* from $P_1,...,P_k$. The third argument appeals to the fact that we have the evidence $\{P_1,...,P_n\}$ and it is a larger set of evidence than $\{P_1,...,P_k\}$, and it infers an undercutting defeater for the second argument. The first two arguments rebut one another, and the third argument undercuts the

[13] Normally, the collaborating defeaters will be reasons for different conclusions, viz., ~*P* and (*P*⊗*Q*). But in some cases they can both be reasons for (*P*⊗*Q*), in which case this has a similar flavor to the accrual of reasons. It is not anything that simple, however, because each defeater is the strongest defeater (not a sum) from a whole class of defeaters — the dependent and independent ones.

second argument. However, the first and third arguments are weaker than the second argument, because they depend upon all of $P_1,...,P_n$ and hence depend upon a weaker link than the second argument. Without collaborative defeat, the second argument would defeat the first argument, and the first argument and third argument would each do nothing but diminish the justification of the conclusion *R* of the second argument. Thus we would be led draw the conclusion *R* on the basis of a subset of the evidence rather than drawing the conclusion *Q* that is supported by the total evidence. That is the intuitively wrong answer. Collaborative defeat, however, can produce the right answer. First, argument three supports the *R*-independent undercutting defeater, which diminishes the degree of justification of *R*. Then it is that weaker conclusion that squares off against the support of *Q* by the first argument (i.e., the *R*-dependent rebutting defeater), and hence *Q* can be justified and *R* unjustified.

To illustrate this general phenomenon, consider a case of direct inference. In direct inference we reason from general probabilities (symbolized using "prob") to single case probabilities (symbolized using "PROB"). The basic idea behind direct inference was first articulated by Hans Reichenbach (1949): in determining the probability that an individual *c* has a property *F*, we find the narrowest reference class *X* for which we have reliable statistics and then infer that PROB(*Fc*) = prob(*Fx*/*x*∈ *X*). For example, insurance rates are calculated in this way. Although there is general agreement that direct inference is based upon some such principle as this, there is little agreement about the precise form the theory should take.[14] In my (1990) I proposed reconstructing direct inference as defeasible reasoning that proceeds primarily in terms of the following two principles:

(DI) "prob(*F*/ *G*) = *r* & **J**(*Gc*) & **J**(*P* ↔ *Fc*)" is a defeasible reason for "PROB(*P*) = *r* ".

(SD) "prob(*F*/ *H*) ≠ prob(*F*/ *G*) & **J**(*Hc*) &□∀(*H* → *G*)" is an undercutting defeater for (DI).

Here "**J***P*" means "*P* is justified", ∀*P* is the universal closure of *P*, and □ is logical necessity.[15] The defeaters provided by (SD) are *subproperty defeaters* (reminiscent of the subproperty defeaters for the statistical syllogism). To illustrate, suppose you are an insurance actuary computing auto insurance rates for Helen. You know that Helen is female, and the probability of a female driver having an accident within one year is .03. This gives you a defeasible reason for concluding that the probability of Helen having an accident is .03. But you also know that Helen is a reckless driver, and the probability of a female reckless driver having an accident in a year is .1. This gives you a defeasible reason for the conflicting conclusion that the probability of Helen having an accident is .1. Because the latter inference is based upon more information, you will accept it and reject the former inference. Formally, this is because the latter inference is based upon information that provides a subproperty defeater for the former inference, so the former inference is defeated and the latter is left undefeated. Let us examine this formal reconstruction carefully. There are three arguments involved:

Argument 1 — for the conclusion that the probability of Helen having an accident is .03, based

---

[14] For instance, see Kyburg (1974) and Levi (1977).

[15] These two principles do not by themselves provide a complete account, but see my (1990) for more details.

upon the fact that Helen is female.

Argument 2 — for the conclusion that the probability of Helen having an accident is .1, based upon the fact that Helen is female and a reckless driver.

Argument 3 — for the conclusion that Helen is a reckless driver, where the probability of a female driver having an accident is different from the probability of a female reckless driver having an accident.

Arguments 1 and 2 provide rebutting defeaters for each other. If they were of equal strength and there were no other relevant arguments, then both would be defeated and you would be unable to draw any conclusion about the probability of Helen having an accident. However, the conclusion of argument 3 supports an undercutting defeater for argument 1. So if all the arguments are of equal strength, argument 1 will be defeated, leaving argument 2 undefeated. The inference-graph representing all three arguments can be diagrammed as in figure 4.

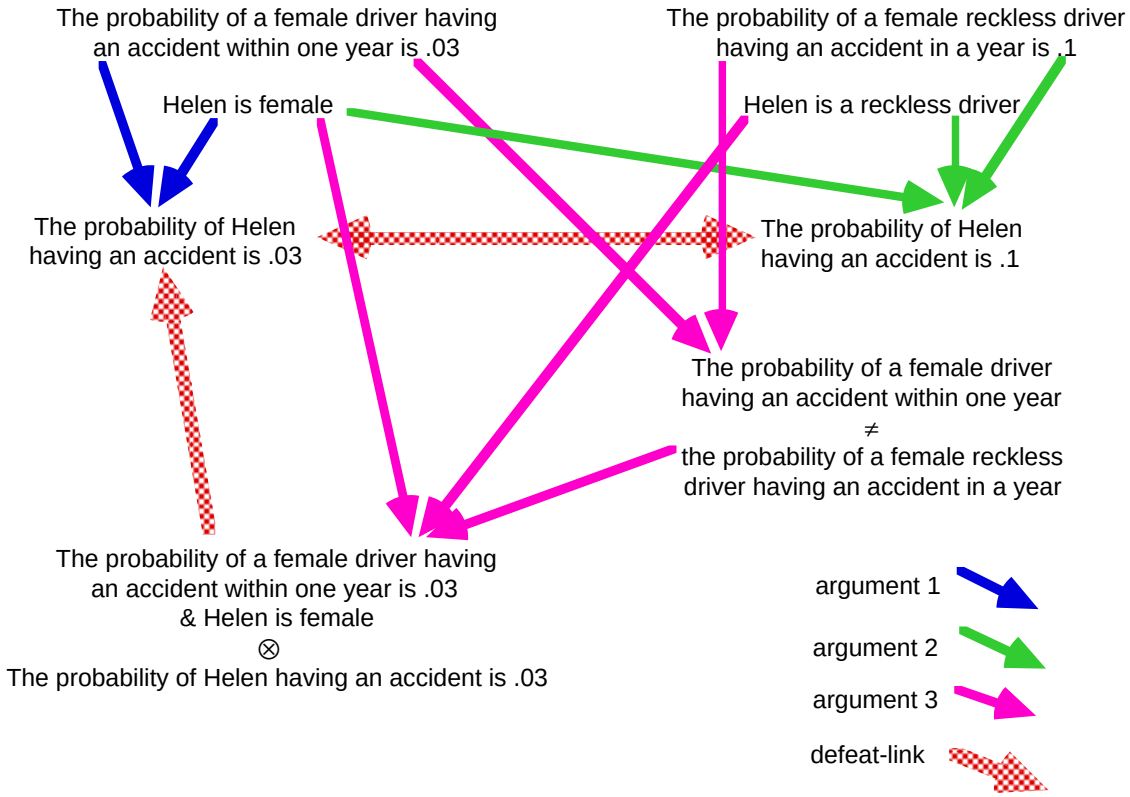


Figure 4. Direct inference

The preceding analysis of the problem assumes that we are equally justified in believing that Helen is female and that Helen is a reckless driver, but typically we will be better justified in believing the former (although adequately justified in believing the latter as well). Arguments 2 and 3 both depend upon knowing that Helen is a reckless driver. Argument 1 depends only upon knowing that Helen is female, which is more strongly justified. This implies that the

argument strengths for arguments 2 and 3 are both less than the argument strength for argument 1. The result is that neither the rebutting defeater provided by argument 2 nor the undercutting defeater provided by argument 3 should be sufficient by itself to defeat argument 1. On the other hand, argument 2 would be defeated outright by the more strongly supported argument 1. This is the intuitively wrong result.

It appears that the only way to get argument 1 defeated and argument 2 undefeated is to allow the rebutting defeater provided by argument 2 and the undercutting defeater provided by argument 3 to work in unison rather than separately. That is precisely what (DJ) accomplishes. Because the undercutting defeater is undefeated, it lowers the degree of support for the conclusion of argument 1. It is that lowered support that should then be compared with the support for the rebutting defeater provided by argument 2, and as the latter is greater, argument 1 is defeated. So the intuitively correct answer is reached by collaborative defeat.

There are numerous other cases in which the reasoning has an analogous structure and variations in the degrees of justification of the premises cause similar prima facie difficulties. In each case, collaborative defeat resolves the prima facie difficulties. For example, the Yale Shooting Problem has played an important role in discussions of the Frame Problem.[16] In the Yale Shooting Problem, we are given that a gun is initially loaded. Then it is pointed at Jones and the trigger is pulled. We suppose we know (simplistically) that if a loaded gun is pointed at someone and the trigger pulled, that person will shortly become dead. The conclusion we are supposed to draw in this case is that Jones will die. The Yale Shooting Problem is the problem of showing how this conclusion can be justified. There are two problems. First, our reason for thinking that the gun is loaded when the trigger is pulled is that it was loaded moments earlier when we checked it. So we are employing temporal projection. Temporal projection provides a defeasible reason for thinking that if something is true at one time then it will still be true later. Some form of temporal projection has been endorsed by most recent authors discussing the frame problem.[17] Using temporal projection, we can construct:

Argument 1 — for the conclusion that Jones will be dead after the shooting, based upon causal knowledge and the temporal projection that the gun will still be loaded when the trigger is pulled.

However, as Hanks and McDermott (1986, 1987) were the first to note, with the help of temporal projection we can also construct a second argument supporting a conflicting conclusion:

Argument 2 — for the conclusion that Jones will still be alive after the shooting, based upon temporal projection from the fact that Jones was initially alive.

In the absence of further arguments, these arguments will defeat each other, leaving us with no justified conclusion to draw about the state of Jones' health. To get the intuitively correct answer, we need a third argument:

---

[16] Hanks and McDermott (1986).

[17] This was proposed by Sandewall (1972), and subsequently endorsed by McDermott (1982), McCarthy (1986), and virtually all subsequent authors.

Argument 3 — supporting an undercutting defeater for argument 1.

The Yale Shooting Problem is resolved by explaining the details of argument 3. I have proposed such a solution in my (1998). For present purposes, the details are not important. Suffice it to say that the undercutter gives inferences based upon causal knowledge priority over those based upon temporal projection and turns on the premise that the gun is still loaded when the trigger is pulled.

A problem derives from the fact that the strength of an inference by temporal projection decreases as the time interval increases. That is, temporal projection gives us a reason for thinking that things won't change, but the greater the time interval the weaker the reason. This is the *temporal decay* of temporal projection discussed at the beginning of section three (see my 1998). If we ignore the temporal decay of temporal projection, the foregoing constitutes a solution to the Yale Shooting Problem. But now suppose, as will typically be the case, that we observe Jones to be alive at a later time (normally, right up to the time we pull the trigger) than we observe the gun to be loaded. I cannot observe the latter at the time I pull the trigger without shooting myself in the face. In this case the strengths of arguments 1 and 3, depending as they do on inferring that the gun is still loaded when the trigger is pulled, may both be less than the strength of argument 2.

The temporal profile we should get is the following. If we observe the gun to be loaded long before observing Jones to be alive (e.g., years ago), our justification for believing Jones to remain alive might be somewhat weakened but not defeated. On the other hand, if we observe the gun to be loaded just shortly before observing Jones to be alive, we should be able to conclude that Jones will die. The intuitive rationale for this profile seems to be as follows. First, we have the undefeated argument 3 for the undercutting defeater for argument 2, but it is weaker than argument 2. Instead of defeating argument 2 outright, it weakens it seriously. This leaves us with only a weak reason for thinking that Jones remains alive. Then argument 1 provides a reason for thinking Jones is dead. If argument 1 turns upon a temporal projection from the distant past, it will not be strong enough to defeat even the weakened argument 2, but if the temporal projection is from a recent observation then argument 1 will be strong enough to defeat the weakened argument 2. This is exactly the computation that results from collaborative defeat.

The upshot is that what seemed initially like a suspicious consequence of (DJ1) and (DJ2) turns out to be a very important logical phenomenon that is often crucial for computing defeat statuses correctly.

9. Measuring Strengths[18]

It has been assumed throughout that reason-strengths and degrees of justification can be measured using the extended reals, although nothing has been said about how that is done. If we are to take strength seriously, we must have some way of measuring it. One way is to compare reasons with a set of standard equally good reasons that have numerical values associated

[18] This section corrects section 9 of my (2001).

with them in some determinant way. I propose to do that by taking the set of standard reasons to consist of instances of the statistical syllogism (SS). For any proposition *p*, we can construct a standardized argument for **~***p* on the basis of the pair of suppositions "prob(*F*/*G*) ≥ *r* & *Gc*" and "(*p* ↔ **~***Fc*)":

1. Suppose prob(*F*/*G*) ≥ *r* & *Gc*.
2. Suppose (*p* ↔ **~***Fc*).
3. *Fc* from 1.
4. **~***p* from 2,3.

where the strength of the argument is a function of *r*. We can measure the strength of a defeasible reason for *p* in terms of that value of *r* such that the conflicting argument from the suppositions "prob(*F*/*G*) ≥ *r* & *Gc*" and "(*p* ↔ **~***Fc*)" exactly counteracts it. The value *r* determines the reason-strength in the sense that the reason-strength is some function **dj**(*r*) of *r*. It is tempting to identify **dj**(*r*) with *r*, but that will not work. The difficulty is that by identifying *J* with **~**, we have required reason-strength to be a cardinal measure that can be meaningfully added and subtracted. Adding and subtracting reason-strengths may not be the same thing as adding and subtracting the corresponding probabilities. To illustrate the general point, should it turn out (it won't) that the reason-strength corresponding to a probability *r* is given by log(*r*), then adding reason-strengths would be equivalent to multiplying probabilities rather than adding them.

I do not have an a priori argument to offer regarding what function **dj**(*r*) produces the reason-strength corresponding to *r*. The only way to determine this is to look for proposals that work plausibly in concrete examples. Perhaps the most illuminating example is that of the biased lotteries diagrammed in figures 2 and 3. Suppose reason-strengths could be identified with the corresponding probabilities. In lottery 2, the probability of ticket 1 being drawn is .000001, and the probability of any other ticket being drawn is .111111. We wanted to conclude in this case that we are justified in believing that ticket 1 will not be drawn. The probability corresponding to the argument-strength for this conclusion is .999999, however the probability corresponding to the counter-argument for the conclusion that ticket 1 will be drawn (because no other ticket will) is .888889. The difference between these probabilities is .11111, which is a very low probability. If probabilities and degrees of justification could be identified, i.e., **dj**(*r*) = *r*, this would produce too low a degree of justification for it to be reasonable to believe that ticket 1 will not be drawn. So apparently we cannot compute degrees of justification by adding and subtracting probabilities.

There is statistical lore suggesting that in probabilistic reasoning degrees of justification can be compared in terms of likelihood ratios.[19] When (as in the biased lotteries) we have an argument for *P* based on a probability *r*, and a weaker argument for **~***P* based on a probability *r**, the likelihood ratio is $(1 - r)/(1 - r^*)$. The suggestion is that the degree of justification for **~***P* is determined by the likelihood ratio. For example, in lottery 2 the likelihood ratio is .000009, while in lottery 3 it is .09. Note that likelihood ratios are defined so that higher likelihood ratios

[19] This is known as the likelihood principle. It is due to R. A. Fisher (1922), and versions of it have been endorsed by a variety of authors, including G. A. Barnard (1949 and 1966), Alan Birnbaum (1962), A. W. F. Edwards (1972), and Ian Hacking (1965).

correspond to lower degrees of justification. An equivalent but more intuitive way of measuring degrees of justification is by using the inverse of the likelihood ratios.

The reason likelihood ratios produce plausible comparisons in the case of the biased lotteries is that by looking at $1/(1-r)$, when the probabilities are close to 1 the differences in likelihood produced by small differences in probability are large. For example, $1/(1-.999999)$ is 1,000,000, but $1/(1-.999989) = 90{,}909$ and $1/(1-.888889) = 9$.

In my (1990) I argued that likelihood ratios seem to yield the intuitively correct answers in many cases of statistical and inductive reasoning, and on that basis I am prepared to tentatively endorse their use in measuring degrees of justification. If we take the degree of justification $\mathbf{dj}(r)$ resulting from an application of the statistical syllogism with probability r to be $\log(1/(1-r))$, then the result of subtracting degrees of justification is the same as taking the logarithm of the inverse of the likelihood ratio, i.e., $\mathbf{dj}(r) - \mathbf{dj}(r^*) = \log(1/(1-r)) - \log(1/(1-r^*)) = \log((1-r^*)/(1-r))$.

However, this definition of **dj** fails to satisfy the obvious constraint that for $r \leq .5$, $\mathbf{dj}(r) = 0$. Furthermore, for $r > .5$, $\mathbf{dj}(r)$ should approach 0 as r approaches .5. In my (2001) I proposed to handle this by defining $\mathbf{dj}(r) = \log(.5) - \log(1-r)$ (for $r \geq .5$). But I have subsequently come to realize that this way of normalizing degrees of justification does not work. Degrees of justification are supposed to correspond to probabilities in the sense that if d is any possible degree of justification then there is an r between .5 and 1 such that $d = \mathbf{dj}(r)$. Furthermore, if d and d^* are possible degrees of justification and $d \geq d^*$ then $d - d^*$ should also be a possible degree of justification. Thus if $x \geq y$ and x and y are between .5 and 1, then there should be an r between .5 and 1 such that $\mathbf{dj}(x) - \mathbf{dj}(y) = \mathbf{dj}(r)$. If we consider the lottery 2, this would require that

$$\log(.5) - \log(1-r) = \mathbf{dj}(.999999) - \mathbf{dj}(.999989) = -1.0414$$

and hence $r = -4.5$, which is not an allowable value.

I propose instead that we normalize $\mathbf{dj}(r)$ by defining:

$$\mathbf{dj}(r) = \begin{cases} \log(1/(1-r) - 1) \text{ for } r \geq .5 \\ 0 \text{ for } r < .5 \end{cases}$$

Equivalently, $\mathbf{dj}(r) = \log(r/(1-r))$ for $r \geq .5$. This has the consequence that $\mathbf{dj}(.5) = 0$, and it has the automatic consequence that when x and y are between .5 and 1 and $x \geq y$ then there is an r between .5 and 1 such that $\mathbf{dj}(r) = \mathbf{dj}(x) - \mathbf{dj}(y)$.

When r and r^* are close to 1, $\mathbf{dj}(r) - \mathbf{dj}(r^*)$ is approximately equal to $\log((1-r^*)/(1-r))$, so this definition has the effect of evaluating comparative reason-strengths in terms of likelihood ratios. On the other hand, if r and r^* are closer to .5 then the comparison can diverge from likelihood ratios. However, this seems to be required to make the normalization work, and the values of $\mathbf{dj}(r)$ produced by this definition are intuitively reasonable. For example, if we consider lottery 2,

$$\mathbf{dj}(.999999) - \mathbf{dj}(.888889) = 5.0969 = \mathbf{dj}(.999992).$$

This accords with our intuition that we are justified in concluding that ticket 1 will not be drawn. It is analogous to concluding (defeasibly) that something having a 1 in 200,000 chance of occurring will not occur. When we turn to lottery 3,

dj(.999999) – **dj**(.999989) = 1.0414 = **dj**(.9167).

This accords with our intuition that we are no longer justified in concluding that ticket 1 will not be drawn — if something has a 1 in 10 chance of occurring, we cannot reasonably conclude (even defeasibly) that it will definitely not occur.

An important consequence of this definition of **dj**(*r*) is that computations of the values of linear combinations of degrees of justification can be done entirely "inside" the logarithms. For example, $\mathbf{dj}(x) + \mathbf{dj}(y) - \mathbf{dj}(z) = \log([x/(1-x)]\cdot[y/(1-y)]\cdot[(1-z)/z])$. Thus comparisons of the values of linear combinations are independent of the base of the logarithm.

My proposal for calibrating degrees of justification is thus that we employ **dj**:

> If *X* is a defeasible reason for *p*, the strength of this reason is $\log(r/(1-r))$ where *r* is that real number such that an argument for **~***p* based upon the suppositions "$\text{prob}(F/G) \geq r$ & *Gc*" and "(*p* ↔ **~***Fc*)" and employing the statistical syllogism exactly counteracts the argument for *p* based upon the supposition *X*.

Note that $\mathbf{dj}(1) = \infty$. That is, the strongest reasons have infinite reason-strength. This could create problems if we ever wanted to subtract the strengths of such arguments from each other, because $\infty - \infty$ is undefined, but in fact we will never have occasion to do that.

10. Simplifying the Computation

Principles (DJ1) and (DJ2) provide a recursive characterization of degrees of justification relative to an inference-graph. However, this characterization does not lend itself well to implementation because it requires the construction of modified inference-graphs. The objective of this section is to produce an equivalent recursive characterization that appeals only to the given inference-graph.

Recall that a defeat-link or support-link extends from its *root* to its *target*. The root of a defeat-link is a single node, and the root of a support-link is a set of nodes. Let us define precisely:

> An *inference/defeat-path* from a node φ to a node θ is a sequence of support-links and defeat-links such that (1) if the first link is a defeat-link, its root is φ; and if it is a support-link then φ is a member of its root; (2) the target of the last link in the path is θ; (3) the last link in the path is a defeat-link; (4) the root of each defeat-link after the first member of the path is the target of the preceding link; (5) some member of the basis of each support-link after the first member of the path is the target of the preceding link; and (6) the path does not contain an internal loop, i.e., no two links in the path have the same root.

> θ is φ-*dependent* iff there is an inference / defeat-path from φ to θ.

We defined a defeat-link to be φ–critical in an inference-graph *G* iff it is a member of some

minimal set of defeat-links such that the removal of all the links in the set results in φ no longer having a φ–dependent defeater in *G*. A necessary condition for a defeat-link *L* to be φ–critical is that it lie on an inference/defeat-path from φ to φ.. But that is not a sufficient condition. In general, there can be diverging and reconverging paths with several "parallel" defeat-links, as in figure 5. The set of parallel defeat-links may then be a minimal set of defeat-links such that the removal of all the links in the set results in φ no longer having a φ–dependent defeater in *G*, in which case the defeat-links are all φ–critical. The only way in which a defeat-link on an inference/defeat-path can fail to be φ–critical is when there is a path around it consisting entirely of support-links, as diagrammed in figure 6. This is what happens in inference-graph (7), and it is crucial to the computation of degrees of justification that such defeat-links not be regarded as φ-critical.

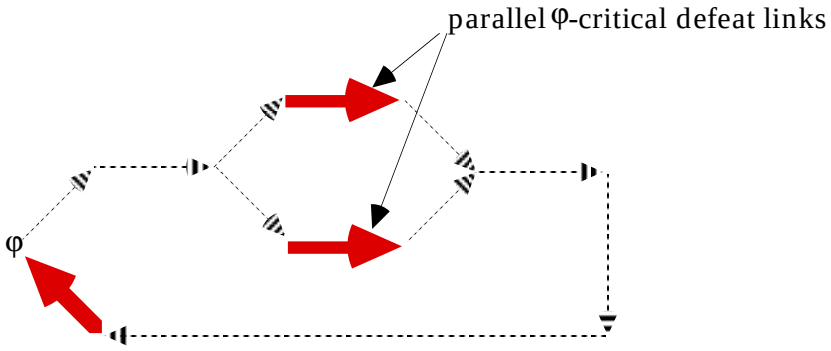


Figure 5. Parallel φ-critical defeat-links

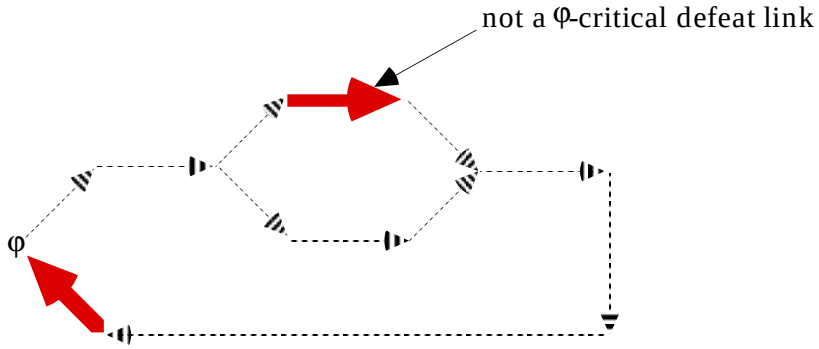


Figure 6. Defeat link that is not φ-critical

Let us say that a node α *precedes* a node β on an inference/defeat-path iff α and β both lie on the path and either α = β or the path contains a subpath originating on α and terminating on β. *Inference-descendants* of a node are nodes that can be reached by following support-links. We can characterize the φ-critical defeat-links as follows:

A defeat-link *L* is φ-critical in *G* iff (1) *L* lies on an inference/defeat-path σ in *G* from φ to φ, and (2) there is no node α preceding the root of *L* on σ and node β preceded by the target of *L*

on σ such that β is an inference-descendant of α in *G*.

This characterization makes it straightforward to search for φ-critical defeat-links in an inference-graph.

Recall that (DJ) was formulated as follows:

(DJ) If *P* is inferred from the basis $\{B_1,...,B_n\}$ in an inference-graph *G* in accordance with a reason-scheme of strength ρ, $D_1,...,D_k$ are the *P*-independent defeaters for *P*, $D_{k+1},...,D_m$ are the *P*-dependent defeaters of *P*, and G_P results from deleting all *P*-critical defeat-links from *G* and making all *P*-independent nodes φ initial with $j(\varphi,G_P) = j(\varphi,G)$, then $j(P,G) = \min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim [\max\{j(D_1,G),...,j(D_k,G)\} + \max\{j(D_{k+1},G_P),...,j(D_m,G_P)\}]$.

In most of the cases we have considered, G_φ is an inference-graph in which no node θ has a θ-dependent defeater. Thus $j(\theta,G_\varphi)$ can be computed using just (DJ1) in G_φ. This is equivalent to applying (DJ1) in *G* but ignoring φ-critical defeat-links:

> If ψ is inferred from the basis $\{B_1,...,B_n\}$ in the inference-graph *G* in accordance with a reason-scheme of strength ρ, $D_1,...,D_k$ are the defeaters for ψ in *G* that are not φ-critical, and no D_i is ψ-dependent, then
> $j(\psi,G_\varphi) = \min\{\rho,j(B_1,G),...,j(B_n,G)\} \sim \max\{j(D_1,G),...,j(D_k,G)\}$.

So for this computation, it is unnecessary to modify the inference-graph.

In those cases in which there is a node θ having a θ-dependent defeater in G_φ, we must instead apply (DJ2):

> If ψ has ψ-dependent defeaters $D_1,...,D_k$ in G_φ and $G_{\psi,\varphi}$ results from deleting all ψ-critical defeat-links from G_φ and making all ψ-independent nodes θ initial with $j(\theta,G_{\psi,\varphi}) = j(\theta,G_\varphi)$, then
> $j(\psi,G_\varphi) = j(\psi,G_{\psi,\varphi}) \sim \max\{j(D_1,G_{\psi,\varphi}),...,j(D_k,G_{\psi,\varphi})\}$.

Applying this computation recursively leads to a sequence of inference-graphs of the form $G_{\varphi_1,...,\varphi_n}$. The inference-graph *G* is finite because it represents the actual state of an agent's reasoning. It follows that the computation eventually terminates because the inference-graph *G* has only finitely many defeat-links, and each subsequent $G_{\varphi_1,...,\varphi_n}$ has fewer defeat-links than its predecessor. The sequence of inference-graphs constructed in this way can be characterized as follows. Where $\varphi_1,...,\varphi_n$ are nodes of an inference-graph *G*, define recursively:

> G_φ results from deleting all φ-critical defeat-links from *G* and making all nodes θ that are φ-independent in *G* initial with $j(\theta,G_\varphi) = j(\theta,G)$;
>
> $G_{\varphi_1,...,\varphi_n}$ results from deleting all φ_1-critical defeat-links from $G_{\varphi_2,...,\varphi_n}$ and making all nodes θ that are φ_1-independent in $G_{\varphi_2,...,\varphi_n}$ initial with $j(\theta,G_{\varphi_1,...,\varphi_n}) = j(\theta,G_{\varphi_2,...,\varphi_n})$.

To reformulate the recursion so as to avoid constructing modified inference-graphs, we define

some new concepts. Where $\varphi_1,...,\varphi_n$ are nodes of an inference-graph G, define recursively:

> A defeat-link δ of G is $\langle\varphi_1,...,\varphi_n\rangle$-*critical* in G iff (1) δ lies on an inference / defeat-path μ in G from φ to φ containing no $\langle\varphi_2,...,\varphi_n\rangle$-critical defeat-links and (2) there is no node α preceding the root of δ on μ and node β preceded by the target of δ on μ such that β is an inference-descendant of α in G.
>
> A defeat-link δ of G is *hereditarily*-$\langle\varphi_1,...,\varphi_n\rangle$-*critical* in G iff either δ is $\langle\varphi_1,...,\varphi_n\rangle$-*critical* in G or δ is hereditarily-$\langle\varphi_2,...,\varphi_n\rangle$-*critical* in G.
>
> A node (defeater) of G is *hereditarily*-$\langle\varphi_1,...,\varphi_n\rangle$-*critical* in G iff it is the root of a hereditarily-$\langle\varphi_1,...,\varphi_n\rangle$-critical defeat-link in G.

Obviously:

Theorem 3: δ is hereditarily-$\langle\varphi_1,...,\varphi_n\rangle$-critical in G iff δ is φ_1-critical in $G_{\varphi_2,...,\varphi_n}$ or φ_2-critical in $G_{\varphi_2,...,\varphi_n}$ or ... or φ_n-critical in G_{φ_n}.

A defeat-link that is φ_i-critical in $G_{\varphi_i,...,\varphi_n}$ does not exist in $G_{\varphi_j,...,\varphi_n}$ for $j > i$, so:

Theorem 4: δ is φ_1-critical in $G_{\varphi_2,...,\varphi_n}$ iff δ is $\langle\varphi_1,...,\varphi_n\rangle$-critical in G but not $\langle\varphi_2,...,\varphi_n\rangle$-critical in G.

Where $\theta,\varphi_1,...,\varphi_n$ are nodes of an inference-graph G, define:

> θ is $\langle\varphi\rangle$-*independent* of ψ in G iff there is no inference / defeat-path in G from ψ to θ containing a φ-critical defeat-link.
>
> θ is $\langle\varphi_1,...,\varphi_n\rangle$-*independent* of ψ in G iff there is no inference / defeat-path in G from ψ to θ containing a hereditarily-$\langle\varphi_1,...,\varphi_n\rangle$-critical defeat-link.

Theorem 5: θ is $\langle\varphi_1,...,\varphi_n\rangle$-*independent* of ψ in G iff θ is ψ-independent in $G_{\varphi_1,...,\varphi_n}$.

Where ψ is an initial node in G, let $j_0(\psi,G)$ be its assigned value. Let us define recursively:

Definition:

(a) If ψ is initial in G then $j_{\varphi_1,...,\varphi_n}(\psi,G) = j_0(\psi,G)$;

(b) If ψ is inferred from the basis $\{B_1,...,B_n\}$ in an inference-graph G in accordance with a reason-scheme of strength ρ, $D_1,...,D_k$ are the defeaters for ψ that are $\langle\varphi_1,...,\varphi_n\rangle$-independent of ψ in G and are not $\langle\varphi_1,...,\varphi_n\rangle$-critical in G, and $D_{k+1},...,D_m$ are the defeaters for ψ that are $\langle\varphi_1,...,\varphi_n\rangle$-dependent on ψ in G and are not hereditarily-$\langle\varphi_1,...,\varphi_n\rangle$-critical in G, then

$$j_{\varphi_1,...,\varphi_n}(\psi,G) = \min\{\rho, j_{\varphi_1,...,\varphi_n}(B_1,G),...,j_{\varphi_1,...,\varphi_n}(B_n,G)\}$$
$$\sim [\max\{j_{\varphi_1,...,\varphi_n}(D_1,G),...,j_{\varphi_1,...,\varphi_n}(D_k,G)\}$$
$$+ \max\{j_{\psi,\varphi_1,...,\varphi_n}(D_{k+1},G),...,j_{\psi,\varphi_1,...,\varphi_n}(D_m,G)\}].$$

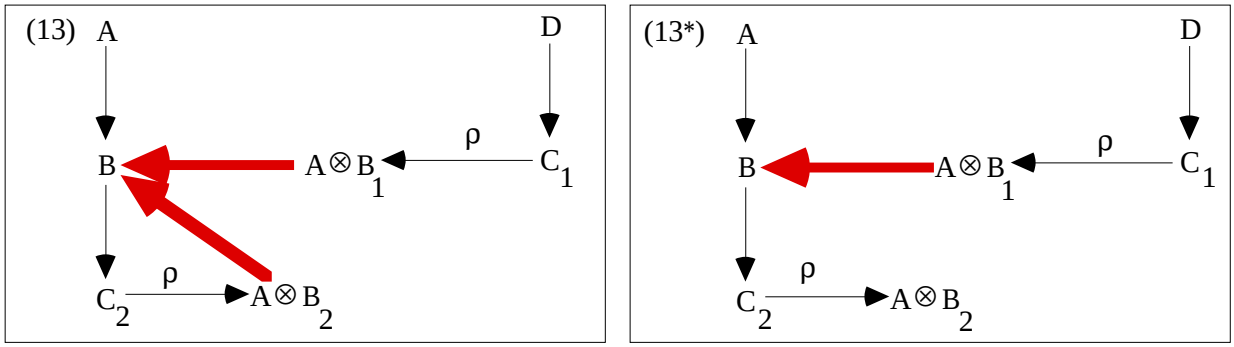
The reason this is a recursive definition is that we always reach an n at which there are no more $\langle\varphi_1,...,\varphi_n\rangle$-dependent defeaters, and then the values of all nodes are computed recursively in terms of the values assigned to initial nodes.

It is now trivial to prove by induction:

Theorem 6: $j(\psi,G_{\varphi_1,...,\varphi_n}) = j_{\varphi_1,...,\varphi_n}(\psi,G)$.

Thus we have a recursive definition of the degree of justification of a node that that computes the degrees of justification entirely by reference to the given inference-graph rather than by building a sequence of modified inference-graphs in accordance with the original analysis.

To illustrate the computation provided by theorem 6, consider inference-graph (13). Here we have two separate arguments for C, from which $A\otimes B$ is inferred with reason-strength ρ. Nodes supporting the same conclusions are distinguished by subscripting the conclusions. To compute $j(B,G_{13})$ using (DJ1) and (DJ2), we remove the B-critical link to get (13*). Then



$j(B,G_{13}) = j_0(A,G_{13}) \sim [j(A\otimes B_1,G_{13}) + j(A\otimes B_2,G_{13^*})]$.

$j(A\otimes B_1,G_{13}) = \min\{\rho,j(C_1,G_{13})\} = \min\{\rho,j_0(D,G_{13})\}$. $j(A\otimes B_2,G_{13^*}) = \min\{\rho,j(C_2,G_{13^*})\} = \min\{\rho,j(B,G_{13^*})\} = \min\{\rho, (j_0(A,G_{13}) \sim \min\{\rho,j_0(D,G_{13})\})\}$.

Thus

$j(B,G_{13}) = j_0(A,G_{13}) \sim [\min\{\rho,j_0(D,G_{13})\} + \min\{\rho, (j_0(A,G_{13}) \sim \min\{\rho,j_0(D,G_{13})\})\}]$.

For example, if $j_0(A,G_{13}) = j_0(D,G_{13})$, and $\rho < .5\cdot j_0(A,G_{13})$. Then $j(B) = j_0(A,G_{13}) \sim 2\rho$.

To repeat the computation using theorem 6,

$j(B,G_{13}) = j_0(A,G_{13}) \sim [j(A\otimes B_1,G_{13}) + j_B(A\otimes B_2,G_{13})]$.

$j(A\otimes B_1,G_{13}) = \min\{\rho,j(C_1,G_{13})\} = \min\{\rho,j_0(D,G_{13})\}$.

$j_B(A\otimes B_2,G_{13}) = \min\{\rho,j_B(C_2,G_{13})\} = \min\{\rho,j_B(B,G_{13})\}$.

$A\otimes B_1$ is the only B-independent defeater of B, and it is not B-critical. Thus

$j_B(B,G_{13}) = j_B(A,G_{13}) \sim j_B(A \otimes B_1,G_{13}) = j_0(A,G_{13}) \sim j(A \otimes B_1,G_{13}) = j_0(A,G_{13}) \sim \min\{\rho,j_0(D,G_{13})\}$.

Therefore, $j_B(A \otimes B_2,G_{13}) = \min\{\rho, j_0(A,G_{13}) \sim \min\{\rho,j_0(D,G_{13})\}\}$.

Hence once again,

$j(B,G_{13}) = j_0(A,G_{13}) \sim [\min\{\rho,j_0(D,G_{13})\} + \min\{\rho, j_0(A,G_{13}) \sim \min\{\rho,j_0(D,G_{13})\}\}]$.

Theorem 6 constitutes the desired recursive characterization of $j(\varphi,G)$. This could be implemented straightforwardly. However, as the next section shows, further improvements are possible.

11. Inference-Hypergraphs

11.1 Two Kinds of Inference-Graphs

In the interest of theoretical clarity, inference-graphs were defined in such a way that different arguments for the same conclusion are represented by different nodes. This made it clearer how the algorithm for computing defeat status works. However, for the purpose of implementing defeasible reasoning, using different nodes to represent different arguments for the same conclusion is an inefficient representation, because it leads to needless duplication. If we have two arguments supporting a single conclusion, then any further reasoning from that conclusion will generate two different nodes. If we have two arguments for each of two conclusions, and another inference proceeds from those two conclusions, the latter will have to be represented by four different nodes in the inference-graph, and so on. This is illustrated in figure 7, where *P* and *Q* are each inferred in two separate ways, and then *R* is inferred from *P* and *Q*.

A more efficient representation of reasoning would take the inference-graph to be a hypergraph rather than a standard graph. In a hypergraph, nodes are linked to *sets* of nodes rather than individual nodes.[20] This is represented diagramatically by connecting the links with arcs. In an inference-hypergraph, when we have multiple arguments for a conclusion, the single node representing that conclusion will be tied to different bases by separate groups of links. This is illustrated in figure 8 by an inference-hypergraph encoding the same reasoning as the standard inference-graph in figure 7. In an inference-hypergraph, a support-link will be represented by a *set* of supporting-arrows connected with an arc.

---

[20] This observation derives from my (1994). This section adapts the approach taken there to the new definition of "status assignment". I take the term "hypergraph" from Pearl (1988), who observes that Bayesian nets are hypergraphs. I previously called hypergraphs "and/or graphs".

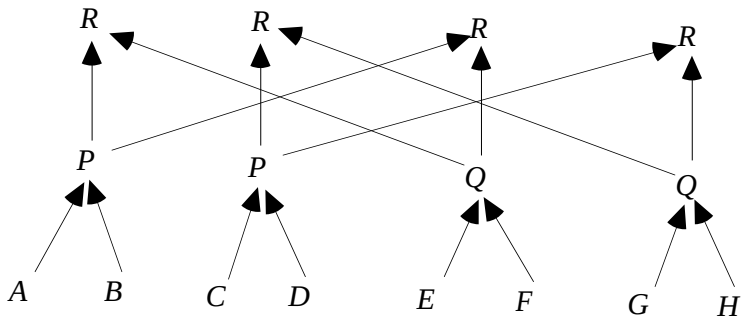


Figure 7. Inference-graph with multiple arguments for a single conclusion.

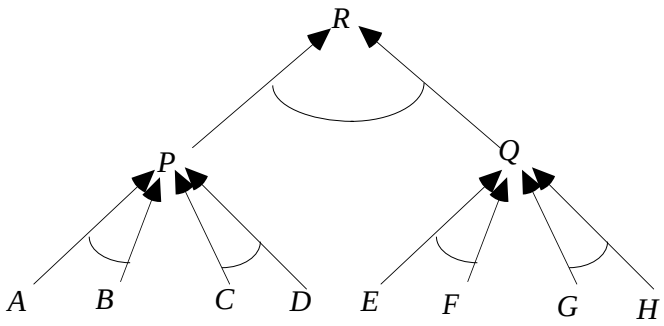


Figure 8. An inference-hypergraph

Although hypergraphs provide an efficient representation of reasoning, they complicate the computation of degrees of justification. Using simple inference-graphs, we can use principles (DJ1) and (DJ2) to compute degrees of justification. If we are to use hypergraphs in the implementation, we must find a computation for hypergraphs that is equivalent to that for simple inference-graphs. A simple inference-graph can be rewritten as a hypergraph in which each node of the hypergraph corresponds to a set of nodes of the simple graph. A node of the simple inference-graph corresponds to an *argument* in the hypergraph. An argument is a kind of connected sub-tree of the graph. More precisely:

> An *argument in* an inference-hypergraph G *for* a node φ is a minimal subset A of the nodes and support-links of the graph such that (1) if a node in A has any support-links in G, exactly one of them is in A, (2) if a support-link is in A then the nodes in its support-link-basis are also in A, and (3) φ is in A.

Nodes in the simple inference-graph correspond one-one to arguments in the inference-hypergraph.

11.2 Computing Degrees of Justification

It was argued in section five that the degree of justification of a conclusion should be the maximum of the degrees of justification supplied by the different arguments supporting that

conclusion. Thus the result we want for inference-hypergraphs is the following *Correspondence Principle*:

> The degree of justification of a node of the inference-hypergraph is equal to the maximum of the degrees of justification of the corresponding nodes of the simple inference-graph.

To accomplish this, it helps to assign degrees of justification to support-links as well as nodes. Some nodes will be initial, in which case they are simply assigned a degree of justification $j_0(P,G)$ in an inference-graph G. I assume that such nodes have no support-links. If a node is not initial, its degree of justification is the maximum of the degrees of justification of its support-links. If a node is not initial and it has no support-links, we stipulate that its degree of justification is zero. This cannot happen naturally, but certain constructions that occur in the computation of degrees of justification can produce inference-graphs having such nodes. Having characterized the degree of justification of a node of the inference-graph in terms of the degrees of justification of its support-links, the remaining problem is how to compute the degrees of justification of support-links.

In an inference-hypergraph, undercutting defeat-links attach to support-links rather than nodes. A rebutting defeat-link could be regarded as attaching to either a node or to all defeasible support-links for the node. It will be more convenient to adopt the latter convention so that both undercutting and rebutting defeaters can be treated uniformly. Where λ is a support-link, let $\otimes\lambda$ be its undercutting defeater and let $\sim\lambda$ be its rebutting defeater. That is, if the root of λ is $\{B_1,...,B_n\}$ and the target is φ, $\otimes\lambda$ is $[(B_1\&...\&B_n) \otimes \varphi]$ and $\sim\lambda$ is $\sim\varphi$.

In inference-hypergraphs, we can define inference/defeat-paths pretty much as we did for simple inference-graphs, except that they are paths from support-links to support-links rather than from nodes to nodes:

> An *inference/defeat-path* from a support-link γ to a support-link λ in an inference-hypergraph HG is a sequence of support-links and defeat-links in HG such that (1) the first link is γ; (2) the last link in the path is a defeat-link and its target is λ; (3) the root of each defeat-link after the first member of the path is the target of the preceding link; (4) some member of the basis of of each support-link after the first member of the path is either the target of the preceding link or has a support-link that is the target of the preceding link; and (5) the path does not contain an internal loop, i.e., no two links in the path have the same root.

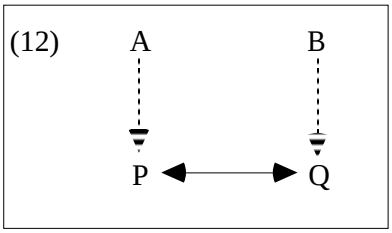
Let us say that a simple graph *corresponds* to a hypergraph iff the corresponding nodes support the same conclusions and the support-links and defeat-links are analogous. Representing links by ordered pairs, we can define precisely:

> A simple inference-graph G *corresponds to* an inference-hypergraph HG iff there is a function μ mapping the nodes of G onto the nodes of HG such that:
> (a) if α is a node of G and β is a node of HG, $\mu(\alpha) = \mu(\beta)$ iff α and β support the same proposition;
> (b) if α is a node of G and $\Gamma = \{\beta \mid \langle\beta,\alpha\rangle$ is a support-link for a in $G\}$ then $\langle\{\mu(\gamma) \mid \gamma\in\Gamma\},\mu(\alpha)\rangle$ is a support-link in HG;

(c) if $\langle\Gamma,\alpha\rangle$ is a support-link in $\mathcal{H}G$ then for all nodes β,γ in G, if $\mu(\beta)\in\Gamma$ and $\mu(\gamma) = \alpha$ then $\langle\beta,\gamma\rangle$ is a support-link in G;
(d) $\langle\beta,\alpha\rangle$ is a defeat-link in G iff, if $\langle\gamma_1,\alpha\rangle,...,\langle\gamma_n,\alpha\rangle$ are the support-links for α in G then $\langle\mu(\beta),\langle\{\mu(\gamma_1),...,\mu(\gamma_n)\},\mu(\alpha)\rangle\rangle$ is a defeat-link in $\mathcal{H}G$.

Then given an inference / defeat-path in the simple graph, the corresponding sequence of links in the hypergraph is a corresponding inference / defeat-path in the hypergraph. Furthermore, given an inference / defeat-path in the hypergraph, every corresponding path in the simple graph is an inference / defeat-path.

(12)



Inference-paths must be defined somewhat differently in inference-hypergraphs than in simple inference-graphs. The difficulty is that if we simply define inference-paths to be sets of linked support-links, they can be circular. E.g., suppose *P* and *Q* are logically equivalent. Then if we have independent reasons for *P* and for *Q*, we can derive each from the other. This makes perfectly good sense. If the independent argument for *P* is subsequently defeated, the argument that derives *P* from *Q* will still support *P* as long as the argument supporting *Q* is not defeated, and vice versa. Thus we can have an inference-hypergraph like (12) "inference-loops". These are inference-hypergraphs in which some node is an inference-descendant of itself (where the inference-descendants of a node are those nodes that can be reached by following sequences of support-links). In this inference-graph, we do not want to count paths like $A \rightarrow P \rightarrow Q \rightarrow P$ as inference-paths. We can handle this by defining:

> An *inference-path* from a node φ to a node θ is a sequence of support-links such that (1) φ is a member of the root of the first link, (2) the target of the last is θ, (3) some member of each support-link after the first is the target of the preceding link, and (4) the path is non-circular, i.e., $\varphi \neq \theta$ and no two links in the path have the same root.

Then we can define, for hypergraphs, where λ is a support-link:

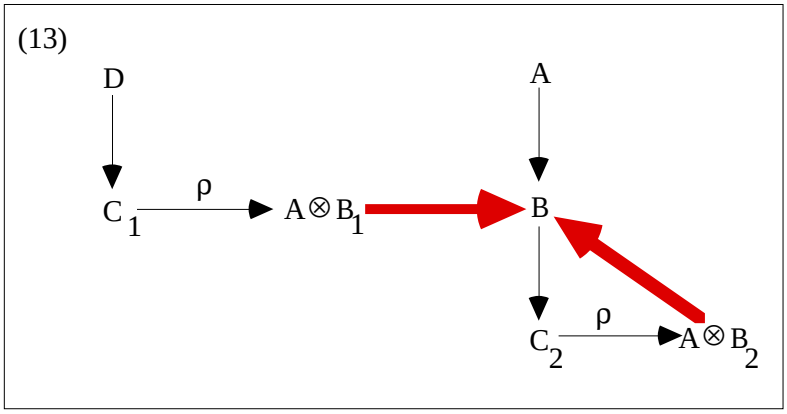
> A defeat-link δ is λ-*critical* in G iff (1) δ lies on an inference / defeat-path μ in G from λ to λ, and (2) there is no node α preceding the root of δ on μ and node β preceded by the target of the target of δ on μ and an inference-path ν from α to β such that the path resulting from splicing ν into μ in place of the path from α to β is still an inference / defeat-path.

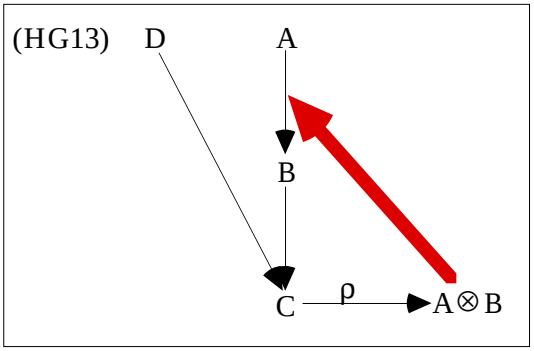
The somewhat simpler definition (in terms of inference-descendants) employed in simple inference-graphs still works. The path resulting from splicing σ^* into σ may contain an internal loop, but if

we simply omit the loop we then have an inference-path going around the defeat-link:

Theorem 7: A defeat-link δ is λ-critical in *G* iff (1) δ lies on an inference / defeat-path μ in *G* from λ to λ, and (2) there is no node α preceding the root of δ on μ and node β preceded by the target of the target of δ on μ such that β is an inference-descendant of α.

Relating inference-hypergraphs to simple inference-graphs is complicated by the fact that a defeat-link in an inference-hypergraph can correspond to several different defeat-links in the simple inference-graph. Consider the simple inference-graph (13) again and the corresponding inference-hypergraph (*HG*13). Here ρ is the reason-strength for the inference from *C* to *A*⊗*B*. The defeat-link ⟨*A*⊗*B*,*B*⟩ in (*HG*13) corresponds to the two defeat-links ⟨*A*⊗B_1,*B*⟩ and ⟨*A*⊗B_2,*B*⟩ in (13). If a defeat-link δ is λ-critical in the hypergraph and φ is the target of λ, then *some* corresponding defeat-link is φ-critical in the simple graph. More precisely, if δ is λ-critical in the hypergraph by virtue of lying on the inference / defeat-path σ, then for every corresponding inference / defeat-path Σ in the simple graph, the corresponding defeat-link in Σ is φ-critical in the simple graph. However, there can also be corresponding defeat-links in the simple graph that do not lie on inference / defeat-paths corresponding to σ. For example, writing the support-link from *A* to *B* as an ordered pair, ⟨*A*⊗*B*,*B*⟩ is ⟨*A*,*B*⟩-critical in (*HG*13), but only ⟨*A*⊗B_2,*B*⟩, not ⟨*A*⊗B_1,*B*⟩, is *B*-critical in (13).



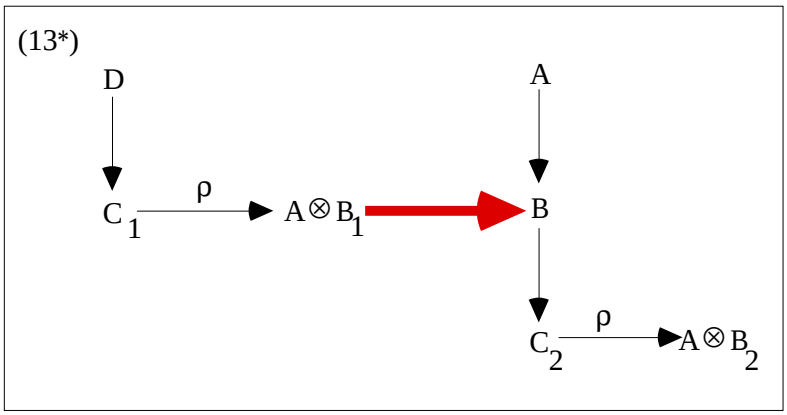


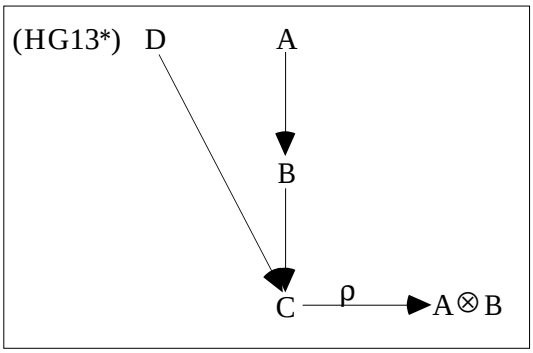
Our objective is to find a way of computing degrees of justification in the inference-hypergraph that agrees with the computation in the simple inference-graph. To compute $j(B,G_{13})$ we removed the *B*-critical link to get (13*), and then computed that $j(B,G_{13}) = j(B,G_{13*}) \sim j(A\otimes B_2,G_{13*}) = j(A,G_{13})$

~ [min{ρ,$j(D,G_{13})$} + min{ρ, ($j(A,G_{13})$ ~ min{ρ,$j(D,G_{13})$})}]. For example, if $j(A,G_{13}) = j(D,G_{13})$, and ρ < .5·$j(A,G_{13})$, then $j(B,G_{13}) = j(A,G_{13})$ ~ 2ρ. It might be supposed that we can compute degrees of justification analogously in the inference-hypergraph by deleting ⟨*A*,*B*⟩-critical defeat-links. However, this does not work. Removing ⟨*A*,*B*⟩-critical defeat-links would produce inference-graph (*HG*13*). The computation analogous to that employed in (13) would yield

$$\begin{aligned} j(B,HG13) &= j(B,HG13^*) \sim j(A\otimes B,HG13^*) \\ &= j(A,HG13) \sim \min\{\rho, j(C,HG13^*)\} \\ &= j(A,HG13) \sim \min\{\rho,\max\{j(A,HG13),j(D,HG13)\}\}. \end{aligned}$$

If ρ < $j(A,HG13) = j(D,HG13)$ then $j(B,HG13) = j(A,HG13)$ ~ ρ. So this does not validate the Correspondence Principle.

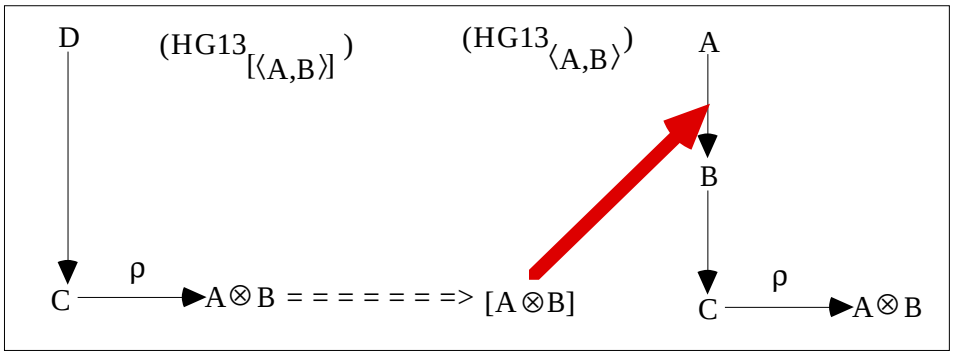




This illustrates a general point. In an inference-hypergraph *HG*, if a defeat-link for a support-link λ with target *P* has multiple arguments supporting it, some λ-dependent and others not, then it represents both λ-dependent and λ-independent defeaters in the corresponding simple inference-graph *G*, and so by (DJ) it will be counted twice in computing degrees of justification in *G*. The computation for *HG* must work similarly. We must compute both a λ-dependent and a λ-independent degree of justification for the defeat-link, and then subtract their sum from the argument-strength for the basis of λ. The λ-independent degree of justification for a defeater *D* in *HG* should be the maximum degree of justification of *P*-independent defeaters corresponding to *D* in *G*, and the λ-dependent degree of justification for a defeater *D* in *HG* should be the maximum degree of justification in G_P of *P*-dependent defeaters corresponding to *D* in *G*. The

Correspondence Principle will then follow.

Using simple inference-graphs we were able to compute the *P*-independent and *P*-dependent values for nodes by modifying the inference-graph. The *P*-independent values were computed in the original inference-graph *G*, and then the *P*-dependent values were computed in the inference-graph G_P. However, the analogous strategy does not work for inference-hypergraphs. For example, there is no way to represent the simple inference-graph (13*) as an inference-hypergraph. The best we can do is construct (*HG*13), but that corresponds to (13), not (13*). This is because in an inference-hypergraph, there is no way to remove just one of the two defeat-links between *A*⊗*B* and *B*. To get around this difficulty, we must construct "hybrid" inference-graphs which are inference-hypergraphs except that we allow undercutting defeaters to be represented by two separate nodes, one having the *P*-independent value and the other having the *P*-dependent value. Let me explain.



We can compute λ-independent values for nodes by deleting arguments containing support-links having λ-dependent defeaters. Thus in inference-graph (*HG*13), the ⟨*A*,*B*⟩-independent strength of *A*⊗*B* can be computed by removing from the inference-graph all support-links having ⟨*A*,*B*⟩-dependent defeaters, producing ($HG13_{[\langle A,B\rangle]}$). To compute the ⟨*A*,*B*⟩-dependent strength of *A*⊗*B*, we use the ⟨*A*,*B*⟩-independent strength to attenuate the strength of the support-link ⟨*A*,*B*⟩, as in ($HG13_{\langle A,B\rangle}$). The latter inference-graph is constructed from (*HG*13) by (1) removing ⟨*A*,*B*⟩-independent arguments for *A*⊗*B*, (2) removing ⟨*A*,*B*⟩-critical defeat-links, and (3) adding a new node [*A*⊗*B*] that is treated as an initial node whose value is inherited from inference-graph ($HG13_{[\langle A,B\rangle]}$). By way of explanation, note that separate ⟨*A*,*B*⟩-independent arguments for *A*⊗*B* in (*HG*13) correspond to separate *B*-independent defeaters for *B* in the simple inference-graph (13), and the strength assigned to [*A*⊗*B*] is then the maximum strength of those *B*-independent defeaters for ⟨*A*,*B*⟩. The strength computed for *A*⊗*B* in ($HG13_{\langle A,B\rangle}$) will then be the maximum strength of the ⟨*A*,*B*⟩-dependent defeaters for ⟨*A*,*B*⟩. It then follows by (DJ) that $j(B,HG13) = j(A,HG13) \sim [j(A\otimes B,HG13_{[\langle A,B\rangle]}) + j(A\otimes B,HG13_{\langle A,B\rangle})]$. This should then be the value computed for $j(\langle A,B\rangle,HG13)$.

In constructing ($HG13_{[\langle A,B\rangle]}$) we remove ⟨*A*,*B*⟩-dependent arguments for *A*⊗*B*. This is done by removing support-links that occur in ⟨*A*,*B*⟩-dependent arguments but not also in ⟨*A*,*B*⟩-independent arguments. Let us say that a support-link is an ⟨*A*,*B*⟩-*link* iff it occurs in some ⟨*A*,*B*⟩-dependent argument. Precisely:

A *support-path* in an inference-hypergraph *HG* is an inference-path in *HG* from an initial

node to some other node.

A support-link γ is λ-independent in $\mathbf{HG}$ iff there is no inference-defeat path in $\mathbf{HG}$ from λ to γ.

A support-link is a λ-*link* in an inference-hypergraph $\mathbf{HG}$ iff it occurs in some support-path in $\mathbf{HG}$ every member of which lies on an inference / defeat-path μ to $\otimes\lambda$ and some member of μ is the target of a λ-dependent defeat-link.

A support-link is a $[\lambda]$-*link* in an inference-hypergraph $\mathbf{HG}$ iff it occurs in a support-path μ in $\mathbf{HG}$ every member of which lies on an inference / defeat-path to $\otimes\lambda$ no member of μ is the target of a λ-dependent defeat-link.

The $\langle A,B\rangle$-links are the support-links included in ($\mathbf{HG}13_{\langle A,B\rangle}$) and the $[\langle A,B\rangle]$-links are the support-links included in ($\mathbf{HG}13_{[\langle A,B\rangle]}$). So, for example $\langle C,A\otimes B\rangle$ is both a $[\langle A,B\rangle]$-link and a $\langle A,B\rangle$-link, but $\langle D,C\rangle$ is only a $[\langle A,B\rangle]$-link and $\langle A,B\rangle$ and $\langle B,C\rangle$ are only $\langle A,B\rangle$-links.

Where $\mathbf{HG}$ is an inference-hypergraph, and λ is a support-link in $\mathbf{HG}$, we can then define:

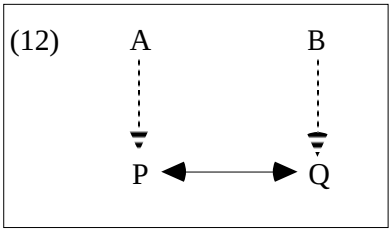
$\mathbf{HG}_{[\lambda]}$ results from (1) removing all support-links that are not $[\lambda]$-links from $\mathbf{HG}$, (2) removing all non-initial nodes having no support-link that is a $[\lambda]$-link, and (3) making all nodes θ that are λ-independent in $\mathbf{HG}$ initial with $j(\theta,\mathbf{HG}_{[\lambda]}) = j(\theta,\mathbf{HG})$.

$\mathbf{HG}_{\lambda}$ results from (1) removing all support-links that are not λ-links from $\mathbf{HG}$, (2) removing all non-initial nodes having no support-link that is a λ-link, (3) removing λ-critical defeat-links, (4) for each support-link γ that is the target of a λ-critical defeat-link, adding a node $\langle\otimes\gamma\rangle$ and making it the target of λ-links pointing to $\otimes\gamma$ in $\mathbf{HG}$; (4) making $\otimes\gamma$ initial with $j(\otimes\gamma,\mathbf{HG}_{\lambda}) = j(\otimes\gamma,\mathbf{HG}_{[\gamma]})$ and a defeat-link from $\otimes\gamma$ to γ, and (6) making all nodes θ that are λ-independent in $\mathbf{HG}$ initial with $j(\theta,\mathbf{HG}_{\lambda}) = j(\theta,\mathbf{HG})$.

We can then replicate the computation of degrees of justification in most simple inference-graphs by adopting the following principle for inference-hypergraphs that do not contain inference-loops:

(DJH) If λ is a support-link with basis $\{B_1,...,B_n\}$ and reason-strength ρ, then
$j(\lambda,\mathbf{HG}) = \min\{\rho,j(B_1,\mathbf{HG}),...,j(B_n,\mathbf{HG})\} \sim [j(\langle\otimes\lambda\rangle,\mathbf{HG}_{\lambda}) + \max\{j(\sim\lambda,\mathbf{HG}_{\lambda}), j(\otimes\lambda,\mathbf{HG}_{[\lambda]})\}]$.

(DJH) is to be understood so that if any of $\otimes\lambda$, $\sim\lambda$, or $\langle\otimes\lambda\rangle$ is not present in $\mathbf{HG}$ then their contributions are omitted from the summation. Assuming that the initial nodes in the inference-graph representing an agent's cognitive state cannot be defeaters, the dummy defeaters in the inference-graphs constructed by this computation can be detected easily by the fact that they are initial nodes in the constructed inference-graphs.



There is one case in which the computation of degrees of justification by (DJH) does not work. This occurs in an inference-hypergraph like (12) that contains an inference-loop. In such an inference-hypergraph, the computation described by (DJH) will not terminate. To compute $j(P,HG12)$ we would first have to compute $j(Q,HG12)$, and to compute $j(Q,HG12)$ we would first have to compute $j(P,HG12)$. For now, let us content ourselves with describing how to compute degrees of justification in inference-hypergraphs that do not contain inference-loops. In section twelve, I will extend the analysis to handle inference-loops.

To illustrate the use of (DJH), consider the inference-graph ($HG13$) again.

$j(B,HG13) = j(\langle A,B\rangle,HG13) = j_0(A,HG13) \sim [j(A\otimes B,HG13_{[\langle A,B\rangle]}) + j(\langle A\otimes B\rangle,HG13_{\langle A,B\rangle})]$.

$j(A\otimes B,HG13_{[\langle A,B\rangle]}) = j(\langle C,A\otimes B\rangle,HG13_{[\langle A,B\rangle]}) = \min\{\rho,j(C,HG13_{[\langle A,B\rangle]})\} = \min\{\rho,j(\langle D,C\rangle,HG13_{[\langle A,B\rangle]})\}$
$= \min\{\rho,j(D,HG13_{[\langle A,B\rangle]})\} = \min\{\rho,j(D,HG13)\} = \min\{\rho,j_0(D,HG13)\}$.

$j(\langle A\otimes B\rangle,HG13_{\langle A,B\rangle}) = j(\langle C,\langle A\otimes B\rangle\rangle,HG13_{\langle A,B\rangle}) = \min\{\rho,j(C,HG13_{\langle A,B\rangle})\}$
$= \min\{\rho,j(\langle B,C\rangle,HG13_{\langle A,B\rangle})\} = \min\{\rho,j(B,HG13_{\langle A,B\rangle})\} = \min\{\rho,j(\langle A,B\rangle,HG13_{\langle A,B\rangle})\}$.

$j(\langle A,B\rangle,HG13_{\langle A,B\rangle}) = j(A,HG13_{\langle A,B\rangle}) \sim j(A\otimes B,HG13_{\langle A,B\rangle}) = j(A,HG13_{\langle A,B\rangle}) \sim j(A\otimes B,HG13_{[\langle A,B\rangle]})$
$= j_0(A,HG13_{\langle A,B\rangle}) \sim \min\{\rho,j_0(D,HG13)\}$.

Thus

$j(B,HG13) = j_0(A,HG13) \sim [\min\{\rho,j_0(D,HG13)\} + \min\{\rho,[j_0(A,HG13_{\langle A,B\rangle}) \sim \min\{\rho,j_0(D,HG13)\}]\}]$.

11.3 Refining the Computation

To refine the recursive computation and make it more readily implementable we can more or less repeat the analysis given earlier for simple inference-graphs. Where σ is a sequence of support-links λ and bracketed support-links $[\lambda]$, let σ_1 be the first member of σ and let σ^* be the rest of σ (its cdr). Where $\sigma = \langle\sigma_1,...,\sigma_n\rangle$, let $\lambda\wedge\sigma = \langle\lambda,\sigma_1,...,\sigma_n\rangle$. Applying (DJH) repeatedly generates inference-graphs G_σ that can be characterized recursively. Where HG is an inference-hypergraph, λ is a support-link in HG, σ is a finite sequence of support-links and / or bracketed support-links, and φ is the target of λ:

$HG_\varnothing = HG$;

$HG_{[\lambda]\wedge\sigma}$ results from 1) removing all support-links that are not $[\lambda]$-links from HG_σ, (2) removing all nodes that are not initial in HG_σ and have no support-link that is a $[\lambda]$-link

in HG_σ, and (3) making all nodes θ that are λ-independent in HG_σ initial with $j(\theta,\mathit{HG}_{[\lambda]\wedge\sigma}) = j(\theta,\mathit{HG}_\sigma)$.

$\mathit{HG}_{\lambda\wedge\sigma}$ results from (1) removing all support-links that are not λ-links from HG_σ, (2) removing all nodes that are not initial in HG_σ and have no support-link that is a λ-link in HG_σ, (3) removing λ-critical defeat-links from HG_σ, (4) for each support-link γ that is the target of a λ-critical defeat-link in HG_σ, adding a node $\langle\otimes\gamma\rangle$ and making it the target of λ-links that point to $\otimes\gamma$ in HG_σ; (4) making $\otimes\gamma$ initial with $j(\otimes\gamma,\mathit{HG}_{\lambda\wedge\sigma}) = j(\otimes\gamma,\mathit{HG}_{[\gamma]\wedge\sigma})$ and a defeat-link from $\otimes\gamma$ to γ, and (6) making all nodes θ that are λ-independent in HG_σ initial with $j(\theta,\mathit{HG}_{\lambda\wedge\sigma}) = j(\theta,\mathit{HG}_\sigma)$.

To reformulate the recursion so as to avoid constructing modified inference-graphs, we define by simultaneous recursion:

A defeat-link δ of *HG* is σ-*critical* in *HG* iff (1) σ_1 is a support-link λ (not a bracketed support-link), (2) δ lies on a σ∗-inference / defeat-path μ in *HG* from λ to λ, and (3) there is no node α preceding the root of δ on μ and node β preceded by the target of the target of δ on μ such that there is a σ∗-support-path in *HG* from α to β.

A defeat-link δ of *HG* is *hereditarily-σ-critical* in *HG* iff $\sigma \neq \varnothing$ and (1) δ is σ-*critical* in *HG* or (2) δ is hereditarily-σ∗-*critical* in *HG*.

μ is a *σ–inference/defeat-path* in *HG* iff μ is an inferencedefeat-path in *HG* consisting entirely of σ-support-links and σ-defeat-links in *HG*.

A support-link *L* is a *σ-link* in *HG* iff either:

(1) $\sigma_1 = [\lambda]$ for some support-link λ in *HG* and *L* occurs in a σ∗-support-path in *HG* every member of which lies in a σ∗-inference / defeat-path to $\otimes\lambda$ in *HG* and no member of which is the target of a λ^σ∗-dependent defeat-link; or

(2) $\sigma_1 = \lambda$ for some support-link λ in *HG* and *L* occurs in a σ∗-support-path in *HG* every member of which lies in a σ∗-inference / defeat-path to $\otimes\lambda$ in *HG* and some member of which is the target of a λ^σ∗-dependent defeat-link

A defeat-link δ is a *σ-defeat-link* in *HG* iff either:

(1) $\sigma_1 = [\lambda]$ for some support-link λ in *HG* and the target of δ is a σ-link in *HG*; or

(2) $\sigma_1 = \lambda$ for some support-link λ in *HG* and the target of δ is a σ-link in *HG* and δ is not hereditarily-σ-critical in *HG*.

μ is a *σ-support-path* in *HG* iff μ is a support-path in *HG* and μ consists entirely of σ-support-links in *HG*.

A defeat-link or support-link γ is *σ-independent* of a support-link λ in *HG* iff there is no σ-inference / defeat-path from λ to γ in *HG*.

A node ψ is σ-*independent* of a support-link λ in HG iff every σ-link for ψ is σ-independent of λ in HG.

It is then trivial to prove by induction:

Theorem 8: If HG is an inference-hypergraph and σ is a finite sequence of support-links and bracketed support-links in HG then:

(1) a defeat-link L of HG is σ-critical in HG iff it is σ_1-critical in $HG_{\sigma*}$;
(2) μ is a σ–inference / defeat-path in HG iff μ is an inferencedefeat-path in HG_σ;
(3) L is a σ-link in HG iff L is a support-link in HG_σ;
(4) δ is a σ-defeat-link in HG iff δ is a defeat-link in HG_σ;
(5) μ is a σ-support-path in HG iff μ is a support-path in HG_σ;
(6) a defeat-link or support-link γ is σ-dependent in HG iff $\sigma_1 = \lambda$ for some support-link λ and γ is λ-dependent in $HG_{\sigma*}$.

We then define recursively:

Definition: If HG is an inference-hypergraph:

(a) If ψ is initial in HG then $j_\sigma(\psi,HG) = j_0(\psi,HG)$;
(b) If $\sigma_1 = \lambda$ or $\sigma_1 = [\lambda]$ and ψ is $\sigma*$-independent of λ in HG then $j_\sigma(\psi,HG) = j_{\sigma*}(\psi,HG)$;
(c) If ψ is the root of a σ-critical defeat-link whose target is λ (so $\psi = \otimes\lambda$) then $j_\sigma(\psi,HG) = j_{[\lambda]^\frown\sigma*}(\psi,HG)$;
(d) If ψ is any other inference-node in HG then $j_\sigma(\psi,HG) = \max\{ j_\sigma(\lambda,HG) | \lambda$ is a σ-link for $\psi\}$;
(e) If λ is a support-link with basis $\{B_1,...,B_n\}$ and reason-strength ρ, and λ is σ-independent of λ, then $j_\sigma(\lambda,HG) = \min\{\rho,j_\sigma(B_1,HG),...,j_\sigma(B_n,HG)\} \sim j_\sigma(\otimes\lambda,HG)$;
(f) If λ is a support-link with basis $\{B_1,...,B_n\}$ and reason-strength ρ, and λ is σ-dependent on λ, then
$j_\sigma(\lambda,HG) = \min\{\rho,j_\sigma(B_1,HG),...,j_\sigma(B_n,HG)\} \sim [j_{[\lambda]^\frown\sigma}(\otimes\lambda,HG) + \max\{j_{\lambda^\frown\sigma}(\sim\lambda,HG),j_{\lambda^\frown\sigma}(\otimes\lambda,HG)\}]$.

The reason this constitutes a recursive definition is that when σ becomes long enough there will cease to be any σ-links.

It is now trivial to prove by induction that:

Theorem 9: $j(\psi,HG_\sigma) = j_\sigma(\psi,HG)$.

Theorem 9 constitutes our desired recursive definition of $j_\sigma(\psi,HG)$. Because $j(\psi,HG) = j_\varnothing(\psi,HG)$, this also constitutes a recursive definition of $j(\psi,HG)$.

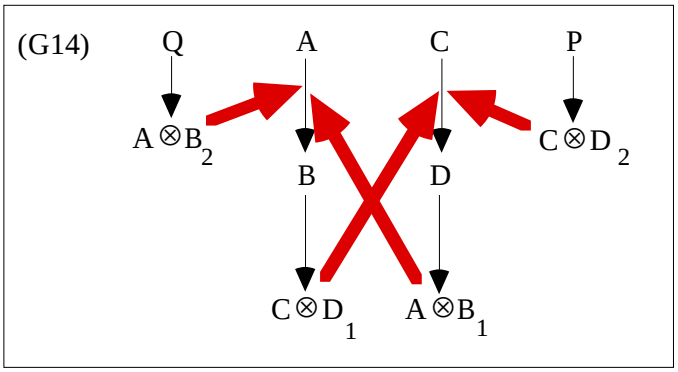
To illustrate this recursive computation, consider the simple inference-graph *G14*.Let us compute the degree of justification for *B*. To do this, we construct the inference-graph $G14_B$. By (DJ),

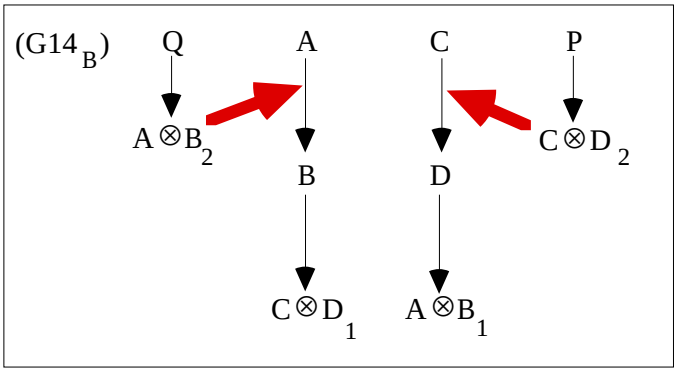
$$j(B,G14) = j_0(A,G14) \sim [j(A\otimes B_2,G14) + j(A\otimes B_1,G14_B)].$$

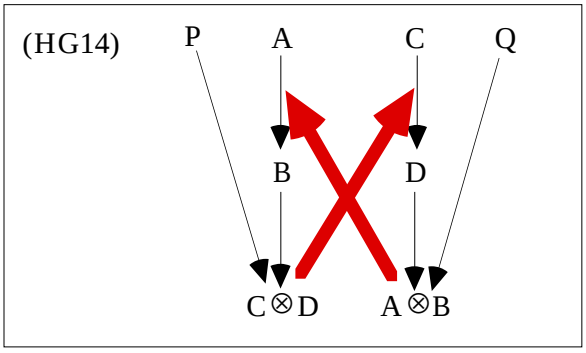
$$j(A\otimes B_1,G14_B) = j(D,G14_B) = j_0(C,G14) \sim j(C\otimes D_2,G14_B) = j_0(C,G14) \sim j_0(P,G14).$$

$j(A\otimes B_2,G14) = j_0(Q,G14)$.

So $j(B,G14) = j_0(A,G14) \sim [j_0(Q,G14) + (j_0(C,G14) \sim j_0(P,G14))]$.







The equivalent inference-hypergraph is *HG14*. To compute $j(B,HG14)$ we construct $HG14_{\langle A,B\rangle}$. By (DJH),

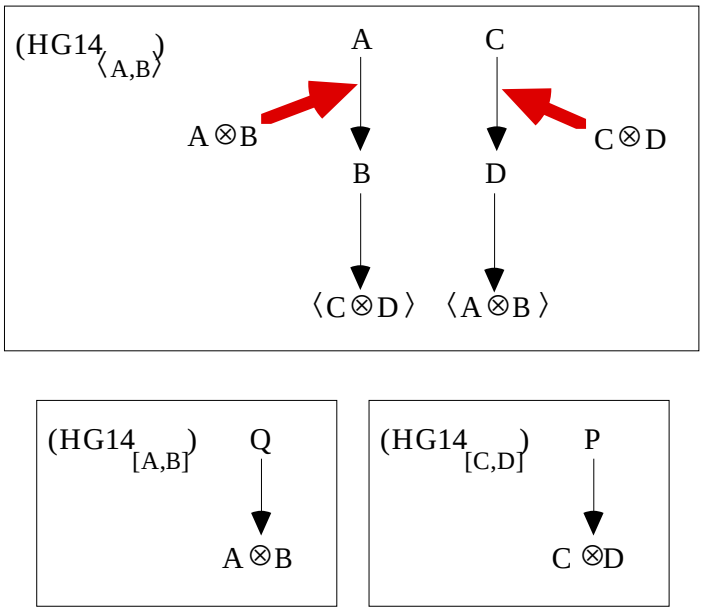
$j(B,HG14) = j_0(A,HG14) \sim [j(\langle A\otimes B\rangle,HG14_{\langle A,B\rangle}) + j(A\otimes B),HG14_{[\langle A,B\rangle]})]$.

$j(\langle A\otimes B\rangle,HG14_{\langle A,B\rangle}) = j(D,HG14_{\langle A,B\rangle}) = j_0(C,HG14) \sim j(C\otimes D,HG14_{\langle A,B\rangle})$.

$j(C\otimes D,HG14_{\langle A,B\rangle}) = j(C\otimes D),HG14_{[\langle C,D\rangle]}) = j_0(P,HG14)$.

$j(A\otimes B),HG14_{[\langle A,B\rangle]}) = j_0(Q,HG14)$.

So $j(B,HG14) = j_0(A,HG14) \sim [j_0(Q,HG14) + (j_0(C,HG14) \sim j_0(P,HG14))]$.



Finally, doing the computation recursively, by clauses (d) and (f):

$j(B,HG14) = j(\langle A,B\rangle,HG14) = j_0(A,HG14) \sim [j_{[\langle A,B\rangle]}(A\otimes B,HG14) + j_{\langle A,B\rangle}(A\otimes B,HG14)]$.

The only $\langle\langle A,B\rangle\rangle$-link to $A\otimes B$ is $\langle D,A\otimes B\rangle$, and the only $\langle[\langle A,B\rangle]\rangle$-link to $A\otimes B$ is $\langle Q,A\otimes B\rangle$. Thus by (d) and (f):

$j_{\langle A,B\rangle}(A\otimes B),HG14) = j_{\langle A,B\rangle}(\langle D,A\otimes B\rangle,HG14) = j_{\langle A,B\rangle}(D,HG14) = j_{\langle A,B\rangle}(\langle C,D\rangle,HG14)$.

$j_{[\langle A,B\rangle]}(A\otimes B,HG14) = j_{[\langle A,B\rangle]}(\langle Q,A\otimes B\rangle,HG14) = j_0(Q,HG14)$.

$\langle C,D\rangle$ is $\langle\langle A,B\rangle\rangle$-independent of $\langle C,D\rangle$, so by (e),

$j_{\langle A,B\rangle}(\langle C,D\rangle,HG14) = j_0(C,HG14) \sim j_{\langle A,B\rangle}(C\otimes D,HG14)$.

All of the defeat-links in *HG14* are $\langle\langle A,B\rangle\rangle$-critical, so $C\otimes D$ is the root of an $\langle\langle A,B\rangle\rangle$-critical defeat-link in *HG14*, and hence by (c),

$j_{\langle A,B\rangle}(C\otimes D,HG14) = j_{[\langle C,D\rangle]}(C\otimes D,HG14) = j_0(P,HG14)$.

Thus

$j_{\langle A,B\rangle}(\langle C,D\rangle,HG14) = j_0(C,HG14) \sim j_0(P,HG14)$

and

$$j(B,HG14) = j_0(A,HG14) \sim [j_0(Q,HG14) + (j_0(C,HG14) \sim j_0(P,HG14))].$$

12. Conclusions

The topic of degrees of justification resulting from defeasible reasoning is virtually unexplored in AI. There is a massive literature on degrees of probability (and the related Dempster-Shafer theory), but this paper partly argues and partly assumes (referring to arguments given elsewhere) that degrees of justification are not probabilities, in the sense that they do not conform to the probability calculus.

The starting point of the present theory is the observation that defeaters that are too weak to defeat an inference may nevertheless diminish the degree of justification of its conclusion. The search for a semantics for defeasible reasoning that is compatible with diminishing leads to a new way of computing degrees of justification recursively. A consequence of this analysis is the principle of collaborative defeat, wherein a pair of defeaters can defeat an inference when they are individually too weak to do that. Work is currently underway to implement this new computation of degrees of justification in OSCAR.

References

Barnard, G. A.

1949 "Statistical inference", *Journal of the Royal Statistical Society B, II*, 115-149.

1966 "The use of the likelihood function in statistical practice", *Proceedings v Berkeley Symposium on Mathematical Statistics and Probability I*, 27-40.

Birnbaum, Allan

1962 "On the foundations of statistical inference", *Journal of the American Statistical Association 57*, 269-326.

Chesñevar, Carlos, Ana Babriela Maguitman, and Ronald Loui

2000 "Logical models of argument", *ACM Computing Surveys* **32**, 337-383.

Covington, Michael, Donald Nute, and Andre Vellino

1997 *Prolog Programming in Depth*. Second edition. Prentice-Hall, Englewood Cliffs, NJ.

Dempster, A. P.

1968 "A generalization of Bayesian inference", *Journal of the Royal Statistical Society, Series B* **30**, 205-247.

Dung, P. M.

1995 "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*-person games", *Artificial Intelligence* **77**, 321-357.

Edwards, A. W. F.

1972 *Likelihood*. Cambridge: Cambridge University Press.

Fisher, R. A.
1922 "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society A*, *222*, 309-368.
Hacking, Ian
1965 *Logic of Statistical Inference*. Cambridge: Cambridge University Press.
Hanks, Steve, and McDermott, Drew
1986 "Default reasoning, nonmonotonic logics, and the frame problem", AAAI-86.
1987 "Nonmonotonic logic and temporal projection", *Artificial Intelligence* **33**, 379-412.
Kyburg, Henry
1961 *Probability and the Logic of Rational Belief*. Middletown, Conn.: Wesleyan University Press.
1974 *The Logical Foundations of Statistical Inference*. Dordrecht: Reidel.
Levi, Isaac
1977 "Direct inference". *Journal of Philosophy* **74**, 5-29.
Makinson, D. and Schlechta, K.
1991 "Floating conclusions and zombie paths: Two deep difficulties in the 'directly skeptical' approach to inheritance nets", *Artificial Intelligence* **48**, 199-209.
McCarthy, John
1986 "Applications of circumscription to formalizing common sense knowledge." *Artificial Intelligence* **26**, 89-116.
McDermott, Drew
1982 "A temporal logic for reasoning about processes and plans", *Cognitive Science* **6**, 101-155.
Nute, Don
1992 "Basic defeasible logic." In L. Farias del Cerro and M. Penttonen (eds.), *Intensional Logics for Programming*, Oxford University Press, 125-154.
1999 "Norms, priorities, and defeasibility." In P. McNamara and H. Prakken (eds.), *Norms, Logics and Information Systems*, IOS Press, Amsterdam, 201-218.
Peal, Judea
1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann.
Pollock, John
1970 "The structure of epistemic justification", *American Philosophical Quarterly*, monograph series 4: 62-78.
1974 *Knowledge and Justification*, Princeton University Press.
1983 "Epistemology and probability", *Synthese* **55**, 231-252.
1987 *Contemporary Theories of Knowledge*, Rowman and Littlefield.
1990 *Nomic Probability and the Foundations of Induction*, Oxford University Press.
1994 "Justification and defeat", *Artificial Intelligence* **67**, 377-408.
1995 *Cognitive Carpentry*, MIT Press.
1997 "Reasoning about change and persistence: a solution to the frame problem", *Nous* **31**, 143-169.
1998 "Perceiving and reasoning about a changing world", *Computational Intelligence* **14**, 498-562.
1998a "Degrees of Justification", in P. Weingartner, G. Schurz and G. Dorn (Eds.), *The Role of Pragmatics in Contemporary Philosophy*. *(Proceedings of the 20th International Wittgenstein Symposium 1997 Kirchberg/Wechsel, Austria)*, Hoelder-Pichler Tempsky publishers, Vienna,

207-223.
1998b "The logical foundations of goal-regression planning in autonomous agents", *Artificial Intelligence* **106**, 267-335.
2001 "Defeasible reasoning with variable degrees of justification", *Artificial Intelligence*, forthcoming.
Pollock, John, and Joe Cruz
1999 *Contemporary Theories of Knowledge*, 2nd edition. Rowman and Littlefield.
Poole, David
1988 "A logical framework for default reasoning", *Artificial Intelligence* **36**, 27-47.
Prakken, H. and G.A.W. Vreeswijk
2002 "Logics for Defeasible Argumentation", to appear in *Handbook of Philosophical Logic*, 2nd Edition, vol. 5, ed. D. Gabbay and F. Guenthner, Kluwer: Dordrecth.
Reichenbach, Hans
1949 *A Theory of Probability*. Berkeley: University of California Press. (Original German edition 1935).
Reiter, Raymond
1980 "A logic for default reasoning", *Artificial Intelligence* **13**, 81–132.
Sandewall, Erik
1972 "An approach to the frame problem and its implementation". In B. Metzer & D. Michie (eds.), *Machine Intelligence* **7**. Edinburgh: Edinburgh University Press.
Shafer, G.
1976 *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
Simari, G. R., and Loui, R. P.
1992 "A mathematical treatment of defeasible reasoning and its implementation", *Artificial Intelligence* **53**, 125–158.
Touretzky, David
1984 "Implicit orderings of defaults in inheritance systems", *Proceedings of AAAI-84*.
Touretzky, David, John Horty, and Richmond Thomason
1987 "A clash of intuitions: the current state of nonmonotonic multiple inheritance systems", *IJCAI87*, 476-482.
Verheij, Bart
1996 *Rules, Reasons, Arguments*. PhD dissertation, University of Maastricht, The Netherlands.