Reasoning and probability

JOHN L. POLLOCK†
Department of Philosophy, University of Arizona, Tucson, AZ 85721, USA

[Revised on 24 May 2007; accepted on 25 May 2007]

Presented at the workshop on ‘Graphic and visual representations of evidence and inference in legal settings’ at Cardozo School of Law, New York City, 28–29 January 2007.

Reasoning justifies its conclusions to varying degrees, depending on the strength of the reasons and the justification of the evidence. It is often supposed that the resulting degrees of justification can be viewed as probabilities, but there are simple reasons why that cannot be right. This paper explores an alternative way of relating probabilities and degrees of justification.

Keywords: argument; evidence; probability; defeasible; reasoning.

1. Some puzzles about defeasible reasoning

Legal reasoning is, a fortiori, reasoning. No doubt, legal reasoning has many important properties not shared by all reasoning. But it also has many features that are common to most reasoning. Some of those general features are philosophically problematic, and if we do not understand them, that will affect our understanding of legal reasoning as well. An important feature of most reasoning is that it is defeasible. That is, our premises (evidence) make it reasonable to draw various conclusions, but they do not logically guarantee that the conclusions are true. The considerations that mandate the retraction of the conclusion are defeaters for the initial argument.

The term ‘defeasible reasoning’ originates with Hart (1951) writing in the philosophy of law. Legal scholars like Schum (1994) have continued to investigate varieties of defeasible legal reasoning, like reasoning from evidence. In the last 50 years, the ubiquity of defeasible reasoning has become generally appreciated and it has been the subject of intensive investigation both in philosophical logic and in artificial intelligence. However, I am unhappy with all existing work on defeasible reasoning, including my own (Pollock, 1987, 1995, 2002). I have two main problems with the existing work. First, some arguments are better than others, providing stronger support for their conclusions. Conclusions supported by better arguments are more highly justified, and we believe them with greater conviction. Furthermore, when we have arguments for competing conclusions, the strengths of the arguments are relevant to deciding which conclusion to accept (which is undefeated). Everyone acknowledges these facts, but most semantics for reasoning ignore them, and I believe that those semantics that try to accommodate varying degrees of justification get it wrong. This must be rectified if the formal semantics are to be useful in analysing the actual use of arguments.

We can think of a semantics for defeasible reasoning as having three parts. First, there must be an account of the reasoning itself. This consists of a set of rules telling us how arguments can be

† Email: pollock@arizona.edu
Arguments are built out of premises and inference schemes. The rules for argument construction tell us how to combine these into arguments. At the very least, we should be able to build and extend arguments by employing inference schemes to draw new conclusions from the premises and previously drawn conclusions. There may be other kinds of rules as well. For instance, we might be able to construct arguments by making hypothetical suppositions and then discharging them later using rules like conditionalization.

The second part of a semantics for defeasible reasoning consists of an account (hopefully an algorithm) for computing how strongly the conclusion of an argument is supported by the argument. Let us call this the argument strength of the argument. The argument strength should be a function of the degrees of justification of the initial premises and of what we might call the reason strengths of the inference schemes employed in the arguments. The reason strength of an inference scheme is a measure of how strongly it supports its conclusions. My first criticism of the existing work on the semantics of defeasible reasoning is that most semantics do not even attempt to provide an account of argument strengths. I (Pollock, 1995, 2002) did attempt to provide an account of argument strengths, but I am no longer convinced that my own account is correct.

In the absence of defeaters, the degree of justification of a conclusion should be determined by the argument strengths of the argument (or arguments) supporting it. However, things get more complicated when we have not only arguments supporting a conclusion but also arguments supporting defeaters for some of the steps of the first arguments. Then, how do we determine the degree of justification of the conclusion? The third part of a semantics for defeasible reasoning attempts to answer this question. We can combine all our arguments into a single structure, called an inference graph. Figures 1–3 are simple examples of inference graphs. Then, a semantics for defeasible reasoning attempts to compute the degrees of justification for all the nodes of the graph (i.e. all the conclusions drawn by any of the arguments). These degrees of justification should be determined by the initial premises of the arguments, the reason strengths of the inference schemes, and the structure of the graph which determines what supports what and what defeats what.

Existing work on defeasible reasoning has focussed largely on this third aspect of a semantics for defeasible reasoning, generally just looking at the simplified case in which all premises have the same degree of justification and all inference schemes have the same reason strength.¹ This leads to the second problem that I have with the existing work on the semantics for defeasible reasoning. This is that existing semantics for defeasible reasoning are largely ad hoc. They are carefully gerrymandered to yield the intuitively correct answer regarding what reasoning is semantically correct,

¹ Pollock (1995, 2002) tried to give a more general account that took account of varying degrees of justification and varying reason strengths.
but they provide no underlying rationale for their structure. For instance, if we look at a simple example like that in Fig. 1, the answer seems clear. Here, we suppose that we have equally justified initial premises $A$ and $B$ and $A$ gives us a reason for $P$ while $B$ gives us an equally good reason for $\sim P$. For instance, suppose that Jones is a potential witness to a crime. But let $A$ be ‘Paul says that Jones was looking the other way and did not see what happened’. Let $P$ be ‘Jones did not see what happened’. So $A$ gives us a reason for believing $P$. But let $B$ be ‘Jacob says that Jones was watching carefully and had a clear view of the gunman’. $B$ gives us a reason for believing $\sim P$, viz., that Jones did see what happened. If we regard Paul and Jacob as equally reliable, their conflicting testimonies defeat one another. What should we believe about whether Jones saw what happened? It seems clear that, in the absence of any other information, we should withhold belief, believing neither $P$ nor $\sim P$. In other words, the conclusions $P$ and $\sim P$ should be completely unjustified.

Semantics for defeasible reasoning have been crafted in this way, by eliciting our intuitions about particular cases and then looking for general semantical rules that capture those intuitions. This is not automatically a bad way to go about designing a semantics although it would be better to have a semantics whose details are motivated by some explanation for why our intuitions go the way they do. But the real problem with this approach is that our intuitions are unclear in many cases. Different semantics have been constructed that agree with respect to those cases in which our intuitions are perfectly clear, but differ with respect to some cases in which people lack clear intuitions. An example is what Makinson & Schlechta (1991) called ‘zombie arguments’. Consider the inference graph diagrammed in Fig. 2. Again, the heavy arrows indicate defeat relations. In this example, as in Fig. 1, we have equally good arguments for $P$ and $\sim P$ and an argument from $P$ to a defeater for a third argument. Everyone agrees that the arguments for $P$ and $\sim P$ should be defeated.

A possible exception is circumscription which is based on an intuitive idea. However, circumscription does not withstand the test of supporting the intuitively right reasoning, and it does not have anything to say about degrees of justification.
This is a case of ‘collective defeat’. But it is unclear whether \( P \) should retain the ability to defeat the third argument. If so, the argument for \( P \) is a kind of ‘zombie’—it is dead, but it can get you anyway. Most formal semantics for defeasible reasoning license zombie arguments, but it is not clear they should. I, at least, have no strong intuitions about this one way or the other. To illustrate the phenomenon, consider again the case of Jones who is a potential witness to a crime. Let \( C \) be ‘Jones says that the gunman had a mustache’, and let \( Q \) be ‘The gunman had a mustache’. \( C \) gives us a reason for believing \( Q \). Again, let \( A \) be ‘Paul says that Jones was looking the other way and did not see what happened’ and let \( P \) be ‘Jones did not see what happened’. So \( A \) gives us a reason for believing \( P \) and \( P \) is a defeater for the inference \( C \) to \( A \). But letting \( B \) be ‘Jacob says that Jones was watching carefully and had a clear view of the gunman’, \( B \) gives us a reason for believing \( \sim P \), viz., that Jones did see what happened. Once again, if we regard Paul and Jacob as equally reliable, their conflicting testimonies defeat one another. Given the conflict, should we accept Jones’ claim that the gunman had a mustache? I am not sure. Most researchers on defeasible reasoning disavow any clear intuitions about this case. But then, how do we decide between a semantics that licenses zombie arguments and one that does not?

A different kind of puzzle concerns how argument strengths interact. Consider once more the simple case of collective defeat diagrammed in Fig. 1. For example, \( A, B \) and \( P \) might be as in the preceding example. But now suppose that we regard Paul as more reliable than Jacob. Then, the argument for \( P \) is stronger than the argument for \( \sim P \). Most people have the intuition that the justification for \( P \) should be ‘diminished’, viz., if the difference between the strengths of the arguments is significant, \( P \) might still be justified but to a lesser degree. But how much less? How should we combine the reason strengths to compute the overall degrees of justification? I find untutored intuition of little use on this question.

A related question concerns what happens if we have two independent arguments for a conclusion, as in Fig. 3. Should this increase the justification for the conclusion or should the justification for the conclusion be the same as if we just had the stronger of the two arguments? Most people have the intuition that the justification should increase, but if so, how much?

Semantics based on untutored intuition seem to be incapable of answering these questions in a compelling way. This has led me to look for a better motivated way of evaluating proposed semantics for defeasible reasoning. This paper presents some preliminary results that may, hopefully, eventuate in a better motivated semantics.

2. Degrees of justification and probabilities

2.1 Bayesian epistemology

There is a common intuition that degrees of justification should reflect how probable a conclusion is, and this has led some philosophers to identify degrees of justification with probabilities. Such an identification underlies Bayesian epistemology, and it would seem to provide an answer to both the preceding problems. However, there is a simple (and familiar) argument that convinces me that such an identification is untenable. The argument turns on the observation that, by the probability calculus, necessary truths have probability 1. Necessary truths include truths of logic, truths of mathematics and anything else that is not just true as a contingent matter of fact. If degrees of justification were probabilities, it would follow that we are always completely justified in believing anything that is in fact a necessary truth, even if we have no argument for it and no reason to think that it is a necessary truth. For example, we now know that Fermat’s conjecture is a truth of mathematics, and hence a
necessary truth. Its being a necessary truth is just dependent on its being a truth of mathematics, not on its having been proven. It would follow that we were justified in believing Fermat’s conjecture all along, before Andrew Wiles found his famous proof of it because it was always a mathematical truth. And had it been false rather than true, it would follow that we would have been justified in believing it false even without a proof to that effect. But we were not. Nobody had a justified belief about it prior to Wiles’s proof.

2.2 A criterion of adequacy?

The preceding objection seems compelling. Degrees of justification cannot have the mathematical structure of the probability calculus, and accordingly cannot be probabilities. However, there are other more indirect ways that probabilities could be related to reasoning. Let us step up a level and suppose that we are designing a cognitive agent that reasons about its world and acts on its beliefs. Some designs are going to be better than others, and we can evaluate them in terms of how well the agent gets along in its world. One desideratum for a system of reasoning is that, given accurate input, it will tend to produce true beliefs. It cannot be guaranteed to produce true beliefs—that would require the agent to reason only deductively, and we know that deductive reasoning is inadequate for an agent functioning in a sufficiently complex environment. Such an agent has to be able to draw conclusions provisionally and be prepared to withdraw them later if further information makes them unreasonable. In other words, the agent must be able to reason defeasibly. Although we cannot require the agent to employ a system of reasoning that guarantees the truth of its conclusions, it seems that a constraint on the agent’s system of defeasible reasoning should be that it tends to produce true beliefs. How, exactly, should we understand this constraint?

It is tempting to require that a system of reasoning should be organized so that conclusions drawn on the basis of some initial set of premises must always have high probability conditional on those premises and the degree of justification attached to the conclusion should reflect that conditional probability. This avoids the preceding difficulty because it assigns no degree of justification to a necessary truth for which we have not constructed an argument. However, this still has the consequence that, restricted to the set of conclusions drawn by the system of reasoning, degrees of justification are probabilities. This implies that if we do have an argument for a necessary truth, then because its probability is 1, we are completely justified in believing it. This overlooks the fact that we can have weak arguments for necessary truths. For instance, if one of my undergraduate students tells me that a particular mathematical principle is true, that gives me a reason for believing it. But if one of my colleagues in the mathematics department tells me the same thing, that surely gives me a better reason for believing it. And if I actually work through a proof of the principle, that may give me an even better reason. The degree of justification for my belief must reflect the argument supporting the belief and not just the fact that the belief is a necessary truth. However, what argument leads me to the belief is irrelevant to whether the probability of the mathematical principle is 1, so the latter cannot determine the degree of justification.

The insight to be gleaned from these considerations is that the degree of justification of a belief should be determined by what arguments have been mustered in its support (or more generally, to take account of defeat, by the inference graph consisting of all the cognizer’s reasoning to date) and should be independent of the content of the belief. Of course, the content will indirectly affect what arguments the cognizer constructs, but aspects of the content that are not reflected in the arguments should not affect the degree of justification. The difficulty with simple probabilistic accounts of
degrees of justification is that the probability of the conclusion is sensitive to aspects of the content that need not be reflected by the arguments. In the limiting case of necessary truths, the content determines that the probability is 1, even if the cognizer has no good arguments for the conclusion.

Still, it seems reasonable to require the agent’s system of defeasible reasoning to be such that it tends to produce true beliefs. Let us see whether there is a way of understanding this constraint that will relate probabilities and degrees of justification in a less problematic way.

2.3 Argument strengths

When we reason, the individual steps of inference proceed in terms of inference schemes (either deductive or defeasible) that convey greater or lesser degrees of justification. The inference schemes have reason strengths associated with them, and the degree of justification of the conclusion should be a function of the reason strength and the degrees of justification of the premises to which the inference scheme is applied. It is plausible to insist that in a well-designed system of reasoning, the reason strength of an inference scheme should reflect the general probability of its producing a true conclusion given that the premises are true. In a deductive argument, these conditional probabilities may all be 1. But when I believe a necessary truth because I am told by a mathematics professor that it is true, the conditional probability is high but less than 1, and when I believe it on the basis of the testimony of my undergraduate student, the conditional probability is lower still. This suggests that it is these conditional probabilities that ought to be reflected in the computation of the degrees of justification of conclusions.

In developing this suggestion, it is important to distinguish between general or statistical probabilities and what are sometimes called single-case probabilities. Historically, there have been two general approaches to probability theory. The most familiar takes ‘definite’ or ‘single-case’ probabilities to be basic. Definite probabilities attach to closed formulas or propositions. I write them using small caps: \( \text{PROB}(P) \) and \( \text{PROB}(P/Q) \). To be contrasted with definite probabilities are ‘indefinite’ or ‘general’ probabilities (sometimes called ‘statistical probabilities’). The indefinite probability of an \( A \) being a \( B \) is not about any particular \( A \), but rather about the property of being an \( A \). In this respect, its logical form is the same as that of relative frequencies. I write indefinite probabilities using lower case ‘prob’ and free variables: \( \text{prob}(Bx/Ax) \). For instance, we can talk about the probability of a person with certain symptoms having a particular disease. This probability is about the properties of having certain symptoms and having the disease. This probability is not directly about any particular person having those symptoms. If we examine a particular patient and determine conclusively that he has the disease, then the definite probability of his having the disease is 1.0, but that does not alter the indefinite probability of persons in general having the disease if they have the symptoms.

The distinction between definite and indefinite probabilities is commonly overlooked by contemporary probability theorists, perhaps because of the popularity of subjective probability (which has no way to make sense of indefinite probabilities). But most objective approaches to probability tie probabilities to relative frequencies in some essential way, and the resulting probabilities have the same logical form as the relative frequencies. That is, they are indefinite probabilities. The simplest theories identify indefinite probabilities with relative frequencies (Russell, 1948; Braithwaite, 1953; Kyburg, 1961, 1974b; Sklar, 1970, 1973).³ The simplest objection to such ‘finite-frequency theories’

³ Kneale (1949) traces the frequency theory to R. L. Ellis, writing in the 1840s, and Venn (1888) and C. S. Peirce in the 1880s and 1890s.
is that we often make probability judgements that diverge from relative frequencies. For example, we can talk about a coin being fair (and so, the indefinite probability of a flip landing heads is 0.5) even when it is flipped only once and then destroyed (in which case, the relative frequency is either 1 or 0). For understanding such indefinite probabilities, we need a notion of probability that talks about possible instances of properties as well as actual instances. Theories of this sort are sometimes called ‘hypothetical frequency theories’. C. S. Peirce was perhaps the first to make a suggestion of this sort. Similarly, the statistician R. A. Fisher (1922, p. 311), regarded by many as ‘the father of modern statistics’, identified probabilities with ratios in a ‘hypothetical infinite population, of which the actual data is regarded as constituting a random sample’. Popper (1956, 1957, 1959) endorsed a theory along these lines and called the resulting probabilities propensities. Kyburg (1974a) was the first to construct a precise version of this theory (although he did not endorse the theory), and it is to him that we owe the name ‘hypothetical frequency theories’. Kyburg (1974a) also insisted that von Mises should also be considered a hypothetical frequentist. There are obvious difficulties for spelling out the details of a hypothetical frequency theory. More recent attempts to formulate precise versions of what might be regarded as hypothetical frequency theories are van Fraassen (1981), Bacchus (1990), Halpern (1990), Pollock (1990, 2008) and Bacchus et al. (1996). I will be making essential use of some aspects of the theory of Pollock (1990, 2008), some parts of which I will sketch briefly in Section 3.

The distinction between definite and indefinite probabilities gives us the mechanism for avoiding the problem that necessary truths have probability 1. That is a remark about definite probabilities. I propose instead that when we talk about the probability of an inference scheme producing a true conclusion from true premises, this should be understood as an indefinite (general) probability. It is about conclusions in general that are inferred in this way. This has the effect of insulating the probability from the content of the conclusion because we are not talking about particular conclusions. The conclusion is represented by a free variable in the relevant indefinite probability. If we ask instead about the probability that a particular proposition is true given that it is inferred in this way from true premises, that probability is sensitive to the content of the conclusion. When the proposition is a necessary truth, then the latter conditional probability is 1. Thus, we cannot distinguish between different cases of believing a necessary truth and assign different degrees of justification to our beliefs if we focus only on definite probabilities. It is the different indefinite probabilities associated with different inference schemes that lead to the differences between the cases.

2.4 A criterion of adequacy—second try

If we can evaluate the individual steps of inference in terms of the indefinite conditional probability of the conclusion being true given that the premises are true, it seems plausible that we might evaluate the entire system of reasoning in the same way. This suggests the following probabilistic criterion of adequacy for a system of reasoning:

The indefinite probability should be high of a conclusion being true given that it is supported by an argument proceeding from true premises.

As formulated, this is not a probabilistic constraint on individual inferences. It is a constraint on the entire system of reasoning. It does not say that individual argument schemes have high indefinite probabilities of producing true conclusions. However, the only obvious way to guarantee that this constraint is satisfied is to require it to be true of all argument schemes produced by the
system. Furthermore, if there were particular argument schemes for which this is not true, we could increase the probability with which the system produces true conclusions by disallowing those argument schemes. So it seems reasonable to impose this requirement on all the argument schemes produced by the system. Let us make this more precise. Let us suppose that $P_1, \ldots, P_n \vdash Q$ is an inference scheme built into the reasoning system. It will be important that inference schemes are always general schemes. That is, $P_1, \ldots, P_n$ and $Q$ are open formulas containing free variables. When we apply the inference scheme, we instantiate the variables. Then, there is some number $r$ such that $\text{prob}(Q / P_1 \& \cdots \& P_n) = r$. Because $P_1, \ldots, P_n$ and $Q$ are open formulas, this is an indefinite (general, not single case) probability. Let us call this probability the reliability of the inference scheme.

The agent employs this reasoning scheme without having to have independent evidence for the high reliability of its inference schemes. The inference schemes are simply built into his reasoning system, and for him, proper reasoning consists in reasoning in accordance with these inference schemes. However, from the perspective of the designer of the cognitive system, an inference scheme would be included in the agent architecture because of its high reliability. Furthermore, assuming that the system assigns reason strengths to inference schemes and uses those to adjudicate conflicts and in deciding questions of defeat, the reason strength should track the value of this reliability. For present purposes then, let us suppose that reason strengths are calibrated in such a way that they purport to be the associated reliabilities. Part of our criterion of adequacy will then be that if $P_1, \ldots, P_n \vdash Q$ is one of the reasoner’s built-in inference schemes and its reason strength is $r$, then if the reasoner is given the information $P_1, \ldots, P_n$ (and nothing else) and reasons from that to $Q$, the probability of his conclusion conditional on his evidence is $r$.

An individual argument is built, using the principles of the system of reasoning, by combining particular premises and specific inference schemes. Let an argument scheme be the schematic structure of the argument. Different concrete arguments result from instantiating the argument scheme with different premises and different inference schemes. It seems initially reasonable to suppose that there should be a fixed function associated with an argument scheme which computes an argument strength on the basis of the reason strengths of the inference schemes and the structure of the argument scheme. If the reason strengths of the inference schemes reflect their probabilities of producing true conclusions from true premises, then we might require that the computed argument strength should reflect the conditional indefinite probability of the conclusion of such an argument being true given that the premises are true. Let the latter probability be the reliability of the argument scheme.

Once we realize that arguments have argument strengths attached to them and some do not purport to be strong arguments, it is all right if the system allows the construction of arguments having low reliability as long as they are also assigned low argument strengths. So it seems that we should modify our previously proposed criterion of adequacy to require that the argument strengths of argument schemes track their reliabilities.

### 2.5 Degrees of justification

Our criterion of adequacy still needs tweaking. As stated so far, it overlooks two facts. First, it has been proposed that the argument strength should track the indefinite probability of the conclusion being true conditional on the premises being true. However, the agent may have other arguments that are relevant to the conclusion, and the probability of the conclusion conditional on that larger set of arguments need not be the same as the argument strength. Put in familiar terms, the larger set of arguments may support defeaters for the first argument.
Expanding on this first point, what an agent should believe is determined not by the individual arguments he has constructed but by the inference graph consisting of the totality of arguments he has constructed. This should already be clear in light of familiar work on defeasible reasoning, where distinct arguments interact and defeat one another. Rather than computing argument strengths for individual arguments, what we need to compute are the degrees of support of conclusions relative to entire inference graphs. This can be understood on analogy to argument strengths. Let the inference graph scheme consist of the argument schemes for the arguments comprising the inference graph. Then, the probabilistic degree of support of a conclusion relative to the inference graph is the indefinite probability of an arbitrary conclusion (occupying that place in the inference graph scheme) being true conditional on (1) the premises of the inference graph scheme being true and (2) the constituent inference schemes having reliabilities equal to their stipulated reason strengths. A semantics for reasoning (both deductive and defeasible) is supposed to provide a way of computing degrees of justification for conclusions given inference graphs. So as a first pass, it seems we should require the computed degree of justification to be the same as the probabilistic degree of support.

This is not quite right, however. There is a second difficulty for the proposed criterion of adequacy. Whether the cognitive agent accepts a conclusion should depend not just on the degree of support, as just defined, but also on the degrees of justification of the premises. In evaluating a system of reasoning, we can suppose that we supply premises having certain definite probabilities (viewed from our perspective—i.e. from the perspective of the external evaluator). We assign those definite probabilities as the stipulated degrees of justification of the premises. Then, we might try redefining the probabilistic degree of support of a conclusion relative to the inference graph to be the indefinite probability of an arbitrary conclusion (occupying that place in the inference graph scheme) being true conditional on (1) the premises of the inference graph scheme having the stipulated definite probabilities and (2) the constituent inference schemes having reliabilities equal to their stipulated reason strengths. To illustrate, consider the simple inference graph diagrammed, along with the associated probabilities, in Fig. 4.

The proposal is that the probabilistic degree of support of the conclusion in Fig. 4 should be

\[
\begin{align*}
\text{prob}_{A,B,P,Q,c}((Pc & Qc)/\text{PROB}(Ac) = a & \text{PROB}(Bc) = b & \text{prob}_x(Px/Ax) = r & \text{prob}_x(Qx/Bx) = s).
\end{align*}
\]

To evaluate this proposal, we must understand how \text{PROB} and \text{prob} are connected. I have proposed elsewhere (Pollock, 1990, 2006) that definite probabilities can be identified with indefinite probabilities conditional on a knowledge base \(K\). Specifically, \(\text{PROB}(Ac) = \text{prob}(Ax/x = c & K)\). I will assume without argument that this is correct. The degree of justification relative to an inference graph should be independent of the particular knowledge base, so it is natural to reconstrue the

![Fig. 4. A simple inference graph.](https://academic.oup.com/lpr/article-abstract/6/1-4/43/962266)
probabilistic degree of support as involving variable $K$ (where now we roll the ‘$x = c$’ condition into $Kx$):

$$
\text{prob}_{A,B,P,Q,K,x}( (Pc \& Qc)/\text{prob}_x(Ax/Kx) = a \& \text{prob}_x(Bx/Kx) = b \& \text{prob}_x(Px/Ax)
\ = r \& \text{prob}_x(Qx/Bx) = s ).
$$

Once we make the role of $K$ explicit, it seems clear that this cannot be the correct definition. As the probabilities of $Ax$ and $Bx$ are conditional on $Kx$, it seems that the assessment of $(Pc \& Qc)$ should also be conditional on $Kx$. In other words, the probabilistic degree of support should be

$$
\text{prob}_{A,B,P,Q,K,x}( (Px \& Qx)/Kx \& \text{prob}_x(Ax/Kx) = a \& \text{prob}_x(Bx/Kx) = b \& \text{prob}_x(Px/Ax)
\ = r \& \text{prob}_x(Qx/Bx) = s ).
$$

The proposed criterion of adequacy would then be that the reasoning system computes the degrees of justification (relative to an inference graph) that track the values of these probabilities.

It might seem doubtful that such probabilities actually exist. Even if they do, if the system is to be computing these probabilities a priori to use as degrees of justification, they would have to be determined by purely mathematical considerations, and that also may seem dubious. Surprisingly, recent work in probability theory establishes that these probabilities do exist and are determined by purely mathematical considerations. Suppose that we have a set of premises, an inference graph and a conclusion and we are given the reliabilities of the inference schemes and the degrees of justification (in the form of definite probabilities) of the premises. Then, there will be a unique value $r$ that is the indefinite probability of a conclusion being true in an (arbitrary) inference graph having the form of this inference graph scheme, conditional on (1) $Kx$, (2) the reason strengths of its inference schemes and (3) the definite probabilities of its premises. I will suggest that we take the value $r$ of this indefinite probability to be the probabilistic degree of support of the conclusion relative to that inference graph. Furthermore, there is an algorithm for computing the probabilistic degree of support, so defined. For instance, it turns out that the probabilistic degree of support of $(Pc \& Qc)$ relative to the inference graph of Fig. 4 is $(a \cdot r + \frac{1-a}{2}) \cdot (b \cdot s + \frac{1-b}{2})$. The proposed criterion of adequacy for a system of reasoning is then that the computed degrees of justification of conclusions track these probabilistic degrees of support. A digression is in order while I explain where these probabilities come from and how they can be computed.

3. Probable probabilities

The results I will now discuss arise specifically in the theory of nomic probability, as developed in my books and papers (Pollock, 1990, 2006, 2008). However, I will not go into the details of that theory here. The crucial idea that underlies these results is a common and intuitively compelling one. This is that there is a connection between the mathematical structure of probabilities and the behaviour of relative frequencies in finite sets, particularly as the sizes of the sets go to infinity. The theory of ‘probable probabilities’ (Pollock, 2008) derives from a set of straightforward, but often surprising, combinatorial theorems about finite sets. For instance, suppose that we have a set of 10 000 000 objects. I announce that I am going to select a subset, and ask you how many members it will have. Most people will protest that there is no way to answer this question. It could have any number of members from 0 to 10 000 000. However, if you answer ‘approximately 5 000 000’, you will almost
certainly be right. This is because, although there are subsets of all sizes from 0 to 10,000,000, there are many more subsets whose sizes are approximately 5,000,000 than there are of any other size. In fact, 99% of the subsets have cardinalities differing from 5,000,000 by less than 0.08%. To state this result precisely, we need some notation. Let $\#X$ be the cardinality of a set $X$. Given finite sets $X$ and $Y$, let $\rho(X, Y)$ be the proportion of $Y$’s that are $X$’s. That is, $\rho(X, Y) = \frac{\#(X \cap Y)}{\#Y}$. Given open formulas $Fx$ and $Gx$, let $\rho_x(Fx/Gx)$ be the proportion (relative frequency) of $G$’s that are $F$’s. That is, $\rho_x(Fx/Gx) = \frac{\#\{x : (Fx & Gx)\}}{\#\{x : Fx\}}$.

Let ‘$x \approx y$’ mean ‘the difference between $x$ and $y$ is less than or equal to $\delta$’. The general theorem is then the following.

**Finite-indifference principle**

For every $\varepsilon, \delta > 0$, there is an $N$ such that if $U$ is finite and $\#U > N$, then

$$\rho_X \left( \rho(X, U) \approx \frac{0.5}{\delta} \bigg/ X \subseteq U \right) \geq 1 - \varepsilon.$$ 

In other words, the proportion of subsets of $U$ which are such that $\rho(X, U)$ is approximately equal to 0.5, to any given degree of approximation, goes to 1 as the size of $U$ goes to infinity. To see why this is true, suppose $\#U = n$. If $r \leq n$, the number of $r$-membered subsets of $U$ is $C(n, r) = \frac{n!}{r!(n-r)!}$. It is illuminating to plot $C(n, r)$ for variable $r$ and various fixed values of $n$, see Fig. 5. This illustrates that the sizes of the subsets of $U$ will cluster around $\frac{n}{2}$, and they cluster more tightly as $n$ increases. This is precisely what the indifference principle tells us.

The reason the indifference principle holds is that $C(n, r)$ becomes ‘needle-like’ in the limit. As we proceed, I will state a number of similar combinatorial theorems, and in each case, they have similar intuitive explanations. The cardinalities of relevant sets are products of terms of the form $C(n, r)$, and their distribution becomes needle-like in the limit. Within the theory of nomic probability, such limit principles regarding finite sets entail analogous principles about indefinite probabilities. For example, where ‘$X \preceq G$’ means ‘$X$ is a subproperty of $G$’ (i.e. $X$ entails $G$), the finite-indifference principle entails.
Probabilistic indifference principle

For any property \(G\) and for every \(\delta > 0\), \(\operatorname{prob}_{X}(\operatorname{prob}(X/G) \approx 0.5 / X \preceq G) = 1\).

I will formulate this more simply by saying ‘given that \(X \preceq G\), the expectable value of \(\operatorname{prob}(X/G)\) is 0.5’. In my paper (Pollock, 2008), I enumerate a number of interesting limit principles and the corresponding principles of probable probability. Here are a few of the latter.

Principle of expectable statistical independence

If \(\operatorname{prob}(A/C) = r\) and \(\operatorname{prob}(B/C) = s\), the expectable value of \(\operatorname{prob}(A \& B/C) = r \cdot s\).

Non-classical direct inference

If \(\operatorname{prob}(A/B) = r\), the expectable value of \(\operatorname{prob}(A/B \& C) = r\).

Y-Principle

If \(B, C \preceq U\), \(\operatorname{prob}(A/B) = r\), \(\operatorname{prob}(A/C) = s\) and \(\operatorname{prob}(A/U) = a\), then the expectable value of \(\operatorname{prob}(A/B \& C) = Y(r, s, a) = \frac{rs(1-a)}{a(1-r-s)+rs}\).

It turns out that there is a mechanical way of generating results like this. Given a list of variables \(X_1, \ldots, X_n\) ranging over subsets of a set \(U\), Boolean compounds of these sets are compounds formed by union, intersection and set complement. So, e.g. \((X \cup Y) - Z\) is a Boolean compound of \(X, Y\) and \(Z\). Linear constraints on the Boolean compounds either state the values of certain proportions, e.g. stipulating that \(\rho(X, Y) = r\), or relate proportions using linear equations. For example, if we know that \(X = Y \cup Z\), that generates the linear constraint

\[\rho(X, U) = \rho(Y, U) + \rho(Z, U) - \rho(X \cap Z, U)\]

We can prove the following general theorem.

Probable probabilities theorem

Let \(U, X_1, \ldots, X_n\) be a set of variables ranging over sets, and consider a finite set \(\text{LC}\) of linear constraints on proportions between Boolean compounds of those variables. If \(\text{LC}\) is consistent with the probability calculus, then for any pair of Boolean compounds \(P, Q\) of \(U, X_1, \ldots, X_n\), there is a real number \(r\) between 0 and 1 such that for every \(\varepsilon, \delta > 0\), there is an \(N\) such that if \(U\) is finite and \(#U > N\), then

\[\rho_{X_1, \ldots, X_n}(\rho(P, Q) \approx r / \text{LC} \& X_1, \ldots, X_n \subseteq U) \geq 1 - \varepsilon\]

Within the theory of nomic probabilities, this entails the following principle.
Expectable probabilities principle

Let \( U, X_1, \ldots, X_n \) be a set of variables ranging over properties and relations, and consider a finite set \( LC \) of linear constraints on probabilities between truth-functional compounds of those variables. If \( LC \) is consistent with the probability calculus, then for any pair of truth-functional compounds \( P, Q \) of \( U, X_1, \ldots, X_n \), there is a real number \( r \) between 0 and 1 such that, given the constraints \( LC \), the expectable value of \( \text{prob}(P/Q) = r \).

This establishes the existence of expectable values for probabilities under very general circumstances. Given a set of linear constraints, there is an algorithm for generating a set of simultaneous equations which characterize the expectable values for all the variables, and I have written a program in LISP to do this. In some cases, the equations have analytic solutions, which can often be found automatically by computer algebra programs, and this generates principles like the above. More often, they do not have analytic solutions, but they can still be solved numerically to compute the expectable values of probabilities in particular cases.

4. Degrees of justification and probable probabilities

Now, let us return to the suggestion that we might be able to design a reasoning system so that the conclusions will \textit{probably} have high definite probability. We can formulate the suggestion more precisely as follows:

Given an inference graph \( A \), it can be regarded as an instantiation of an inference graph scheme \( A^* \). \( A^* \) contains variables \( x_1, \ldots, x_n \) for inference schemes and variables \( y_1, \ldots, y_k \) for premises. \( A \) results from instantiating \( x_1, \ldots, x_n \) with some particular inference schemes \( R_1, \ldots, R_n \) and instantiating \( y_1, \ldots, y_k \) with some premises \( P_1, \ldots, P_k \). \( R_1, \ldots, R_n \) have reason strengths \( r_1, \ldots, r_n \), which we interpret as above as indefinite probabilities. \( P_1, \ldots, P_k \) have degrees of justification \( a_1, \ldots, a_k \), which we interpret as definite probabilities. The \textit{probabilistic degree of support of the conclusion relative to} \( A \) is the indefinite probability that the corresponding conclusion of an arbitrary inference graph is true given that the inference graph is an instance of \( A^* \) and employs inference schemes with reliabilities \( r_1, \ldots, r_n \) and premises with definite probabilities \( a_1, \ldots, a_k \).

Let me illustrate the suggestion by looking at the argument diagrammed in Fig. 4. In that case, the probabilistic degree of support of the conclusion is the real number \( p \) such that

\[
\text{prob}_{A,B,P,Q,K,x}( (P x & Q x)/K x \& \text{prob}_x(A x/K x) = a \& \text{prob}_x(B x/K x) = b \& \text{prob}_x(P x/A x) = r \& \text{prob}_x(Q x/B x) = s ) = p.
\]

Why should we think that such a number \( p \) exists? This turns on the following theorem of the theory of nomic probability.

Projection principle

If for every \( \delta > 0 \), \( \text{prob}_y \left( \text{prob}_x(F x/G x y) \approx r / H y \right) = 1 \), then \( \text{prob}_{x,y}(F x/G x y \& H y) = r \).
The intuitive idea here is that if for all y’s that are H except for a subset of measure 0, prob_x(Fx/Gxy) has a constant value r, then that is the value of prob_x,Fx,Gxy,Hy. Now, consider an argument for a conclusion Pc. The structure of the argument together with the values of the probabilities associated with the inference schemes and the values of the definite probabilities of the premises determines a set LC of linear constraints on the probabilities relating the variables occurring in the argument scheme. By the expectable probabilities principle, there is an expectable value r for prob_x,Px/Kx relative to LC. That is, for every δ > 0,

\[ \text{prob}_{P,K,\ldots}(Px/Kx) \approx r \bigg/ \text{LC} = 1. \]

But then by the projection principle,

\[ \text{prob}_{P,K,\ldots,x}(Px/Kx & \text{LC}) \approx r. \]

The latter is true for every δ > 0, so this entails that

\[ \text{prob}_{P,K,\ldots,x}(Px/Kx & \text{LC}) = r. \]

The latter is the probabilistic degree of support of Pc relative to that argument. Thus, the probabilistic degree of support is the same as the expectable value of prob_x,Px/Kx, and we know that the latter always exists. So, e.g. the probabilistic degree of support of the conclusion (Pc & Qc) in Fig. 4 is the unique real number p such that

\[ \text{prob}_{A,B,P,Q,K}(\text{prob}_x(Px & Qx/Kx) \approx p \bigg/ \text{prob}_x(Ax/Kx) = a \& \text{prob}_x(Bx/Kx) = b \& \text{prob}_x(Px/Ax) = r \& \text{prob}_x(Qx/Bx) = s) = 1. \]

The unique p satisfying this condition is \((a \cdot r + \frac{1-a}{2}) \cdot (b \cdot s + \frac{1-b}{2})\), so on the proposed definition, this is the probabilistic degree of support of (Pc & Qc) relative to this inference graph. Degrees of justification are supposed to track these probabilistic degrees of support. Thus, degrees of justification become identified with the expectable values of the definite probability of a conclusion.

Note that p is a function of the structure of A (i.e. of the argument scheme A∗), the reason strengths of the inference schemes occurring in A and the definite probabilities of the premises but not of the actual content of the inference schemes or premises. Some arguments (comprising a set of measure 0) having the same form as A and employing inference schemes of the same strength and premises with the same definite probabilities may nevertheless be such that the definite probabilities of their conclusions are different from p. Thus, e.g. an argument supporting Q on the basis of expert testimony (for fixed degree of expertise) will have the same strength p regardless of what Q is. But if Q is a necessary truth, the definite probability of Q will be 1 rather than p. So the degree of justification relative to the argument measures what we should expect the probability of Q to be without taking account of the content of Q. In other words, it tells us how much support the argument gives for Q, independent of the content of Q and hence independent of what other arguments might be mustered for or against Q.
Suppose that we infer a conclusion $Q$ from premises $P_1, \ldots, P_n$ using an argument $A$. If the probabilistic degree of support for $Q$ relative to that argument is $p$, then we have a defeasible reason for expecting that $\text{PROB}(Q) = p$. For example, in the argument diagrammed in Fig. 4, we know that $\text{prob}(P_1/Ax) = r$, $\text{prob}(Q_1/Bx) = s$, $\text{prob}(A_1/Kx) = a$ and $\text{prob}(B_1/Kx) = b$, so we can infer defeasibly that

$$\text{prob}(P_1 \& Q_1/K) = (a \cdot r + \frac{1-a}{2}) \cdot (b \cdot s + \frac{1-b}{2}).$$

Equivalently,

$$\text{PROB}(P_1 \& C) = (a \cdot r + \frac{1-a}{2}) \cdot (b \cdot s + \frac{1-b}{2}).$$

Generally, defeaters for the preceding reasoning will result from having other arguments pertaining to some of the predicates in the argument $A$. If we combine all those arguments into a single inference graph, we can compute the expected probability of the conclusion given that larger set of arguments, and that will then be what we should defeasibly expect the definite probability to be.

I suggest, then, that our criterion of adequacy for a system of reasoning should be that the system computes degrees of justification in such a way that, if the built-in reason strengths of the inference schemes are their associated reliabilities and the stipulated degrees of justification are their definite probabilities, then the computed degrees of justification are the probabilistic degrees of support as defined here. This has the result that, with probability 1, the degree of justification is equal to the definite probability of the conclusion.

5. Conclusions

I have proposed a way of defining degrees of justification in terms of probabilities in a way that avoids many of the difficulties that beset more familiar attempts to give probabilistic accounts of this concept. In particular, degrees of justification are functions just of the forms of the arguments supporting conclusions, and are not sensitive to what propositions occur in the arguments. This is what allows us to avoid the familiar difficulty concerning necessary truths. One reason degrees of justification, so defined, are of interest is that it is defeasibly reasonable to expect the definite probability of the conclusion to be the same as its degree of justification although that is not logically guaranteed. However, this is only a preliminary proposal. It remains to be seen precisely what consequences this has for a semantics of reasoning. This is a topic for future research.

Funding

National Science Foundation (IIS-0412791).

REFERENCES


BACCHUS, FAHIEM, ADAM J. GROVE, JOSEPH Y. HALPERN, DAPHNE KOLLER 1996 From statistical knowledge bases to degrees of belief. Artificial Intelligence 87, 75–143.

POLLOCK, JOHN 2008 Reasoning Defeasibly about Probabilities, in Michael O’Rourke and Joseph Campbell, eds., Knowledge and Skepticism, Cambridge, MA: MIT Press.