

A Recursive Semantics for Defeasible Reasoning

John L. Pollock

1 Reasoning in the Face of Pervasive Ignorance

One of the most striking characteristics of human beings is their ability to function successfully in complex environments about which they know very little. Reflect on how little you really know about all the individual matters of fact that characterize the world. What, other than vague generalizations, do you know about the apples on the trees of China, individual grains of sand, or even the residents of Cincinnati? But that does not prevent you from eating an apple while visiting China, lying on the beach in Hawaii, or giving a lecture in Cincinnati. Our ignorance of individual matters of fact is many orders of magnitude greater than our knowledge. And the situation does not improve when we turn to knowledge of general facts. Modern science apprises us of some generalizations, and our experience teaches us numerous higher-level although less precise general truths, but surely we are ignorant of most general truths.

In light of our pervasive ignorance, we cannot get around in the world just reasoning deductively from our prior beliefs together with new perceptual input. This is obvious when we look at the varieties of reasoning we actually employ. We tend to trust perception, assuming that things are the way they appear to us, even though we know that sometimes they are not. And we tend to assume that facts we have learned perceptually will remain true, at least for awhile, when we are no longer perceiving them, but of course, they might not. And, importantly, we combine our individual observations inductively to form beliefs about both statistical and exceptionless generalizations. None of this reasoning is deductively valid. On the other hand, we cannot be criticized for drawing conclusions on the basis of such non-conclusive evidence, because there is no feasible alternative. Our non-deductive reasoning makes our conclusions reasonable, but does not guarantee their truth. As our conclusions are not guaranteed to be true, we must countenance the possibility that new infor-

John L. Pollock
Department of Philosophy, University of Arizona, Tucson, Arizona 85721

mation will lead us to change our minds, withdrawing previously adopted beliefs. In this sense, our reasoning is “defeasible”. That is, it makes it reasonable for us to form beliefs, but it can be “defeated” by considerations that make it unreasonable to maintain the previously reasonable beliefs.

If we are to understand how rational cognition works, we must know how defeasible reasoning works, or ought to work. This chapter attempts to answer that question.

2 The Structure of Defeasible Reasoning

2.1 Inference Graphs

I assume that much of our reasoning proceeds by stringing together individual inferences into more complex arguments. In philosophy it is customary to think of arguments as linear sequences of propositions, with each member of the sequence being either a premise or the conclusion of an inference (in accordance with some inference scheme) from earlier propositions in the sequence. However, this representation of arguments is an artifact of the way we write them. In many cases the ordering of the elements of the sequence is irrelevant to the structure of the argument. For instance, consider an argument that proceeds by giving a subargument for P and an unrelated subargument for $(P \rightarrow Q)$, and then finishes by inferring Q by modus ponens. We might diagram this argument as in figure 1. The ordering of the elements of the two subarguments with respect to each other is irrelevant. If we write the argument for Q as a linear sequence of propositions, we must order the elements of the subarguments with respect to each other, thus introducing artificial structure in the representation. For many purposes it is better to represent the argument graphically, as in figure 1. Such a graph is an inference graph. The compound arrows linking elements of the inference graph represent the application of inference schemes.

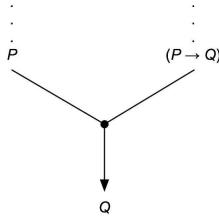


Fig. 1 An inference graph

More generally, we can take the elements of arguments to be Fitch-style *sequents* – ordered pairs of propositions and suppositions (sets of propositions), and inference rules like *conditionalization* can take advantage of that. However, for the purposes of this chapter, I will ignore that sophistication. In deductive reasoning, the inference schemes employed are deductive inference rules. What distinguishes deductive reasoning from reasoning more generally is that the reasoning is not defeasible. More precisely, given a deductive argument for a conclusion, you cannot rationally deny the conclusion without denying one or more of the premises. In contrast, consider an inductive argument. Suppose we observe a number of swans and they are all white. This gives us a reason for thinking that all swans are white. If we subsequently journey to Australia and observe a black swan, we must retract that conclusion. But notice that this does not give us a reason for retracting any of the premises. It is still reasonable to believe that each of the initially observed swans is white. What distinguishes defeasible arguments from deductive arguments is that the *addition* of information can mandate the retraction of the conclusion of a defeasible argument without mandating the retraction of any of the premises or conclusions from which the retracted conclusion was inferred.

2.2 *Rebutting defeaters*

Information that can mandate the retraction of the conclusion of a defeasible argument constitutes a defeater for the argument. There are two kinds of *defeaters*. The simplest are *rebutting defeaters*, which attack an argument by attacking its conclusion. In the inductive example concerning white swans, what defeated the argument was the discovery of a black swan, and the reason that was a defeater is that it entails the negation of the conclusion, i.e., it entails that not all swans are white. More generally, a rebutting defeater could be any reason for denying the conclusion (deductive or defeasible). For instance, I might be informed by Herbert, an ornithologist, that not all swans are white. People do not always speak truly, so the fact that he tells me this does not entail that it is true that not all swans are white. Nevertheless, because Herbert is an ornithologist, his telling me that gives me a defeasible reason for thinking that not all swans are white, so it is a rebutting defeater.

2.3 *Undercutting defeaters*

Not all defeaters are rebutting defeaters. Suppose Simon, whom I regard as very reliable, tells me, "Don't believe Herbert. He is incompetent." That Herbert told me that not all swans are white gives me a reason for believing that not all swans are white, but Simon's remarks about Herbert give me a reason for withdrawing my belief, and they do so without either (1) making me doubt that Herbert said what I took him to say or (2) giving me a reason for thinking it false that not all swans are white. Even if

Herbert is incompetent, he might have accidentally gotten it right that not all swans are white. Thus Simon's remarks constitute a defeater, but not a rebutting defeater. This is an example of an undercutting defeater. The difference between rebutting defeaters and undercutting defeaters is that rebutting defeaters attack the conclusion of a defeasible inference, while undercutting defeaters attack the defeasible inference itself, without doing so by giving us a reason for thinking it has a false conclusion. We can think of an undercutting defeater as a reason for thinking that it is false that the premises of the inference would not be true unless the conclusion were true. More simply, we can think of it as giving us a reason for believing that (under the present circumstances) the truth of the premises does not guarantee the truth of the conclusion. It will be convenient to symbolize this as "premises \otimes conclusion". It is useful to expand our graphical representation of reasoning by including defeat relations. Thus we might represent the preceding example as in figure 2. Here I have drawn the defeat relations using thick grey arrows. Note that the rebutting defeat is symmetrical, but undercutting defeat is not.

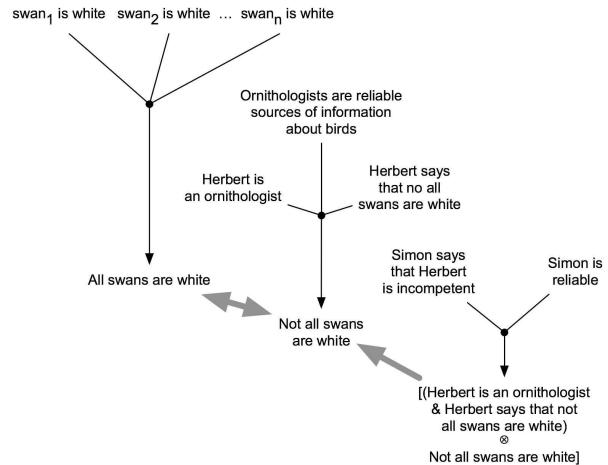


Fig. 2 Inference graph with defeat

2.4 Computing Defeat-statuses

We can combine all of a cognizer's reasoning into a single inference graph and regard that as a representation of those aspects of his cognitive state that pertain to reasoning. The hardest problem in a theory of defeasible reasoning is to give a precise account of how the structure of the cognizer's inference graph determines what he should believe. Such an account is called a "semantics" for defeasible reasoning,

although it is not a semantics in the same sense as, for example, a semantics for first-order logic. If a cognizer reasoned only deductively, it would be easy to provide an account of what he should believe. In that case, a cognizer should believe all and only the conclusions of his arguments (assuming that the premises are somehow initially justified). However, if an agent reasons defeasibly, then the conclusions of some of his arguments may be defeaters for other arguments, and so he should not believe the conclusions of all of them. For example, in figure 2, the cognizer first concludes “All swans are white”. Then he constructs an argument for a defeater for the first argument, at which point it would no longer be reasonable to believe its conclusion. But then he constructs a third argument supporting a defeater for the second (defeating) argument, and that should reinstate the first argument. Obviously, the relationships between interacting arguments can be very complex. We want a general account of how it is determined which conclusions should be believed, or to use philosophical parlance, which conclusions are “justified” and which are not. This distinction enforces a further distinction between beliefs and conclusions. When a cognizer constructs an argument, he entertains the conclusion and he entertains the propositions comprising the intervening steps, but he need not believe them. Constructing arguments is one thing. Deciding which conclusions to accept is another. What we want is a criterion which, when applied to the inference graph, determines which conclusions are defeated and which are not, i.e., a criterion that determines the defeat-statuses of the conclusions. The conclusions that ought to be believed are those that are undefeated. The remainder of the chapter will be devoted to proposing such a criterion.

3 The Multiple-Assignment Semantics

Let us collect all of an agent’s arguments into an inference-graph, where the nodes are labeled by 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 three, which represents two different arguments for $(P \ \& \ Q)$ given the premises, P , Q , A , and $(A \rightarrow (P \ \& \ Q))$. 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.

Because a conclusion can be supported by multiple arguments, it is the arguments themselves to which we must first attach defeat-statuses. Then a conclusion is undefeated iff it is supported by at least one undefeated argument. The only exception to this rule is “initial nodes”, which (from the perspective of the inference graph)

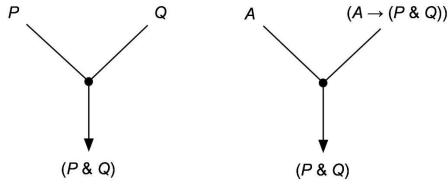


Fig. 3 An inference graph

are simply “given” as premises. Initial nodes are unsupported by arguments, but are taken to be undefeated. Ultimately, we want to use this machinery to model rational cognition. In that case, all that can be regarded as “given” is perceptual input (construed broadly to include such modes of perception as proprioception, introspection, etc.), in which case it may be inaccurate to take the initial nodes to encode propositions. It is probably better to regard them as encoding percepts.¹

The node-basis of a node is the set of roots 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 initial) then the node-basis is empty. The node-defeaters are the roots of the defeat-links having the node as their target. Given an inference-graph, a semantics must determine which nodes encode (the conclusions of) arguments that ought to be accepted, i.e., that are not defeated. This is the defeat-status computation, and nodes are marked “defeated” or “undefeated”. The defeat-status computation is made more complex by the fact that some arguments support their conclusions more strongly than other arguments. For instance, if Jones tells me it is raining, and Smith denies it, and I regard them as equally reliable, then I have equally strong arguments both for believing that it is raining and for believing that it is not raining. In that case, I should withhold belief, not accepting either conclusion. On the other hand, if I regard Jones as much more reliable than Smith, then I have a stronger argument for believing that it is raining, and if the difference is great enough, that is the conclusion I should draw. So argument-strengths make a difference. However, most semantics for defeasible reasoning ignore argument strengths, pretending that all initial nodes are equally well justified and all inference schemes equally strong. I will make this same simplifying assumption in this chapter. What can we say about the semantics in this simplified case?

Let us define:

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

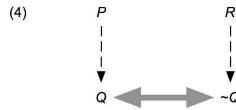
It is initially tempting to try to characterize defeat-statuses recursively using the following two rules:

- (D1) Initial nodes are undefeated.
- (D2) A non-initial node is undefeated iff all the members of its node-basis

¹ See [11] and [12] for a fuller discussion of this.

are undefeated and all node-defeaters are defeated.

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 (4), where dashed arrows indicate defeasible inferences and heavy arrows indicate defeat-links. In computing defeat-statuses in inference-graph (4), we cannot proceed recursively using rules (D1) and (D2), because that would require us to know the defeat-status of Q before computing that of $\sim Q$, and also to know the defeat-status of $\sim Q$ before computing that of Q . The general problem is that a node Q can have an inference/defeasible-descendant that is a defeater of Q , where an inference/defeasible-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 Q -dependent iff it is an inference/defeasible-descendant of a node Q . So the recursion is blocked in inference-graph (4) by there being Q -dependent defeaters of Q and $\sim Q$ -dependent defeaters of $\sim Q$.



Inference-graph (4) 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 Q nor $\sim Q$. In other words, both Q and $\sim Q$ should be defeated. This constitutes a counter-example to rule (D2). So not only do rules (D1) and (D2) 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 (see [9, 10]) that rules (D1) and (D2) might be used to characterize defeat-statuses in another way. Reiter’s default logic [13] 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. There are simple examples showing that this semantics is inadequate for the general defeasible reasoning of epistemic agents (see below), but the idea of having multiple extensions suggested to me that rules (D1) and (D2) might be used to characterize multiple “status assignments”. On this approach, a partial status assignment is an assignment of defeat-statuses to a subset of the nodes of the inference-graph in accordance with (D1) and (D2):

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 non-initial 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”.

My (1995) semantics defined:

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.

I will refer to this semantics as the multiple-assignment semantics. To illustrate, consider inference-graph (4) again. There are two status assignments for this inference graph:

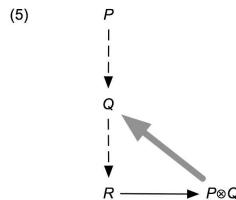
assignment 1:

P “undefeated”
 R “undefeated”
 Q “undefeated”
 $\sim Q$ “defeated”

assignment 2:

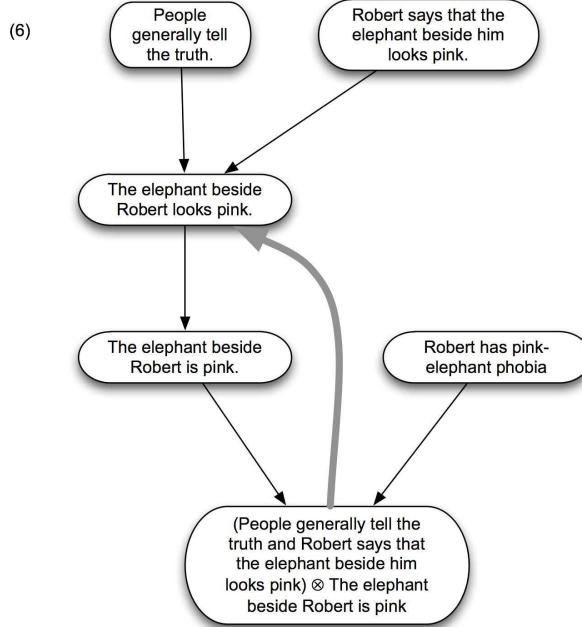
P “undefeated”
 R “defeated”
 Q “defeated”
 $\sim Q$ “undefeated”

P and R are undefeated, but neither Q nor $\sim Q$ is assigned “undefeated” in every assignment, so both are defeated.



The reason for making status assignments “partial” is that there are inference graphs for which it is impossible to construct status assignments assigning statuses to every node. One case in which this happens is when we have “self-defeating arguments”, i.e., arguments whose conclusions defeat some of the inferences leading to those conclusions. A simple example is inference-graph (5). A partial status assignment must assign “undefeated” to P . If it assigned “undefeated” to Q then it would assign “undefeated” to R and $P \otimes Q$, 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 $P \otimes Q$, 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 $P \otimes Q$. Hence there is only one status assignment (i.e., maximal partial status assignment), and it assigns “undefeated” to P and nothing to the other nodes. Accordingly, P is undefeated and the other nodes are defeated. An intuitive example having approximately the same form is shown in inference-graph (6). Here we suppose that people generally tell the truth, and this gives us a reason for believing what they tell us. However, some people suffer from a malady known as “pink-elephant phobia”. In the presence of pink elephants, they become strangely disoriented so that their statements about their surroundings cease to be reliable. Now imagine Robert, who tells us that the elephant beside him looks pink. In ordinary circumstances, we would infer that the elephant beside Robert does look pink, and hence probably is pink. However, Robert suffers from pink-elephant phobia. So if it were true that the elephant beside Robert is pink, we could not rely upon his report to conclude that it is. So we should not conclude that it is pink. We may be left wondering why he would say that it is, but we cannot explain his utterance by supposing that the elephant really is pink. So this gives us no reason at all for a judgment about the color of the elephant. On the other hand, it gives us no reason to doubt that Robert did say that the elephant is pink, or that Robert has pink-elephant phobia. Those are perfectly justified beliefs.

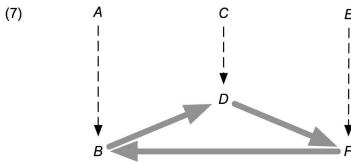


Inference-graphs (5) and (6) constitute intuitive counterexamples to default logic [13] and the stable model semantics [2] because there are no extensions. Hence

on those semantics, P has the same status as Q , R , and $P \otimes Q$. 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.

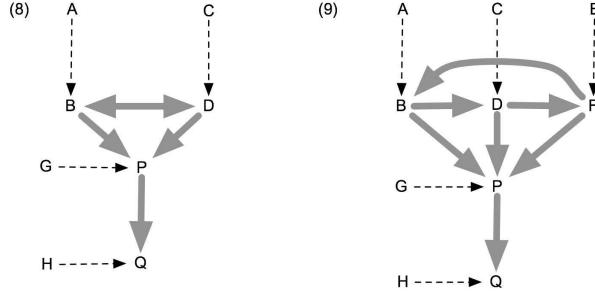
4 A Problem Case

The multiple-assignment semantics produces the intuitively correct answer for many complicated inference-graphs. For a number of years, I thought that, given the simplifying assumption that all arguments are equally strong, this semantics was correct. But I no longer think so. Here is the problem. Contrast inference-graph (4) with inference-graph (7). Inference-graph (7) involves “odd-length defeat cycles”. For an example of inference-graph (7), 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 multiple-assignment semantics 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 to 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 (4). This difference has always bothered me.



That we get the right answer in a different way does not show that the semantics is incorrect. As long as otherwise equivalent inference-graphs containing odd-length and even-length defeat cycles always produce the same defeat-statuses throughout the graphs, there is no problem. However, they do not. Contrast inference-graphs (8) and (9). In inference-graph (8), there are two status assignments, one assigning “defeated” to B and “undefeated” to D , and the other assigning “undefeated” to B and “defeated” to D . On either status assignment, P has an undefeated defeater, so it is defeated on both status assignments, with the result that Q is undefeated on both status assignments. Hence Q is undefeated simpliciter. However, in inference-graph (9), there is only one status-assignment, and it assigns no status to any of B , D , F , P ,

or Q . Thus Q is defeated in inference-graph (9), but undefeated in inference-graph (8). This, I take it, is a problem. Although it might not be clear which inference-graph is producing the right answer, the right answer ought to be the same for both inference-graphs. Thus the semantics is getting one of them it wrong. It is worth noting in passing that, as far as I know, no currently available semantics for defeasible reasoning handles (8) and (9) correctly. I take this to show that we need a different semantics.

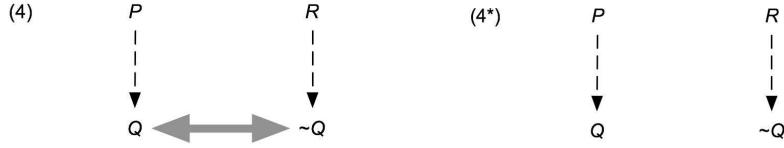


5 A Recursive Semantics

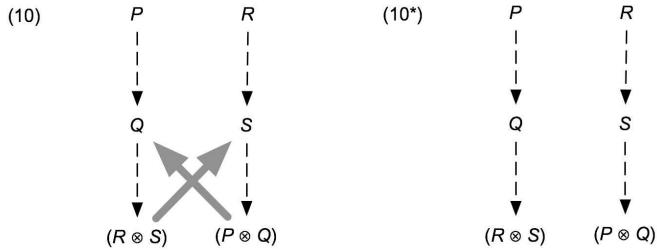
The multiple-assignment semantics is based upon the two rules:

- (D1) Initial nodes are undefeated.
- (D2)] A non-initial node is undefeated if all the members of its node-basis are undefeated and all node-defeaters are defeated.

We have seen that these rules are not true as stated. For example, inference-graph (4) is a counterexample to rule (D2). Both Q and $\sim Q$ should be defeated, but then both have undefeated node-bases but no undefeated defeaters. I tried to avoid this problem by imposing these rules instead on partial-status assignments. But perhaps we should take seriously the fact that these rules are simply wrong. In inference-graph (4), in computing the defeat-status of Q , what is crucial is that (a) its node-basis is undefeated, (b) the node-basis of its defeater is undefeated, and (c) there is no other defeater for $\sim Q$ besides Q itself. We can capture this by asking whether $\sim Q$ would be defeated if it were not defeated by Q . We can test this by removing the mutual defeat-links between Q and $\sim Q$, producing inference-graph (4*). In (4*), $\sim Q$ is undefeated. The proposal is that this should make Q defeated in (4). Note that the defeaters we are removing in constructing inference-graph (4*) are those that are Q -dependent, i.e., those that can be reached by following paths from Q consisting of inference-links and defeat-links.



Consider another example . inference-graph (10). In computing the defeat-status of Q , we note that its node-basis is undefeated, and its defeater $P \otimes Q$ is defeated only by the Q -dependent defeat-link from $R \otimes S$. If we remove the Q -dependent defeat-links from inference-graph (10) we get inference-graph (10*). In inference-graph (10*), $P \otimes Q$ is undefeated, so again, the proposal is that this makes Q defeated in inference-graph (10).



These examples suggest that we might replace rule (D2) by a rule that computes the defeat-statuses of defeat-links in a modified inference-graph from which we have removed those defeat-links that make the computation circular. 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:

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

Definition 2. θ is φ -dependent iff there is an inference/defeat-path from φ to θ .

Definition 3. A *circular inference/defeat-path* from a node φ to itself is an inference/defeat-path from φ to a defeater for φ .

Definition 4. A defeat-link is φ -critical iff it is a member of some minimal set of defeat-links such that removing all the defeat-links in the set suffices to cut all the circular inference/defeat-paths from φ to φ .

It will be convenient to modify our understanding of initial nodes. Previously, I took them to be automatically undefeated, and we can still regard that as the default value, but it will also be useful to be able to stipulate that some of the initial

nodes in a newly-constructed inference-graph are defeated. The construction I am going to propose builds new inference-graphs as subgraphs of preexisting inference-graphs by (1) deleting φ -critical links, and (2) making φ -independent nodes initial, i.e., deleting the arguments for them. The latter nodes, being φ -independent, have defeat-statuses that were computable in the original inference-graph without first having to compute a defeat-status for φ . I want to be able to simply stipulate that these newly-initial nodes have the same defeat-statuses in the new inference-graph as they had in the original. This allows us to define:

Definition 5. If φ is a node of an inference-graph G , let G_φ be the inference-graph that results from deleting all φ -critical defeat-links from G and making all members of the node-basis of φ and all φ -independent nodes initial-nodes (i.e., deleting their support-links and defeat-links) with stipulated defeat-statuses the same as their defeat-statuses in G .

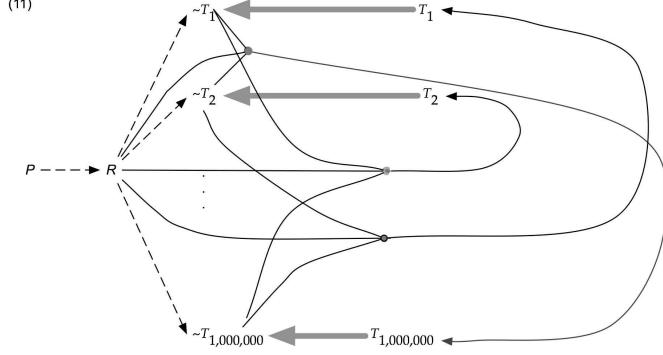
My proposed semantics now consists of two rules:

- (CL1) Initial nodes are undefeated unless they are stipulated to be defeated.
- (CL2) A non-initial node φ is undefeated in an inference-graph G iff all members of the node-basis of φ are undefeated in G and any defeater for φ is defeated in G_φ .

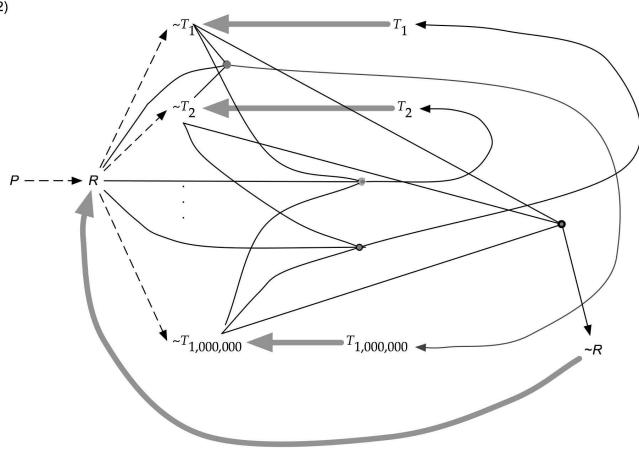
On the assumption that arguments cannot be circular, this pair of rules can be applied recursively to compute the defeat-status of any node in a finite inference-graph. The recursion simply steps through arguments, computing the defeat-status of each node . after the defeat-statuses of the nodes in φ 's node-basis are computed. The problem of circular inference/defeat-paths is avoided by removing the φ -critical defeat-links and evaluating node-defeaters in G_φ . I will refer to this new semantics as *the critical-link semantics*, and contrast it with the multiple-assignment semantics.

I believe that the critical-link semantics gets everything right that the multiple-assignment semantics got right. Consider a more complex example. Inference-graph (11) illustrates the so called “lottery paradox” [3]. 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. That is, the newspaper report gives us a defeasible reason for believing the lottery is fair and has a million tickets. 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 (i) the description R (which entails that some ticket will be drawn) and (ii) 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 (11). The result is that for each i , there is a status assignment on which $\sim T_i$ is assigned “defeated” and the

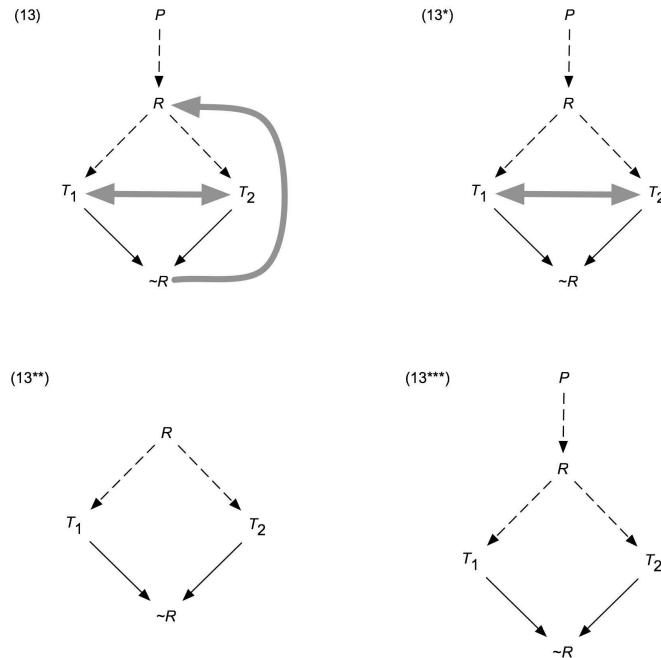
other $\sim T_j$ 'es are all assigned “undefeated”, and hence none of them are assigned “undefeated” in every status assignment.



I believe that all (skeptical) semantics for defeasible reasoning get the lottery paradox right. A more interesting example is the “lottery paradox paradox”, diagrammed in inference-graph (12). 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 $\sim P$ to $\sim R$. This is a 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 [5], 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 [8] also gets this wrong. This was the example that led me to the multiple-assignment semantics. The multiple-assignment semantics gets this right. We still have the same status assignments as in inference-graph (11), 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.



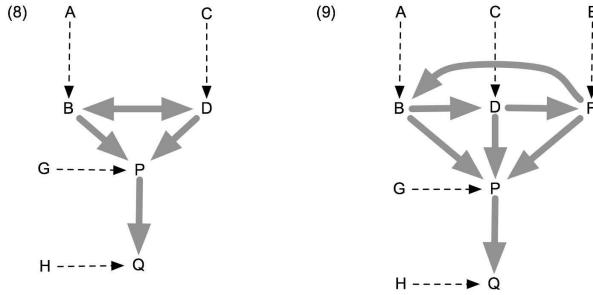
It will be convenient to have a simpler example of an inference-graph with the same general structure as the lottery paradox paradox. For that purpose we can use inference-graph (13). Here P and R should be undefeated, but T_1 , T_2 , and $\sim R$ should be defeated. In the critical link semantics, to compute the defeat-status of R in inference-graph (13), we construct (13*) by removing the only defeat-link whose removal results in R no longer having an R -dependent defeater. In (13*), the triangle consisting of R , T_1 and T_2 is analogous to inference-graph (4), with the result that T_1 and T_2 are both defeated in inference-graph (13*). They constitute the node-basis for $\sim R$, so $\sim R$ is also defeated in inference-graph (13*). Thus by (CL2), R is undefeated in inference-graph (13). Turning to T_1 and T_2 in inference-graph (13), both have R as their node-basis, and R is undefeated. Thus to compute the defeat-status of T_1 or T_2 , we construct inference-graph (13**), and observe that T_1 and T_2 are undefeated there. It then follows by (CL2) that T_1 and T_2 are defeated in inference-graph (13). Then because T_1 and T_2 are defeated, $\sim R$ is defeated in inference-graph (13). So we get the intuitively correct answers throughout.



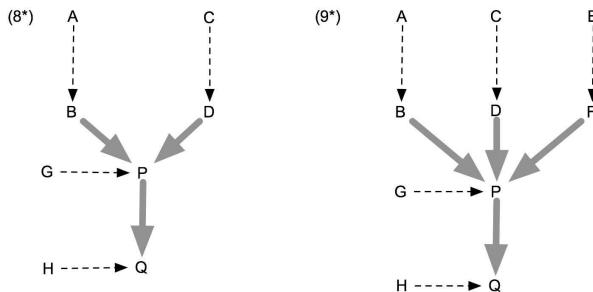
Inference-graph (13) also illustrates why, in constructing G_φ , we remove only the φ -critical defeat-links, and not all of the φ -dependent defeat-links. All of the defeat-links in inference-graph (13) are R -dependent, and if we remove them all we get inference-graph (13***). But in inference-graph (13***), $\sim R$ is undefeated. This would result in R being defeated in inference-graph (13) rather than undefeated. Thus it is crucial to remove only the φ -critical defeat-links rather than all the φ -dependent defeat-links.

6 The Problem Cases

Now let us turn to some cases that the multiple-assignment semantics does not or may not get right. First, consider the pair of inference-graphs that motivated the search for a new semantics. These are inference-graphs (8) and (9). In these inference-graphs, not everyone agrees whether Q should come out defeated or undefeated, but it does seem clear that whatever the right answer is, it should be the same for both inference-graphs. Unfortunately, on the multiple-assignment semantics, Q is undefeated in inference-graph (8) and defeated in inference-graph (9).



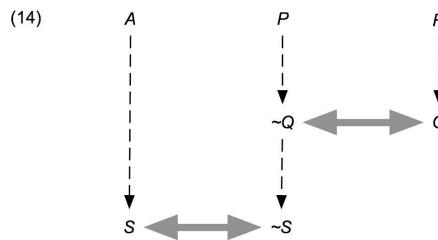
On the critical-link semantics, we compute the defeat-statuses of B and D in inference-graph (8) by constructing inference-graph (8^*) . B and D are undefeated in inference-graph (8^*) , so each defeats the other in inference-graph (8), with the result that B and D are defeated in inference-graph (8). There are no P -critical defeat-links in (8), so removing P -critical defeat-links leaves inference-graph (8) unchanged. B and D are defeated in inference-graph (8), so it follows that P is defeated in inference-graph (8). Then because there are no Q -dependent defeat-links in inference-graph (8), Q is undefeated.



The computation of defeat-statuses in inference-graph (9) works in exactly the same way, via inference-graph (9^*) , again producing the result that Q is undefeated. So on the critical-link semantics, we do not get a divergence between inference-graphs (8) and (9).

Still, we can ask whether the answer we get for inference-graphs (8) and (9) is the correct answer. There is some intuitive reason for thinking so. In inference-graph (8), B and D are defeated, so they should not have the power to defeat P , and hence

P should defeat Q . Similarly, in inference-graph (9), all three of B , D , and F are defeated, and so again, D should not have the power to defeat P , and hence P should defeat Q . However, not everyone agrees that this intuitive reasoning is correct. This issue is closely connected with a question that has puzzled theorists since the earliest work on the semantics of defeasible reasoning. The multiple-assignment semantics, as well as default logic, the stable model semantics, circumscription, and almost every familiar semantics for defeasible reasoning and nonmonotonic logic, supports what I have called [8] “presumptive defeat”.² For example, consider inference-graph (4). On the multiple-assignment semantics, a defeated conclusion like Q that is assigned “defeated” in some status assignment and “undefeated” in another retains the ability to defeat. That is because, in the assignment in which it is undefeated, the defeatee is defeated, and hence not undefeated in all status-assignments. In the case of inference-graph (4) this has the consequence that S is assigned “defeated” in those status-assignments in which Q is assigned “defeated”, but S is assigned “undefeated” and $\sim S$ is assigned “defeated” in those status-assignments in which Q is assigned “undefeated”. Touretzky, Harty, and Thomason [14] called this “ambiguity propagation”, and Makinson and Schlechta [4] called such arguments “Zombie arguments” (they are dead, but they can still get you). However, the critical-link semantics precludes presumptive defeat. It entails that Q , $\sim Q$, and hence S , are all defeated, and $\sim S$ is undefeated. Is this the right answer?



Consider an example. You are sitting with Keith and Alvin, and the following conversation ensues:

Keith: I heard on the news this morning that it is going to rain this afternoon.
 Alvin: Nonsense! I was sitting right beside you listening to the same weather report, and the announcer clearly said that it is going to be a sunny day in Tucson.
 Keith: You idiot, you must have cotton in your ears! It was perfectly clear that he said it is going to rain.
 Alvin: You never pay attention. No one in his right mind could have thought he said it was going to rain. He said it would be sunny.
 ...

At that point, you wander off shaking your head, still wondering what the weather is going to be. Then it occurs to you that it is about time for the noon News, so you

² The only semantics I know about that does not support presumptive defeat are certain versions of Nute's [6] defeasible logic. See also Covington, Nute, and Vellino [1], and Nute [7].

turn on the radio and hear the announcer say, This just in from the National Weather Service. It is going to rain in Tucson this afternoon. Surely, that settles the matter. You will believe, with complete justification, that it is going to rain. The earlier conversation between Keith and Alvin does not defeat your judgment on the basis of the noon broadcast. This example has the form of inference-graph (14) if we let:

$$\begin{aligned} S &= \text{"It is going to rain in Tucson this afternoon"} \\ Q &= \text{"The morning news said that } S\text{"} \\ P &= \text{"Alvin says that } Q\text{"} \\ R &= \text{"Keith says that } Q\text{"} \\ A &= \text{"The noon news says that } S\text{"} \end{aligned}$$

This seems to me to be a fairly compelling example of the failure of presumptive defeat. Formally, presumptive defeat arises for the multiple-assignment semantics from the fact that if a node P is defeated in one assignment and undefeated in another, then P -dependent nodes will also have different defeat-statuses in the different assignments unless one of their inference-ancestors is defeated absolutely (i.e., in all status assignments). A similar problem arises for inference-nodes P that cannot be assigned defeat-statuses in any assignments. This occurs, for example, in cases of self-defeat or when there are odd-length defeat cycles. In this case, no P -dependent node can be assigned a defeat-status either unless one of its inference-ancestors is defeated absolutely. For example, consider once more the sad case of Robert, the pink-elephant-phobic (inference-graph (6)). We observed that Roberts statement that the elephant beside him is pink does not give us a good reason for believing that it really is pink. Now suppose that Robert is accompanied by Herbert, who is also standing beside the elephant. While Robert is blathering about pink-elephants, Herbert turns to you and says, "I read in the newspaper this morning that the President is going to visit China." From this you infer that he did read that in the newspaper, and hence the President is probably going to visit China. Suppose, however, that Herbert also suffers from pink-elephant phobia. Does that make any difference? It does not seem so, because as we observed, Robert's statement gives us no reason to think the elephant is pink, and so no reason to distrust Herbert's statement. This scenario is diagrammed in inference-graph (15). However, on the multiple-assignment semantics,

The elephant beside Robert and Herbert is pink

has no status assignment, and hence neither does

(People generally tell the truth and Herbert says that he read in the newspaper this morning that the President is going to visit China) \otimes Herbert read in the newspaper this morning that the President is going to visit China

or

Herbert read in the newspaper this morning that the President is going to visit China

or

The president is going to visit China.

This seems clearly wrong. On the other hand, on the critical-link semantics,

The elephant beside Robert and Herbert looks link

is defeated, and hence so is

The elephant beside Robert and Herbert is pink

and so is

(People generally tell the truth and Herbert says that he read in the newspaper this morning that the President is going to visit China) \otimes Herbert read in the newspaper this morning that the President is going to visit China.

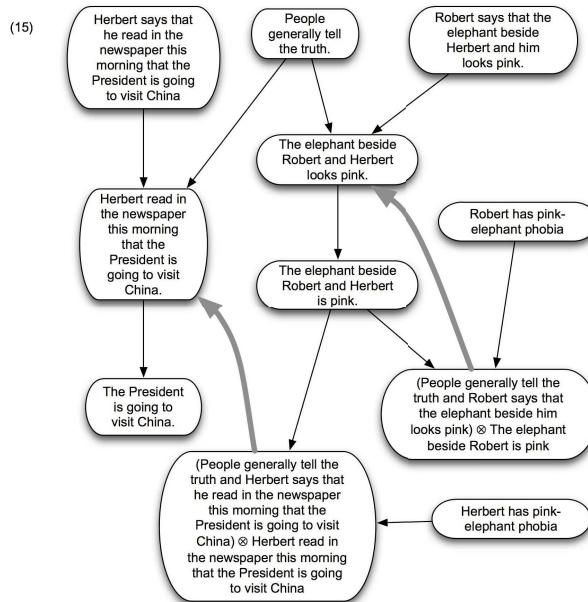
Accordingly,

Herbert read in the newspaper this morning that the President is going to visit China

and

The president is going to visit China

are undefeated, which is the intuitively correct result.



The upshot is that the critical-link semantics agrees with the multiple-assignment semantics on simple cases in which the latter seems to give the right answer, but the critical-link semantics also seems to get right a number of cases that the multiple-assignment semantics gets wrong. The test of a semantics for defeasible reasoning is that it agrees with our intuitions about clear cases. So we have reasonably strong inductive reasons for thinking that the critical-link semantics properly characterizes the semantics of defeasible reasoning.

7 Computing Defeat-Statuses

Principles (CL1) and (CL2) provide a recursive characterization of defeat-status relative to an inference-graph. However, this characterization does not lend itself well to implementation because it requires the construction of modified inference-graphs, which would be computationally expensive. The objective of this section is to produce an equivalent recursive characterization that appeals only to the given inference-graph.

A defeat-link is φ -critical iff it is a member of a minimal set such that removing all the defeat-links in the set suffices to cut all the circular inference/defeat-paths from φ to φ . A necessary condition for a defeat-link L to be φ -critical is that it lie on such a circular path. In general, there can be diverging and reconverging paths with several “parallel” defeat-links, as in figure 16. In figure 16, removing the defeat-link D_3 suffices to cut both circular paths. But the set D_1, D_2 of parallel defeat-links is also a minimal set of defeat-links such that the removal of all the links in the set suffices to cut all the circular inference/defeat-paths from φ to φ . Thus in figure 16, all of the defeat-links are φ -critical. However, lying on a circular inference/defeat-path is not a sufficient condition for being φ -critical. A defeat-link on a circular inference/defeat-path from φ to φ fails to be φ -critical when there is a path around it consisting entirely of support-links, as diagrammed in figure 17. In this case, you must remove D_3 to cut both paths, but once you have done that, removing D_1 is a gratuitous additional deletion. So D_1 is not contained in a minimal set of deletions sufficient for cutting all the circular inference/defeat-paths from φ to φ , and hence D_1 is not φ -critical. This phenomenon is also illustrated by inference-graph (13), and we saw that it is crucial to the computation of degrees of justification in that inference-graph that such defeat-links not be regarded as φ -critical. It turns out that this is the only way a defeat-link on a circular inference/defeat-path can fail to be φ -critical, as will now be proven.

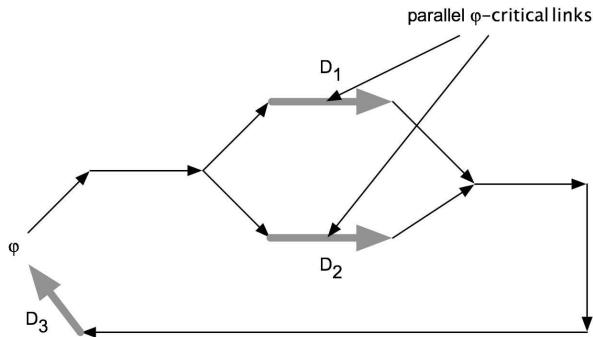


Fig. 16 Parallel φ -critical defeat-links

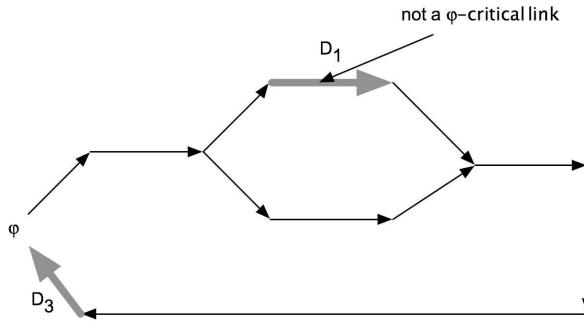


Fig. 17 Defeat link that is not φ -critical

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 $\alpha = \beta$ or the path contains a subpath originating on α and terminating on β . *Node-ancestors* of a node are nodes that can be reached by following support-links backwards. It will be convenient to define:

Definition 6. A defeat-link L is bypassed on an inference/defeat-path μ in G iff there is a node α preceding the root of L on μ and a node β preceded by the target of L on μ such that $\alpha = \beta$ or α is a node-ancestor of β in G .

Definition 7. μ is a φ -circular-path in G iff μ is a circular inference/defeat-path in G from φ to φ and no defeat-link in G is bypassed on μ .

Lemma 1. *Lemma 1: If μ_1 and μ_2 are φ -circular-paths and every defeat-link in μ_1 occurs in μ_2 , then μ_1 and μ_2 contain the same defeat-links and they occur in the same order.*

Proof. Proof: Suppose the defeat-links in μ_1 are $\delta_1, \dots, \delta_n$, occurring in that order. Suppose μ_1 and μ_2 differ first at the i th defeat-link. Then μ_1 and μ_2 look as in figure 18. But every defeat-link in μ_1 occurs in μ_2 , so δ_i must occur later in μ_2 . But then the path from δ_{i-1} to δ_i in μ_1 is a bypass around δ_i^* in μ_2 , which is impossible if it is a φ -circular-path. \square

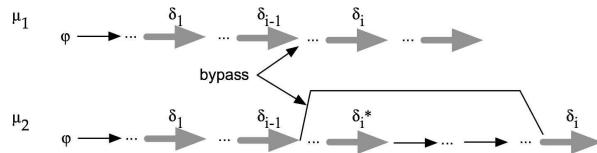


Fig. 18 Paths must agree

Lemma 2. *Every defeat-link in a φ -circular-path is φ -critical.*

Proof. Suppose δ is a defeat-link on the φ -circular-path μ . Let D be the set of all defeaters in the inference-graph other than those on μ . If deleting all members of D is sufficient to cut all φ -circular-paths not containing δ , then select a minimal subset D_0 of D whose deletion is sufficient to cut all φ -circular-paths not containing δ . Adding δ to D_0 gives us a set of defeat-links whose deletion is sufficient to cut all φ -circular-paths. Furthermore, it is minimal, because adding δ cannot cut any paths not containing δ , and all members of D_0 are required to cut those paths. Thus δ is a member of a minimal set of defeat-links the deletion of which is sufficient to cut all φ -circular-paths, i.e., δ is φ -critical. Thus if δ is not φ -critical, there is a φ -circular-path v not containing δ and not cut by cutting all defeat-links not in μ . That is only possible if every defeat-link in v is in μ . But then by the previous lemma, μ and v must contain the same defeat-links, so contrary to supposition, δ is in v . Thus the supposition that δ is not φ -critical is inconsistent with the supposition that it lies on a φ -circular-path. \square

Lemma 3. *If a defeat-link does not occur on any φ -circular-path then it is not φ -critical.*

Proof. For every circular inference/defeat-path μ from φ to φ there is a φ -circular-path v such that every defeat-link in v is in μ . v results from removing bypassed defeat-links and support-links in μ and replacing them by their bypasses. It follows that any set of deletions of defeat-links that will cut all φ -circular-paths will also cut every circular inference/defeat-path from φ to φ . Conversely, φ -circular-paths are also circular-paths from φ to φ , so any set of deletions that cuts all circular-paths from φ to φ will also cut all φ -circular-paths. So the φ -circular-paths and the circular-paths from φ to φ have the same sets of deletions of defeat-links sufficient to cut them, and hence the same minimal sets of deletions. If a defeat-link δ does not occur on any φ -circular-path, then it is irrelevant to cutting all the φ -circular-paths, and hence it is not in any minimal set of deletions sufficient to cut all circular-paths from φ to φ , i.e., it is not φ -critical. \square

Theorem 4 follows immediately from lemmas 2 and 3:

Theorem 4. *A defeat-link is φ -critical in G iff it lies on a φ -circular-path in G .*

A further simplification results from observing that, for the purpose of deciding whether a defeat-link is φ -critical, all we have to know about φ -circular-paths is what defeat-links occur in them. It makes no difference what support-links they contain. So let us define:

Definition 8. A φ -defeat-loop is a sequence μ of defeat-links for which there is a φ -circular-path v such that the same defeat-links occur in μ and v and in the same order.

In other words, to construct a φ -defeat-loop from a φ -circular-path we simply remove all the support-links. We have the following very simple characterization of φ -defeat-loops:

Theorem 5. A sequence $\langle \delta_1, \dots, \delta_n \rangle$ of defeat-links is a φ -defeat-loop iff (1) φ is a node-ancestor of the root of δ_1 , but not of the root of any δ_k for $k > 1$, (2) φ is the target of δ_n , and (3) for each $k < n$, the target of δ_k is equal to or an ancestor of the root of δ_{k+1} , but not of the root of δ_{k+j} for $j > 1$.

The significance of φ -defeat-loops is that by omitting the support-links we make them easier to process, but we still have the simple theorem:

Theorem 6. A defeat-link is φ -critical in G iff it lies on a φ -defeat-loop in G .

In simple cases, G_φ will be an inference-graph in which no node ψ has a ψ -critical defeat-link. But in more complex cases, like inference-graph (13), we have to repeat the construction, constructing first G_φ , and then $(G_{\text{arphi}})_\psi$. Let us define recursively:

Definition 9. $G_{\langle \varphi_1, \dots, \varphi_n \rangle} = (G_{\langle \varphi_2, \dots, \varphi_n \rangle})_{\varphi_1}$

As formulated, the recursive semantics requires us to construct the inference-graphs $G_{\langle \varphi_1, \dots, \varphi_n \rangle}$. To reformulate the semantics so as to avoid this, let us define recursively:

Definition 10.

A defeat-link δ of G is $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G iff (1) δ lies on a φ_1 -defeat-loop μ in G containing no $\langle \varphi_2, \dots, \varphi_n \rangle$ -critical defeat-links.

A defeat-link δ of G is hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G iff either δ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G or δ is hereditarily- $\langle \varphi_2, \dots, \varphi_n \rangle$ in G .

A defeater (i.e., a node) of G is hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G iff it is the root of a hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical defeat-link in G .

Obviously:

Theorem 7. δ is hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G iff δ is φ_1 -critical in $G_{\langle \varphi_2, \dots, \varphi_n \rangle}$ or φ_2 -critical in $G_{\langle \varphi_3, \dots, \varphi_n \rangle}$ or ... or φ_n -critical in G .

Note that a defeat-link that is φ_i -critical in $G_{\langle \varphi_{i+1}, \dots, \varphi_n \rangle}$ does not exist in $G_{\langle \varphi_{j+1}, \dots, \varphi_n \rangle}$ for $j < i$, so:

Theorem 8. δ is φ_1 -critical in $G_{\langle \varphi_2, \dots, \varphi_n \rangle}$ iff δ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G .

Furthermore, a defeat-link still exists in $G_{\langle \varphi_3, \dots, \varphi_n \rangle}$ (i.e., has not been removed) iff it is not $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G .

Where $\theta, \varphi_2, \dots, \varphi_n$ are nodes of an inference-graph G , define:

Definition 11. θ is $\langle \varphi \rangle$ -independent of ψ in G iff there is no inference/defeat-path in G from φ to θ .

θ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -independent in G iff every inference/defeat-path in G from φ_1 to θ contains a hereditarily- $\langle \varphi_2, \dots, \varphi_n \rangle$ -critical defeat-link.

Theorem 9. θ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -independent in G iff θ is φ_1 -independent in $G_{\langle \varphi_2, \dots, \varphi_n \rangle}$.

Let us define recursively:

Definition 12.

- (a) If ψ is initial in G then ψ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -*undefeated* in G iff ψ is undefeated in G ;
- (b) If ψ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -independent in G then ψ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -*undefeated* in G iff ψ is $\langle \varphi_2, \dots, \varphi_n \rangle$ -*undefeated* in G ;
- (c) Otherwise, ψ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -*undefeated* in G iff (1) all members of the node-basis of ψ are $\langle \varphi_1, \dots, \varphi_n \rangle$ -*undefeated* in G , (2) all defeaters for ψ that are $\langle \varphi_1, \dots, \varphi_n \rangle$ -independent of ψ in G and are not hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G (i.e., still exist in $G_{\langle \varphi_3, \dots, \varphi_n \rangle}$) are $\langle \varphi_1, \dots, \varphi_n \rangle$ -defeated in G , and (3) all defeaters for ψ that are $\langle \varphi_1, \dots, \varphi_n \rangle$ -dependent of ψ in G and are not hereditarily- $\langle \varphi_1, \dots, \varphi_n \rangle$ -critical in G (i.e., still exist in $G_{\langle \varphi_3, \dots, \varphi_n \rangle}$) are $\langle \psi, \varphi_1, \dots, \varphi_n \rangle$ -defeated in G ,

The reason this is a recursive definition is that we always reach an n at which there are no more $\langle \varphi_1, \dots, \varphi_n \rangle$ -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 on n that:

Theorem 10. ψ is *undefeated* in $G_{\langle \varphi_1, \dots, \varphi_n \rangle}$ iff ψ is $\langle \varphi_1, \dots, \varphi_n \rangle$ -*undefeated* in G .

Thus we have a recursive definition of the defeat-status of a node that computes defeat-statuses entirely by reference to the given inference-graph rather than by building a sequence of modified inference-graphs in accordance with the original analysis. This is easily implemented with two pages of LISP code.

8 Conclusions

In an environment of real-world complexity, it is impossible to know enough about the world to confine one's reasoning to deductively valid inferences. One has to reason defeasibly, drawing conclusions that are made reasonable by one's evidence, but be prepared to change one's mind in the face of new evidence. The question then arises how defeasible reasoning ought to work. In particular, given a set of defeasible arguments some of which support defeaters for others, how is it determined which conclusions ought to be believed? Most semantics for defeasible reasoning agree with regard to simple cases, and produce intuitively congenial answers. But there are some complex cases that all existing semantics seem to get wrong. This chapter proposes a new semantics, based on the concept of a critical link, that arguably gets those cases right. Furthermore, the semantics is recursive and easily implemented.

References

1. M. A. Covington, D. Nute, and A. Vellino. *Prolog Programming in Depth*. Prentice-Hall, Englewood Cliffs, New Jersey, second edition, 1997.
2. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
3. H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, Connecticut, 1961.
4. D. Makinson and K. Schlechta. Floating conclusions and zombie paths: Two deep difficulties in the “directly skeptical” approach to inheritance nets. *Artificial Intelligence*, 48(2):199–209, 1991.
5. J. McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28(1):89–116, 1986.
6. D. Nute. Basic defeasible logic. In L. F. del Cerro and M. Penttonen, editors, *Intensional Logics for Programming*, pages 125–154. Oxford University Press, USA, 1992.
7. D. Nute. Norms, priorities, and defeasibility. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 201–218. IOS Press, Amsterdam, 1999.
8. J. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.
9. J. Pollock. Justification and defeat. *Artificial Intelligence*, 67(2):377–408, 1994.
10. J. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press, 1995.
11. J. Pollock. Perceiving and reasoning about a changing world. *Computational Intelligence*, 14(4):498–562, 1998.
12. J. Pollock and I. Oved. Vision, knowledge, and the mystery link. *Philosophical Perspectives* 19, pages 309–351, 2005.
13. R. Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13(1,2):81–132, Apr. 1980.
14. D. S. Touretzky, J. F. Harty, and R. H. Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings IJCAI*, pages 476–482, 1987.